

Germline *PARP4* mutations in patients with primary thyroid and breast cancers

Yuji Ikeda¹, Kazuma Kiyotani¹, Poh Yin Yew¹, Taigo Kato¹, Kenji Tamura¹, Kai Lee Yap¹, Sarah M Nielsen¹, Jessica L Mester², Charis Eng^{2,3}, Yusuke Nakamura¹ and Raymon H Grogan⁴

¹Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, Illinois 60637, USA

²Genomic Medicine Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA

³Department of Genetics and Genome Sciences, Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA

⁴Endocrine Surgery Research Program, Section of General Surgery, Department of Surgery, The University of Chicago, 5841 S Maryland Avenue, Chicago, Illinois 60637, USA

Correspondence should be addressed to R H Grogan
Email
rgrogan@surgery.bsd.uchicago.edu

Abstract

Germline mutations in the *PTEN* gene, which cause Cowden syndrome, are known to be one of the genetic factors for primary thyroid and breast cancers; however, *PTEN* mutations are found in only a small subset of research participants with non-syndrome breast and thyroid cancers. In this study, we aimed to identify germline variants that may be related to genetic risk of primary thyroid and breast cancers. Genomic DNAs extracted from peripheral blood of 14 *PTEN* WT female research participants with primary thyroid and breast cancers were analyzed by whole-exome sequencing. Gene-based case-control association analysis using the information of 406 Europeans obtained from the 1000 Genomes Project database identified 34 genes possibly associated with the phenotype with $P < 1.0 \times 10^{-3}$. Among them, rare variants in the *PARP4* gene were detected at significant high frequency (odds ratio = 5.2; $P = 1.0 \times 10^{-5}$). The variants, G496V and T1170I, were found in six of the 14 study participants (43%) while their frequencies were only 0.5% in controls. Functional analysis using HCC1143 cell line showed that knockdown of *PARP4* with siRNA significantly enhanced the cell proliferation, compared with the cells transfected with siControl ($P = 0.02$). Kaplan–Meier analysis using Gene Expression Omnibus (GEO), European Genome-phenome Archive (EGA) and The Cancer Genome Atlas (TCGA) datasets showed poor relapse-free survival ($P < 0.001$, Hazard ratio 1.27) and overall survival ($P = 0.006$, Hazard ratio 1.41) in a *PARP4* low-expression group, suggesting that *PARP4* may function as a tumor suppressor. In conclusion, we identified *PARP4* as a possible susceptibility gene of primary thyroid and breast cancer.

Key Words

- ▶ thyroid
- ▶ breast
- ▶ cancer
- ▶ *PARP4*
- ▶ germline mutation

Endocrine-Related Cancer
(2016) 23, 171–179

Introduction

A cancer of a different site and histologic type that develops in a person who already has a cancer diagnosis is considered a second primary cancer. Certain primary

cancers are associated with a high risk of developing a second primary cancer. Specifically, individuals with a primary thyroid cancer have a high incidence of second

primary cancers even after adjusting for surveillance bias. One study of a cohort of 39 002 research participants suggested that when compared to the general population, thyroid cancer survivors have a 30% increased risk of a second primary breast cancer (Sandeep *et al.* 2006). The incidence of second primary cancers is also high even in participants with thyroid micro-carcinoma (Kim *et al.* 2013, Hsu *et al.* 2014). Because of the high incidence of thyroid cancer in women (Davies & Welch 2014, Siegel *et al.* 2015), it is important to understand why there is an increased risk of second primary breast cancers (Aschebrook-Kilfoy *et al.* 2013).

One possible cause for the increased rate of second primary breast cancers is a common genetic pathway. Cowden syndrome (CS) is an autosomal dominant disorder characterized by developing multiple primary cancers in various sites including thyroid and breast (Eng 1993). *PTEN* is one of the causative genes of CS or Cowden-like syndrome (CLS) (Liaw *et al.* 1997), and germline mutations are found in 25–85% of CS patients and <5% of CLS patients (Marsh *et al.* 1998, Tan *et al.* 2011). Germline mutations in other genes, such as succinate dehydrogenase genes (*SDHx*) (8%), *PIK3CA* (8.8%), *AKT1* (2.2%), and *SEC23B* (4%) are also known as CS/CLS susceptibility genes (Ni *et al.* 2008, Ngeow *et al.* 2011, Orloff *et al.* 2013). CS/CLS is one of the plausible causes of primary thyroid and breast cancers, and the standardized incidence rate ratio of second primary cancers in research participants with thyroid cancer is increased to 5.83 per 10 000 person a year (95% CI=3.01–10.18) by *PTEN* mutations (Ngeow *et al.* 2014). However, *PTEN* germline mutations are responsible for only a small portion of participants who present with non-syndromic primary thyroid and breast cancers (Pal *et al.* 2001, Ngeow *et al.* 2014). We hypothesize that germline mutations in genes other than *PTEN*, *SDHx*, *PIK3CA*, and *AKT* may explain the high rate of primary thyroid and breast cancer. In this study, we sought to identify possible germline mutations associated with thyroid and breast cancers using whole-exome sequencing.

