Genotype analyses in the Japanese and Belarusian populations reveal independent effects of rs965513 and rs1867277 but do not support the role of *FOXE1* polyalanine tract length in conferring risk for papillary thyroid carcinoma

Alyaksandr V. Nikitski[1], Tatiana I. Rogounovitch[1], Andrey Bychkov[1*], Meiko Takahashi[2], Koh-ichiro Yoshiura[3], Norisato Mitsutake[1,4], Takahisa Kawaguchi[5], Michiko Matsuse[1], Valentina M. Drozd[6], Yuri Demidchik[7], Eijun Nishihara[8], Mitsuyoshi Hirokawa[8], Akira Miyauchi[8], Alexander V. Rubanovich[9,10], Fumihiko Matsuda[5], Shunichi Yamashita[1,10], Vladimir A. Saenko[10]

[1] Department of Radiation Medical Sciences, Atomic Bomb Disease Institute, Nagasaki University, Nagasaki, Japan

[2] Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Japan

[3] Department of Human Genetics, Atomic Bomb Disease Institute, Nagasaki University, Nagasaki, Japan

[4] Nagasaki University Research Center for Genomic Instability and Carcinogenesis, Nagasaki 852-8523, Japan

[5] Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan

[6] Department of Endocrinology, Belarusian Academy for Postgraduate Education, Minsk, Belarus

[7] Department of Oncology, Belarusian Academy for Postgraduate Education, Minsk, Belarus

[8] Kuma Hospital, Kobe, Japan

[9] Ecological Genetics Laboratory, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

[10] Department of Radiation Molecular Epidemiology, Nagasaki University, Nagasaki, Japan

*Present address: Department of Pathology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

Alyaksandr V. Nikitski, M.D., Ph.D.

Department of Radiation Medical Sciences, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7116, Fax: +81-95-819-7117, E-mail: nikitski.alyaksandr@gmail.com

Tatiana Rogounovitch, M.D., Ph.D.

Department of Global Health, Medicine and Welfare, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7170, Fax: +81-95-819-7172, E-mail: tatiana@nagasaki-u.ac.jp

Andrey Bychkov, M.D., Ph.D.

Department of Radiation Medical Sciences, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7116, Fax: +81-95-819-7117, E-mail: andrey.bychkov@live.com

Meiko Takahashi, Ph.D.

Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, 53 Shogoin Kawaharacho, Sakyo-Ku, Kyoto 606-8507, Japan

Tel.: +81-75-366-7404, Fax: +81-75-751-4167, E-mail: meiko@genome.med.kyoto-u.ac.jp

Koh-ichiro Yoshiura, M.D., Ph.D.

Department of Human Genetics, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan, Japan

Tel.: +81-95-819-7118, Fax: +81-95-819-7121, E-mail: kyoshi@nagasaki-u.ac.jp

Norisato Mitsutake, M.D., Ph.D.

Department of Radiation Medical Sciences

Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Nagasaki University Research Center for Genomic Instability and Carcinogenesis, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7116, Fax: +81-95-819-7117, E-mail: mitsu@nagasaki-u.ac.jp

Takahisa Kawaguchi, Ph.D.

Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Kyoto 606-8501, Japan

Tel.: +81-(0)75-751-4157, E-mail: tkawa@genome.med.kyoto-u.ac.jp

Michiko Matsuse, Ph.D.

Department of Radiation Medical Sciences, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7116, Fax: +81-95-819-7117, E-mail: michikom@nagasaki-u.ac.jp

Valentina M. Drozd, M.D., Ph.D.

Thyroid Disease Research Group, Belarusian Academy for Postgraduate Education, 3-3 P.Brovki str., Minsk 220013, Belarus

Tel.: +375-17-290-9838; Fax: +375-17-292-2533; E-mail: vm.drozd@gmail.com

Yuri Demidchik, M.D., Ph.D.

Department of Oncology, Belarusian Academy for Postgraduate Education, P.Lesnoi, Minsky reg. 223040, Belarus

Tel.: +375-17-287-9598; Fax: +375-17-265-3552; E-mail: yu.demidchik@gmail.com

Eijun Nishihara, M.D., Ph.D.