Materials and methods

Samples

The study cohort consisted of 14 women known to have both primary breast and thyroid cancer malignancies. They were identified from a cohort of study participants who consented to participate in Cleveland Clinic IRB No.8458. These 14 research participants were selected under informed consent due to characteristics making

them high risk for a possible genetic association. Particularly, they all developed both primary cancers prior to 50 years of age; none had radiation exposure to the head and neck or chest prior to developing the cancers, and there was a high frequency of familial breast and thyroid cancers in the cohort. In addition, all participants had already been ruled out for a *PTEN* germline mutation (Ngeow *et al.* 2014). Genomic DNA was extracted from peripheral blood samples, and the quality and quantity of each DNA sample were examined by Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and 2200 Tape Station (Agilent Technologies, Santa Clara, CA, USA). This study is approved by the Institutional Review Board at the University of Chicago (No. 8962) and Cleveland Clinic (No. 8458).

Whole-exome sequencing

DNA libraries for whole-exome sequencing were prepared from 200–1000 ng genomic DNA using SureSelectXT Human All Exon V5 (Agilent Technologies). The prepared whole-exome libraries were quantified on the 2200 Tape Station (Agilent Technologies), and then sequenced by 150 bp paired-end reads on NextSeq 500 Desktop Sequencer (Illumina, San Diego, CA, USA). All procedures were performed according to the manufacturers' protocol.

Read mapping

After the exclusion of low-quality reads (base quality of <20 for more than 80% of bases) using FASTX toolkit, sequence reads were mapped to the human reference genome GRCh37/hg19 using Burrows–Wheeler Aligner (BWA) (v0.7.10) (Li 2014). After read pairs with a mapping quality of <30 and with mismatches more than 5% of read length were excluded, BAM files were generated using SAMTools (v1.1), and possible PCR duplicated reads were removed using Picard v1.91 (<http://broadinstitute.github.io/picard/>). As a population control, BAM files of 406 European (including Caucasian (CEU), British from England and Scotland (GBR), Iberian populations in Spain (IBS), and Tuscans in Italy (TSI), except for Finnish in Finland (FIN)) subjects, which were indicated to be healthy volunteers, were obtained from the 1000 Genomes Project database (www.1000genomes.org).

Variant calling

BCFtools (v1.1) was used to call variants (single nucleotide variants and indels) with the following criteria of i) sequence depth of ≥ 10 , ii) variant depth of ≥ 4 , and iii) base quality of

≥15 (Li *et al.* 2009). Variants in 14 cases and 406 controls were merged into a single .vcf file, and further filtering was performed to identify high-confidence variants, requiring that each variant i) have a call rate of ≥0.8; ii) found ≤1% in controls; and iii) found ≤1% in the American, African, and Asian populations in 1000 Genomes Project.

Variant annotation

Variants were annotated using the hg19 database in single nucleotide polymorphism (SNP) effect prediction tools (SnpEff) (Cingolani *et al.* 2012). The primary SnpEff genomic effects include splice site acceptor, splice site donor, indel frameshift, indel non-frameshift, non-sense, non-synonymous, and synonymous variants. For variants that have multiple different annotations, the highest impact effect was selected. The predicted impact of amino acid substitutions was annotated using five algorithms of latest release time (LRT) score, MutationTaster, PolyPhen-2 HumDiv, PolyPhen-2 HumVar, and scale invariant feature transform (SIFT). Variants predicted as ‘deleterious’ in LRT, ‘disease causing automatic’ and ‘disease causing’ in MutationTaster, ‘probably damaging’ and ‘possibly damaging’ in PolyPhen-2 HumDiv and PolyPhen-2 HumVar, and ‘damaging’ in SIFT were considered as ‘deleterious’s.

Sanger sequencing

Target regions were amplified by PCR from genomic DNA (Supplementary Table 1, see section on supplementary data given at the end of this article). The amplified products were sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit and 3500 Series Genetic Analyzer (Life Technologies) according to the manufacturer’s protocols. Variants not validated by Sanger sequencing were excluded from the following analysis.

Association analysis

Gene-based association analysis was performed using a burden test with 100 000-time permutation in PLINK/SEQ v0.10. In case of multiple results due to alternative transcripts, we selected the result with the smaller *P* value. *P* values of <0.05 were considered as statistically significant. To adjust multiple testing by the strict Bonferroni correction, we used significance level of $P < 3.65 \times 10^{-6}$ (0.05/13 705) in the gene-based association analysis.

Quantitative real-time PCR

Total RNAs from 18 cell lines previously extracted (Park *et al.* 2010) were reverse transcribed to cDNA

using SuperScript III (LifeTechnologies). Real-time PCR using TaqMan gene expression assay for *PARP4* (Hs00173105_m1) and *GAPDH* (Hs02758991_g1) was conducted on ViiA7 instrument (Life Technologies). mRNA levels of *PARP4* were normalized to *GAPDH* expression.

siRNA transfection and cell proliferation assay

HCC1143 cell line authenticated by short tandem repeat analysis was cultured under the recommendations of their respective depositors. siRNA oligonucleotides for targeting *PARP4* transcripts were purchased from Sigma–Aldrich. siNegative control (Cosmo Bio, Tokyo, Japan) was used as a control siRNA. siRNAs were transfected into cells with Lipofectamine RNAi max (Life Technologies). After 48-h incubation, cell viability assay was performed with the Cell Counting Kit-8 (Dojindo, Kumamoto, Japan).

Statistical analysis

Expression analysis and cell viability experiment were repeated three times at triplicates and evaluated using the Student’s *t* test. The prognostic value of *PARP4* was analyzed by Kaplan–Meier plotter (<http://kmplot.com/>), which includes gene expression data of 4142 breast cancer cases from GEO, EGA, and TCGA using Affymetrix HG-U133A, HG-U133 Plus 2.0, and HG-U133A 2.0 (Gyorffy *et al.* 2010, 2013).

Results

The median age at diagnosis for thyroid and breast cancer was 38 (range: 23–49) and 46 (range: 36–49) years old respectively (Table 1). Five participants were diagnosed with both thyroid and breast cancer at the same age. Six cases were diagnosed to have thyroid cancer before breast cancer diagnosis. In total, 13 thyroid cancers were either follicular or papillary subtype, and 11 breast cancers were a ductal subtype (Table 1). Seven (50%) and five (36%) participants had a family history of thyroid or breast cancer within three generations of the proband respectively, although no additional family members had primary thyroid and breast cancers.

In exome sequencing of the 14 cases, we obtained a total sequencing output with an average read depth of 138× per base. We also analyzed 406 European white controls obtained from the 1000 Genomes Project database using the same algorithm to minimize the calling discordance due to different algorithms. We identified a

Table 1 Demographics of the 14 participants with double primary thyroid and breast cancers

ID no.	Race	Cancer type	Age of diagnosis	Family history of thyroid and breast cancer	Histology	Metastasis
1	White	Thyroid Breast	36 36		Papillary Medullary	Unknown Unknown
2	White	Thyroid Breast	38 48		Papillary Ductal	Unknown Unknown
3	Hispanic	Thyroid Breast	40 40	Mother (63 years) and aunt (maternal, 60s)	Follicular Unknown	Unknown Unknown
4	American Indian	Thyroid Breast	25 43	Aunt (paternal, 50s), cousin (maternal, 60s), and cousin (maternal, 40s)	Follicular Ductal	No No
5	White	Thyroid Breast	47 47		Follicular/papillary Ductal	No No
6	White	Thyroid	25	Daughter (30 years) and sister (45 years)	Follicular/papillary	No
7	White	Breast Thyroid	46 46		Ductal Papillary	No No
8	White	Breast Thyroid	46 37	Uncle (maternal, 51 years)	Ductal Papillary	Lymph node Unknown
9	Unknown	Breast Thyroid	38 48	Aunt (maternal, 49 years)	Mucinous Papillary	Unknown Lymph node
10	White	Breast Thyroid	46 35	Aunt (paternal, 56 years)	Ductal Papillary	No Lymph node
11	Black	Thyroid	41	Sister (19 years) and sister (38 years)	Papillary	Lymph node
12	White	Breast Thyroid	41 NA	Grandmother (paternal, 84 years/ maternal, 45 years) and aunt (paternal, 54 years/maternal, 65 years)	Ductal Hurthle cell	Unknown Unknown
13	American Indian	Thyroid	NA	Aunt (maternal, 50s) and aunt (maternal, 50s)	Ductal	No
14	American Indian	Breast Thyroid	NA 23	Uncle (maternal, 72 years) and cousin (paternal, 28 years)	Papillary	No
		Breast	49	Aunt (maternal; 65 years)	Ductal	No
		Thyroid	49	Aunt (maternal, 53 years)	Papillary	No
		Breast	49		Ductal	No
		Thyroid	48	Uncle (maternal, 40s)	Follicular/papillary	Unknown
		Breast	44		Ductal	No

NA, not available.

total of 105 064 and 345 793 variants in our cases and the 1000 Genomes controls respectively (Table 2). Of these variants, 4.8 and 18.8% were rare (defined as having a minor allele frequency (MAF) of <1% in the controls in both cases and controls). After excluding common SNPs with MAF of $\geq 1\%$ in the controls, 5073 variants in 2909 genes in cases and 65 170 variants in 14 232 genes in controls were identified within the coding regions (Table 2). In the 14 participants, we inspected rare variants known to be responsible for hereditary cancer syndromes such as *SDHx* and *KLLN* (for CS), and *BRCA1/2* (for hereditary breast and ovarian cancer syndrome) (Supplementary Figure 1); however, no participant had more than one of these rare variants identified. In addition, no statistically significant enrichment of variants was observed

in the recurrently somatically mutated genes (known to occur in more than 10% of cancers) that have been previously reported in epithelial thyroid cancer (such as *BRAF*, *RET*, and *TERT*) and breast cancers (such as *PIK3CA*, *TP53*, and *CDH1*) (Supplementary Figure 2).