Kuma Hospital, 8-2-35 Shimoyamate-dori, Chuo-ku, Kobe 650-0011, Japan

Tel.: +81-78-371-3721, Fax: +81-78-371-3645, E-mail: nishihara@kuma-h.or.jp

Mitsuyoshi Hirokawa, M.D., Ph.D.

Department of Diagnostic Pathology, Kuma Hospital, 8-2-35 Shimoyamate-dori, Chuo-ku, Kobe 650-0011, Japan

Tel.: +81-78-371-3721, Fax: +81-78-371-3645, E-mail: mhirokawa@kuma-h.or.jp

Akira Miyauchi, M.D., Ph.D.

Department of Surgery, Kuma Hospital, 8-2-35 Shimoyamate-dori, Chuo-ku, Kobe 650-0011, Japan

Tel.: +81-78-371-3721, Fax: +81-78-371-3645, E-mail: miyauchi@kuma-h.or.jp

Alexander V. Rubanovich, Ph.D.

Ecological Genetics Laboratory, Vavilov Institute of General Genetics, Russian Academy of Sciences, 3 Gubkin Str., Moscow 119991, Russia

Tel.: +7-499-123-8958; Fax: +7-499-132-8962; E-mail: rubanovich@vigg.ru

Department of Health Risk Control, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7122, Fax: +81-95-819-7169

Fumihiko Matsuda, Ph.D.

Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Kyoto 606-8501, Japan

Tel.: +81-75-819-7163, E-mail: fumi@genome.med.kyoto-u.ac.jp

Shunichi Yamashita, M.D., Ph.D.

Department of Radiation Medical Sciences, Department of Radiation Molecular Epidemiology, Atomic Bomb Disease Institute, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7116, Fax: +81-95-819-7117, E-mail: shun@nagasaki-u.ac.jp

Vladimir Saenko, Ph.D.

Department of Radiation Molecular Epidemiology, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7122, Fax: +81-95-819-7169, E-mail: saenko@nagasaki-u.ac.jp

Running title: Effects of *FOXE1* polymorphisms on susceptibility to PTC

Key words: papillary thyroid carcinoma, *FOXE1* polymorphism, genotype, case-control association study, functional analysis of transactivation potential

Abstract

Background: Several functional SNPs at the *FOXE1* locus on chromosome 9q22.33 have been associated with the risk for papillary thyroid carcinoma (PTC). This study set out to elucidate whether their effects are independent, using genotyping results in populations of Asian and European descent.

Methods: Single-nucleotide polymorphisms (SNPs) rs965513 and rs1867277, and a polymorphic region determining the length of FOXE1 polyalanine (poly-Ala) tract were genotyped in 501 patients with PTC and 748 healthy individuals from Japan, and in 660 patients and 820 population controls from Belarus. Functional analysis of transactivation activities of FOXE1 isoforms with varying number of alanine repeats was performed by a dual luciferase assay.

Results: All three polymorphisms were significantly associated with PTC in both populations on univariate analysis. However, conditional analysis revealed independent effects of rs965513 and rs1867277 SNPs, but not of the *FOXE1* poly-Ala polymorphism. The independent effect of the lead rs965513 SNP was observed in both populations, while that of rs1867277 was only identified in the Japanese population, in which linkage disequilibrium between the three polymorphisms is markedly weaker. Despite the strong decrease in transcriptional activity with increasing FOXE1 poly-Ala tract length, no difference in transactivation potential of the FOXE1 poly-Ala isoforms could be seen after adjustment for the minimal promoter activity in the reporter vectors. Plasmids encoding FOXE1 isoforms of increasing poly-Ala tract length were also found to produce less FOXE1 protein after cell transfection.

Conclusions: The functional variants, rs965513 and rs1867277, independently contribute to genetic predisposition to PTC, while a contributing role of the *FOXE1* poly-Ala polymorphism could not be confirmed.