Using five different prediction tools, variants predicted as deleterious substitutions in at least one algorithm were identified in 2316 and 13 554 genes in cases and controls respectively (Table 2). Gene-based association analysis of these deleterious variants identified 34 genes, in which rare variants were enriched with a *P* value of $<1.0 \times 10^{-3}$. However, none of these genes showed a genome-wide significant level of association when we considered the Bonferroni correction for multiple testing ($P < 3.65 \times 10^{-6}$) (Supplementary Figure 3). Since several

Table 2 Summary of number of cases, variants, and genes with rare variants

	Subjects	Variants	Rare variants	Genes with rare variants	Genes with variants predicted as deleterious
Cases	14	105 064	5073	2909	2316
Controls ^a	406	345 793	65 170	14 232	13 554

^aControls are 406 European individuals in the 1000 Genomes Project database.

previous reports indicated that the mutations in the genes involved in DNA repair pathways are genetic risk factors for several types of familial cancer, we focused on the DNA repair pathway and found that 13 of the 14 participants had rare variants in at least one DNA repair-related gene (Fig. 1 and Supplementary Table 2). Among them, *PARP4* showed the most significant association with the risk for primary thyroid and breast cancers (odds ratio=5.2; $P=1.0\times 10^{-5}$; Fig. 1). In *PARP4*, two different variants, G496V and T1170I, were found in six of the 14 (43%) participants, and both of these were validated by Sanger sequencing (Fig. 2a). The majority of these were T1170I substitutions found in five of the 14 (36%) cases including three White, one American Indian, and one unknown race

case. In contrast, this variant was found in only two of the 406 (0.5%) control individuals. Interestingly, among the 17 mutations found in 960 breast cancer cases in TCGA database, seven cases (41%) showed the two-hit mutation: a somatic mutation and loss of one *PARP4* allele. Furthermore, one of these 17 somatic mutations, p.T1170I, is identical to one of germline mutations we detected (Supplementary Table 3). G496V substitution was uniquely found in the remaining one American Indian case and not found in any of the 406 controls (Fig. 2b).

Next, we examined whether *PARP4* shows tumor-suppressive function (Fig. 3). mRNA expression of *PARP4* in 18 breast cancer cell lines was examined by qPCR. Since HCC1143 cells showed the highest expression of *PARP4*

Endocrine-Related Cancer

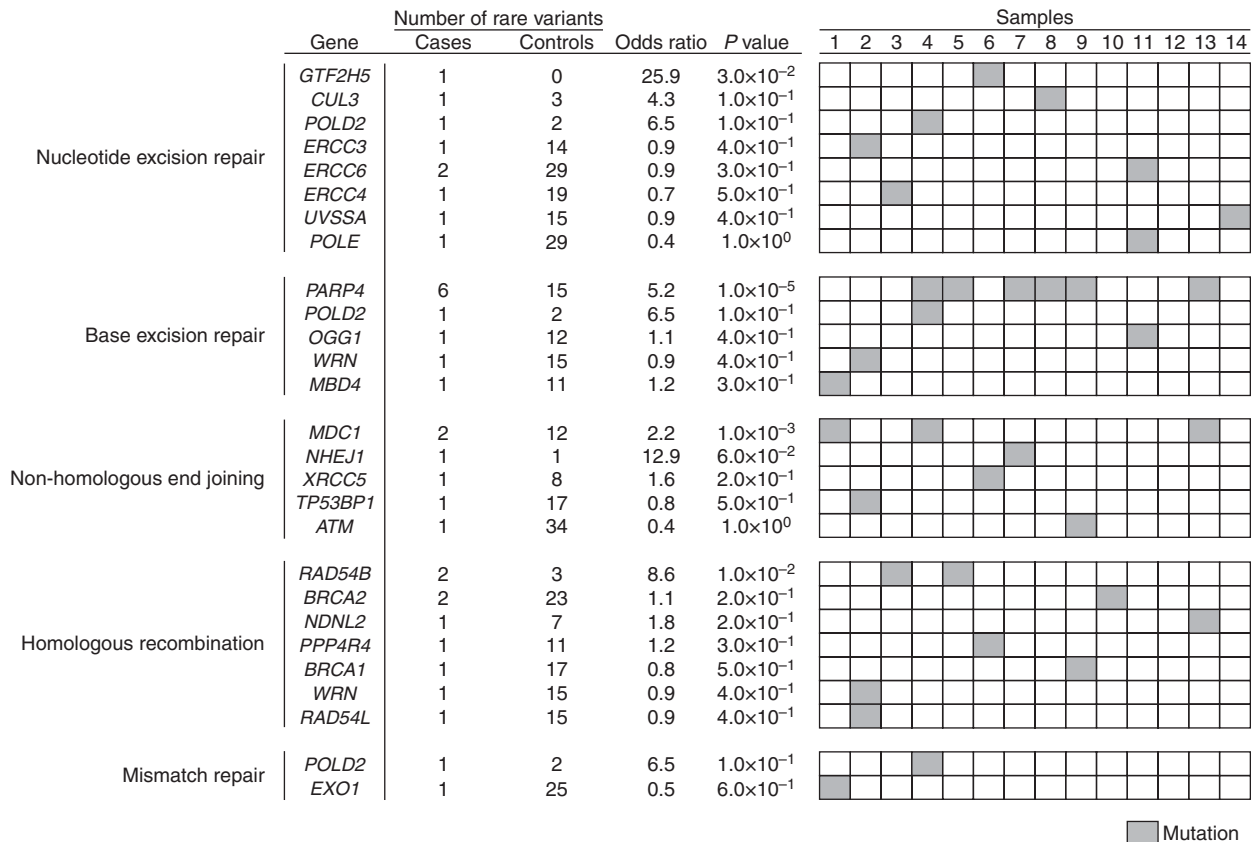
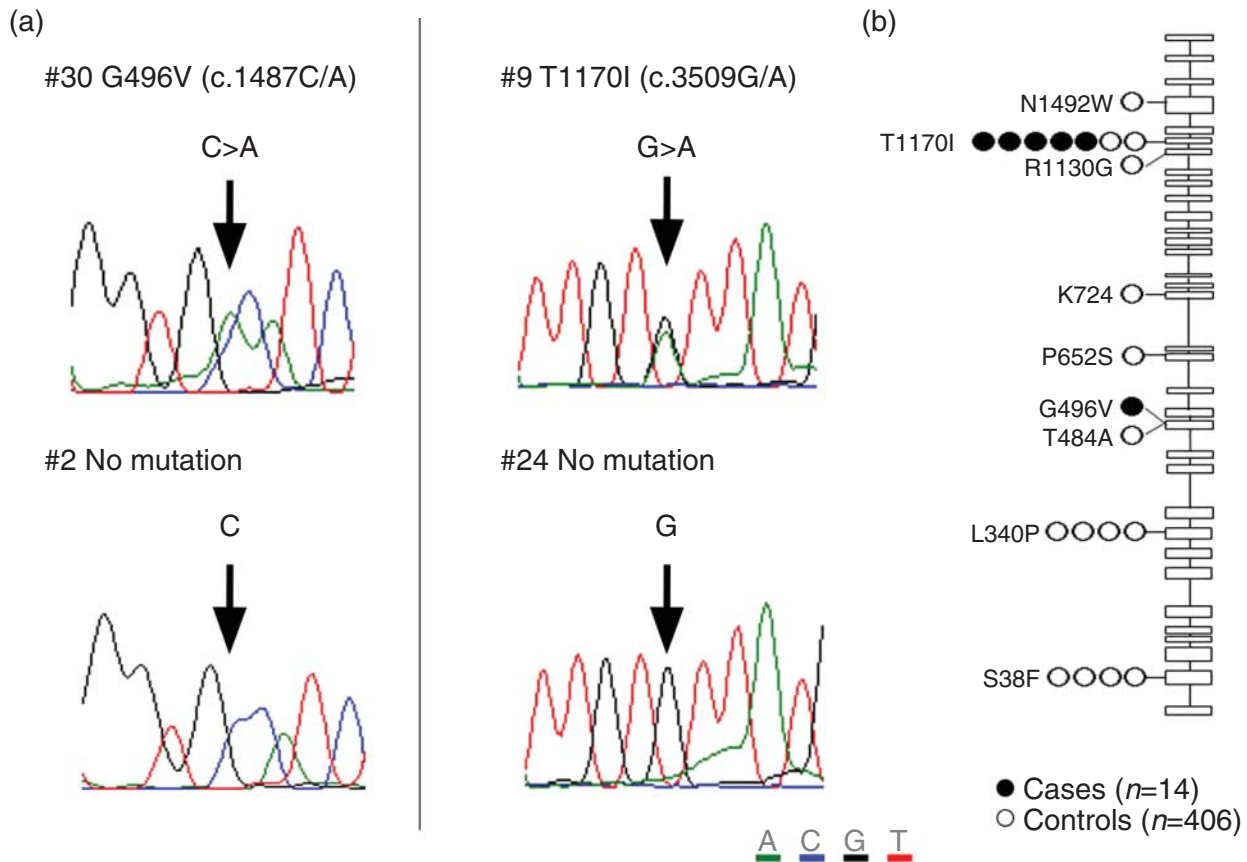


Figure 1

Association of the burden of rare variants in DNA repair-related genes with risk for primary thyroid and breast cancer.

**Figure 2**

Mutations in *PARP4* gene in 14 participants with synchronous thyroid and breast cancers and 406 control individuals. (a) Validation of two variants, G496V and T1170I, by Sanger sequencing. (b) Location of each mutation in