Introduction

There has been extensive progress in the identification of genetic variants affecting susceptibility to differentiated thyroid cancer in humans in recent years. In particular, genome-wide and target gene association studies have identified single-nucleotide polymorphisms (SNPs) on chromosome 9q22.33 with the risk for thyroid cancer, primarily for sporadic and familial papillary thyroid carcinoma (PTC) in non-irradiated or radiation-exposed individuals, across different populations and ethnicities (1-14); these associations have also been confirmed in meta-analyses (15-18). The closest gene in this chromosomal region is *FOXE1* (Forkhead box E1, also known as Thyroid transcription factor 2 (*TTF2I*), gene ID: 2304), an intronless gene encoding a member of the forkhead/winged helix family of evolutionarily conserved transcription factors (19). FOXE1 plays an essential role in thyrocyte precursor migration, thyroid organogenesis and differentiation (20-22).

Localized about 60 kb upstream and centromeric to *FOXE1*, rs966513 was the first SNP reported as a genetic determinant of susceptibility to thyroid cancer in a genome-wide association study (1), but its functional relevance was established only recently (23). The lead rs966513, as well as several other SNPs on 9q22.33 that are in linkage disequilibrium (LD) with rs966513, were shown to modify the activities of long-range enhancers involved in the transcriptional regulation of *FOXE1* and *PTCSC2* (papillary thyroid carcinoma susceptibility candidate 2, gene ID: 101928337), a newly discovered thyroid-specific long intergenic noncoding RNA gene whose chromosomal position partly overlaps with that of the *FOXE1* promoter (24).

Another functional variant, rs1867277, located in the *FOXE1* 5'-UTR (c.-283) has been found to confer risk to differentiated thyroid cancer in the large-scale target gene association study in individuals from Spain and Italy (2). The finding was reproduced by other groups in

different populations (5, 6, 9, 13, 14, 25, 26) and confirmed by meta-analyses [15, 16]. This variant is also involved in the regulation of *FOXE1* expression through differential recruitment of USF1/USF2 transcription factors.

In the coding region, *FOXE1* possesses a multinucleotide polymorphism, which consists of a variable number of trinucleotides (most commonly GCC, less frequently GCT or GCA, all encoding alanine) ranging from 11 to 22 repeats, hereby referred to as the *FOXE1* poly-Ala polymorphism. The most common alleles encode 14 and 16 alanine residues. Polyalanine tracts are a frequent feature of conserved transcription factors, and have been implicated in a number of congenital malformation syndromes (27, 28 for review). Variation in the *FOXE1* polyalanine tract length has been associated with susceptibility to thyroid dysgenesis (29-31) and, more recently, with thyroid cancer. The poly-Ala14 has been shown to be protective, and poly-Ala16 a risk-conferring allele (5, 12, 14, 25, 26). These observations were confirmed by meta-analysis (16). The transcriptional activity of FOXE1 poly-Ala16 was found to be diminished as compared to that of poly-Ala14 (26), although the difference in transactivation potential was not observed between poly-Ala14 and shorter isoforms in an earlier study (29).

To the best of our knowledge, studies on *FOXE1* poly-Ala polymorphism in thyroid cancer have not been performed in individuals of Asian origin. The only information on this genetic variant in corresponding populations is available from a study of 46 Japanese patients with thyroid dysgenesis (29), and 110 cases of idiopatic premature ovarian failure and 110 controls from China (32). Our work, therefore, is the first to characterize the *FOXE1* poly-Ala in a large Japanese cohort.

The objective of our study was to determine whether the three *FOXE1* polymorphisms with functional roles may have independent effects on risk for thyroid cancer. For this purpose we analyzed genotypes of patients with PTC and population controls of diverse ethnic

backgrounds from Japan and Belarus, and performed functional analysis of transactivation activities of five isoforms of FOXE1 with different lengths of polyalanine tract in a normal human thyroid cell line and thyroid cancer cell lines.

Methods

Study populations

A total of 501 patients aged 13-87 years operated for PTC at Kuma hospital (Kobe, Japan) were enrolled (4). As population controls, 748 Japanese individuals aged 20-76 years at sampling were recruited in Nagasaki University. Participants from Belarus included 660 patients with PTC aged 2-22 years at diagnosis and 820 population controls aged 16-49 years at sampling (3). None of the Japanese individuals had a history of radiation exposure. All PTC patients from Belarus and 620 (75.6%) control individuals were exposed to radiation as a result of the Chernobyl accident. Informed consent was obtained from all individual participants included in the study. The protocol of the study was approved by the ethics committees of all participating institutions.