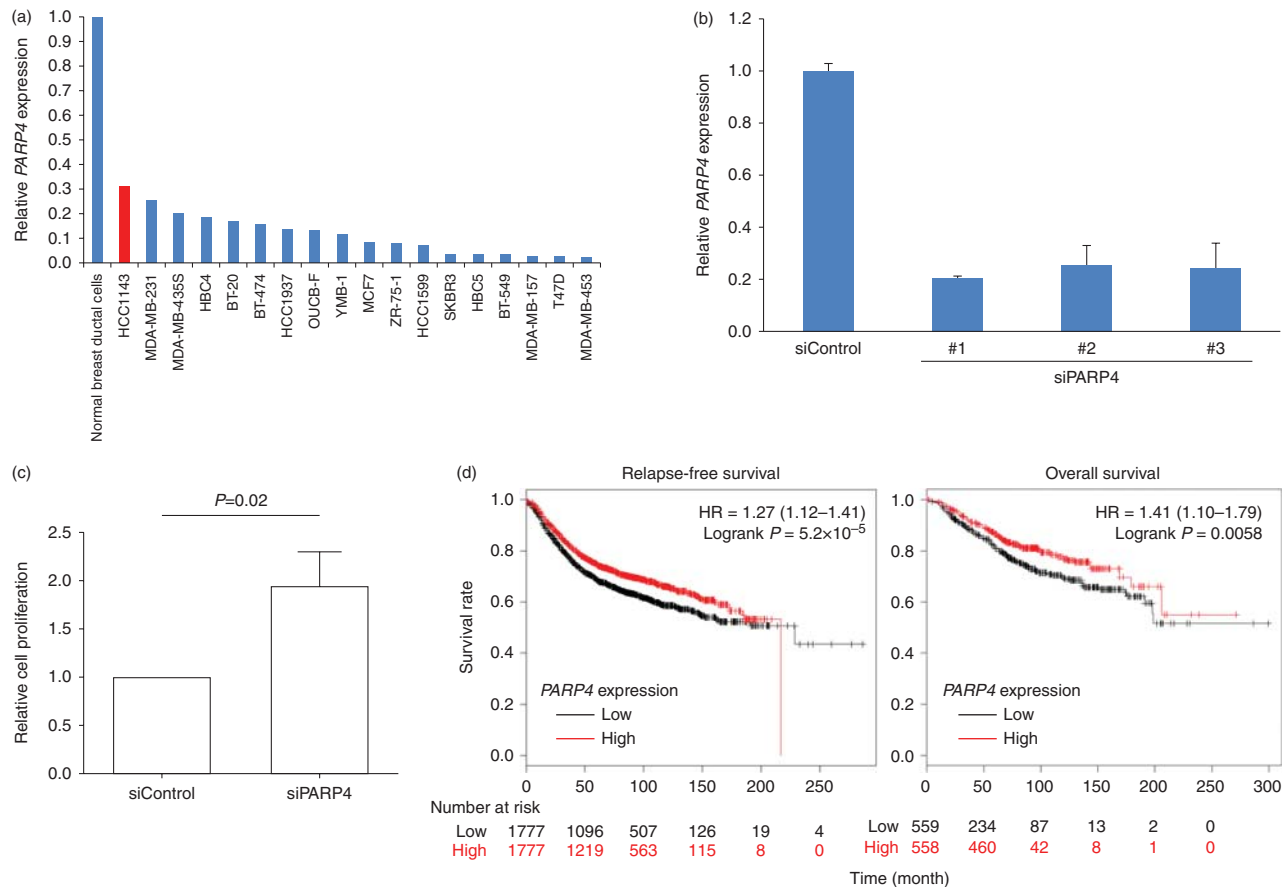
PARP4. Closed and open circles represent the location of *PARP4* mutations found in cases and controls respectively.

among the 18 cell lines tested, we used HCC1143 cells for the further experiments. When *PARP4* expression was knocked down by siRNA, the proliferation of HCC1143 cells was significantly enhanced threefold compared to the cells transfected with siControl ($P=0.02$; Fig. 3c), suggesting that *PARP4* works as a tumor suppressor.

Discussion

In this study, we comprehensively analyzed exome variants in 14 participants with primary thyroid and breast cancer, and identified *PARP4* as a possible susceptibility gene candidate for primary thyroid and breast cancer. *PARP4*, also known as VPARP, is a family member of the poly (ADP-ribose) polymerase (PARP) (Kickhoefer et al. 1999). The PARP protein superfamily has 17 members and controls a wide array of cellular processes such as DNA repair, transcriptional regulation, and RNA interference (Gibson & Kraus 2012). So far, the

function of a few PARPs including PARP1, PARP2, and PARP5 have been well-characterized (Gibson & Kraus 2012). PARP1 and PARP4 contain BRCA1 carboxy-terminal (BRCT) domain repeats, which are thought to bind phosphorylated DNA damage-sensing proteins recruiting PARPs to sites of DNA damage (Manke et al. 2003). Although little is known about the biological function of *PARP4*, it is suspected to be involved in the DNA repair pathway due to its BRCT domain. Many hereditary diseases responsible for synchronous cancers such as Lynch syndrome (*MLH1*, *MSH2*, *MSH6*, and *PMS2* are responsible genes) and hereditary breast and ovarian cancer (*BRCA1* and *BRCA2*) are known to be caused by germline mutations in genes involved in the DNA repair pathway (Lynch et al. 2009). Moreover, several DNA repair genes are reported to be mutated to a significant degree in both aggressive papillary thyroid carcinoma and breast cancers (Cancer Genome Atlas Network 2012, Cancer Genome Atlas Research Network 2014). As shown in

**Figure 3**

Functional analysis of *PARP4*. (a) *PARP4* mRNA expression in 18 breast cancer cell lines. (b) Knockdown of *PARP4* expression by *PARP4*-specific siRNAs in HCC1143 breast cancer cells. (c) Effects of *PARP4* knockdown on the proliferation of HCC1143 cells. (d) Kaplan–Meier analysis of

relapse-free survival and overall survival in 4142 patients with breast cancer patients (from GEO, EGA, and TCGA datasets) according to the expression of *PARP4*. High and low expressions were defined as above and below median respectively; HR, Hazard ratio.

Fig. 3, we suggest that *PARP4* might have a tumor suppressor function. In addition, Kaplan–Meier analysis using gene expression data from 4142 cases of breast cancer which was available on Kaplan–Meier plotter (<http://kmpplot.com/>) showed worse relapse-free survival ($P < 0.001$, Hazard ratio 1.27) and overall survival ($P = 0.006$, Hazard ratio 1.41) in the *PARP4* low-expression group. These lines of evidence support the possibility that *PARP4* plays a critical role in thyroid and breast tumorigenesis.

Radiation therapy is known to increase the risk of second primary cancers (Tucker *et al.* 1991). Generally, a radiation-induced second primary cancer arising at the radiation-associated site typically arises 10 years after the initial radiation exposure (Grantzau & Overgaard 2015). However, since both thyroid and breast cancers were diagnosed prior to any therapeutic radiation exposure, including radioactive iodine (RAI) our cohort does not

reflect a link between radiation-induced secondary primary cancer risk and *PARP4* germline mutations.

In conclusion, through the whole-exome sequencing approach, we have implicated *PARP4* germline mutations as possible susceptibility factors for the risk of synchronous thyroid and breast cancers. It is obvious that our study has a limitation in its sample size; thus, further analysis is warranted with a larger number of samples to verify the results obtained in this study. Although we show evidence of the tumor suppressor function of *PARP4* in breast cancer cells, further analysis is warranted to clarify the development of primary thyroid and breast cancer. Regardless, the high rate of the rare variant of *PARP4* found in our study when compared to the very low rate in the controls is compelling evidence of a possible new genetic syndrome that is responsible for both primary thyroid and breast cancer.

Supplementary data

This is linked to the online version of the paper at <http://dx.doi.org/10.1530/ERC-15-0359>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

Author R H Grogan was supported by Award Number K12CA139160 from the National Cancer Institute. Research samples were accrued, in part, under the Doris Duke Distinguished Clinical Scientist Award and P01124570 from the National Cancer Institute (both to C Eng). The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Cancer Institute or the National Institute of Health. C Eng is the Sondra J and Stephen R Hardis Endowed Chair of Cancer Genomic Medicine at the Cleveland Clinic, and an ACS Clinical Research Professor.

Acknowledgements

Supercomputing resources were provided by the Human Genome Center, the Institute of Medical Science, and the University of Tokyo (<http://sc.hgc.jp/shirokane.html>). Additional thanks to Victoria Raymond and Hannah Pearson for assistance with medical records review.

References

- Aschebrook-Kilfoy B, Schechter RB, Shih YC, Kaplan EL, Chiu BC, Angelos P & Grogan RH 2013 The clinical and economic burden of a sustained increase in thyroid cancer incidence. *Cancer Epidemiology, Biomarkers & Prevention* **22** 1252–1259. (doi:10.1158/1055-9965.EPI-13-0242)
- Cancer Genome Atlas Network 2012 Comprehensive molecular portraits of human breast tumours. *Nature* **490** 61–70. (doi:10.1038/nature11412)
- Cancer Genome Atlas Research Network 2014 Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159** 676–690. (doi:10.1016/j.cell.2014.09.050)
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X & Ruden DM 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6** 80–92. (doi:10.4161/fly.19695)
- Davies L & Welch HG 2014 Current thyroid cancer trends in the United States. *JAMA Otolaryngology – Head & Neck Surgery* **140** 317–322. (doi:10.1001/jamaoto.2014.1)
- Eng C 1993 PTEN Hamartoma Tumor Syndrome (PHTS). In *GeneReviews(R)*. Eds RA Pagon, MP Adam, HH Ardinger, SE Wallace, A Amemiya, LJH Bean, TD Bird, CT Fong, HC Mefford, RJH Smith, et al. Seattle, WA, USA: University of Washington. (available at: <http://www.ncbi.nlm.nih.gov/books/NBK1488/>)
- Gibson BA & Kraus WL 2012 New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nature Reviews. Molecular Cell Biology* **13** 411–424. (doi:10.1038/nrm3376)
- Grantzau T & Overgaard J 2015 Risk of second non-breast cancer after radiotherapy for breast cancer: a systematic review and meta-analysis of 762,468 patients. *Radiotherapy and Oncology* **114** 56–65. (doi:10.1016/j.radonc.2014.10.004)
- Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q & Szallasi Z 2010 An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment* **123** 725–731. (doi:10.1007/s10549-009-0674-9)
- Gyorffy B, Surowiak P, Budczies J & Lanczky A 2013 Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **8** e82241. (doi:10.1371/journal.pone.0082241)
- Hsu CH, Huang CL, Hsu YH, Iqbal U, Nguyen PA & Jian WS 2014 Co-occurrence of second primary malignancy in patients with thyroid cancer. *QJM* **107** 643–648. (doi:10.1093/qjmed/hcu051)
- Kickhoefer VA, Siva AC, Kedersha NL, Inman EM, Ruland C, Streuli M & Rome LH 1999 The 193-kD vault protein, VPARP, is a novel poly (ADP-ribose) polymerase. *Journal of Cell Biology* **146** 917–928. (doi:10.1083/jcb.146.5.917)
- Kim C, Bi X, Pan D, Chen Y, Carling T, Ma S, Udelsman R & Zhang Y 2013 The risk of second cancers after diagnosis of primary thyroid cancer is elevated in thyroid microcarcinomas. *Thyroid* **23** 575–582. (doi:10.1089/thy.2011.0406)
- Li H 2014 Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30** 2843–2851. (doi:10.1093/bioinformatics/btu356)
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & 1000 Genome Project Data Processing Subgroup 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25** 2078–2079. (doi:10.1093/bioinformatics/btp352)
- Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M et al. 1997 Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nature Genetics* **16** 64–67. (doi:10.1038/ng0597-64)
- Lynch HT, Casey MJ, Snyder CL, Bewtra C, Lynch JF, Butts M & Godwin AK 2009 Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management. *Molecular Oncology* **3** 97–137. (doi:10.1016/j.molonc.2009.02.004)
- Manke IA, Lowery DM, Nguyen A & Yaffe MB 2003 BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* **302** 636–639. (doi:10.1126/science.1088877)
- Marsh DJ, Coulon V, Lunetta KL, Rocca-Serra P, Dahia PL, Zheng Z, Liaw D, Caron S, Duboue B, Lin AY et al. 1998 Mutation spectrum and genotype–phenotype analyses in Cowden disease and Bannayan–Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. *Human Molecular Genetics* **7** 507–515. (doi:10.1093/hmg/7.3.507)
- Ngeow J, Mester J, Rybicki LA, Ni Y, Milas M & Eng C 2011 Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-like syndrome characterized by germline PTEN, SDH, or KLLN alterations. *Journal of Clinical Endocrinology and Metabolism* **96** E2063–E2071. (doi:10.1210/jc.2011-1616)
- Ngeow J, Stanuch K, Mester JL, Barnholtz-Sloan JS & Eng C 2014 Second malignant neoplasms in patients with Cowden syndrome with underlying germline PTEN mutations. *Journal of Clinical Oncology* **32** 1818–1824. (doi:10.1200/JCO.2013.53.6656)
- Ni Y, Zbuk KM, Sadler T, Patocs A, Lobo G, Edelman E, Platzer P, Orloff MS, Waite KA & Eng C 2008 Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *American Journal of Human Genetics* **83** 261–268. (doi:10.1016/j.ajhg.2008.07.011)
- Orloff MS, He X, Peterson C, Chen F, Chen JL, Mester JL & Eng C 2013 Germline PIK3CA and AKT1 mutations in Cowden and Cowden-like syndromes. *American Journal of Human Genetics* **92** 76–80. (doi:10.1016/j.ajhg.2012.10.021)
- Pal T, Hamel N, Vesprini D, Sanders K, Mitchell M, Quercia N, Ng Cheong N, Murray A, Foulkes W & Narod SA 2001 Double primary cancers of the

- breast and thyroid in women: molecular analysis and genetic implications. *Familial Cancer* **1** 17–24. (doi:10.1023/A:1011541424424)
- Park JH, Nishidate T, Kijima K, Ohashi T, Takegawa K, Fujikane T, Hirata K, Nakamura Y & Katagiri T 2010 Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer Research* **70** 2759–2769. (doi:10.1158/0008-5472.CAN-09-3911)
- Sandeep TC, Strachan MW, Reynolds RM, Brewster DH, Scelo G, Pukkala E, Hemminki K, Anderson A, Tracey E, Friis S *et al.* 2006 Second primary cancers in thyroid cancer patients: a multinational record linkage study. *Journal of Clinical Endocrinology and Metabolism* **91** 1819–1825. (doi:10.1210/jc.2005-2009)
- Siegel RL, Miller KD & Jemal A 2015 Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians* **65** 5–29. (doi:10.3322/caac.21254)
- Tan MH, Mester J, Peterson C, Yang Y, Chen JL, Rybicki LA, Milas K, Pederson H, Remzi B, Orloff MS *et al.* 2011 A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *American Journal of Human Genetics* **88** 42–56. (doi:10.1016/j.ajhg.2010.11.013)
- Tucker MA, Jones PH, Boice JD Jr, Robison LL, Stone BJ, Stovall M, Jenkin RD, Lubin JH, Baum ES, Siegel SE *et al.* 1991 Therapeutic radiation at a young age is linked to secondary thyroid cancer. The Late Effects Study Group. *Cancer Research* **51** 2885–2888.

Received in final form 18 November 2015

Accepted 23 December 2015

Made available online as an Accepted Preprint

23 December 2015