DNA extraction

In the Japanese cohort, DNA was extracted from formalin-fixed paraffin-embedded tissues of PTC patients (4), and from peripheral blood of control individuals using a QIAamp DNA mini kit (QIAGEN, Tokyo, Japan). In the Belarusian cohort, DNA was extracted from peripheral blood mononuclear cells of all participants with a Puregene kit (Qiagen, Germantown, MD, USA) (3). DNA concentration was measured with a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA), and samples were stored at -80ºC until use.

SNP genotyping and *FOXE1* poly-Ala tract length measurement

In both Japanese and Belarusian cohorts, rs965513 and rs1867277 genotyping was performed with pre-designed custom ABI TaqMan SNP assays (C_1593670_20 and C_11736668_10, respectively) as described before (13).

FOXE1 poly-Ala tract length was measured by resolving PCR products obtained by amplification with a 5'FAM-labeled forward primer and an unlabeled reverse primer flanking the region encoding the *FOXE1* poly-Ala polymorphism in an ABI Prism 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA) in GeneScan mode. Data were processed using GeneMapper version 3.7 software. Direct sequencing of PCR products obtained from 47 randomly chosen samples was used for fine adjustment of GeneScan data. An example of a chromatogram and the detailed protocol are presented in Supplementary Fig. 1 and the corresponding legend.

Cell cultures

The immortalized normal human thyroid cell line Nthy-ori 3-1 and the PTC cell line KTC-1 were grown in RPMI-1640 medium supplemented with 5% FBS and 1% penicillin-streptomycin. The human PTC cells line TPC1 and the follicular thyroid cancer WRO cell line were maintained in DMEM growth medium supplemented with 10% FBS, 1% nonessential amino acids, 1 mM sodium pyruvate and 1% penicillin-streptomycin. All cell lines were cultured in monolayers at 37°C in a humidified 5% CO_2 environment.

Expression vectors and reporter plasmids

PCR products of *FOXE1* with 12, 14, 15, 16 or 19 alanine repeats variants were obtained using corresponding genomic DNA as a template with the forward 5′-ACGCGTATGACTGCCGAGAGCGGGC-3′ and the reverse 5′-CTCGAGCATGGCGGACACGAACCGA-3′ primers (underlined are *Mlu*I and *Xho*I restriction sites). Amplicons were cloned into the pCR4 Blunt-TOPO vector (Invitrogen, Carlsbad, CA, USA), sequenced, and recloned into *Mlu*I/*Xho*I sites of pCMV6-AC-IRES-GFP expression vector (OriGene Technologies, Rockville, MD, USA).

Reporter plasmids were based on the pGL4.23[luc2/minP] luciferase vector (Promega, Madison, WI, USA), which originally contains a 32 bp minimal promoter that regulates the expression of the firefly *luc2* gene. (http://www.promega.jp/~/media/files/resources/protocols/product%20information%20sheets/a/pgl423%20vector.pdf; accessed October 2015).

To prepare the thyroperoxidase (*TPO*) promoter-driven reporter plasmid, a fragment of the *TPO* promoter was PCR-amplified as described before (33) with the following primers: forward 5’-actgGAGCTCGAGCTGCACCCAACCCAAT-3’ and reverse 5’-gcaaCTCGAGAGTAATTTTCACGGCTGT-3’ (underlined: *Sac*I and *Xho*I restriction sites; lower case: 4 bp extensions were added at the 5’-ends to ensure effective endonuclease

digestion), treated with the appropriate enzymes (NEB, Ipswich, MA, USA), and ligated into pGL4.23[luc2/minP] upstream of the minimal promoter.

To prepare the FOXE1 response element (FRE)-driven reporter plasmid, 1 µg of each sense 5'-phospho-tcgaTACTTAAACAAACAGAA-3' and antisense 5'-phospho-tcgaTTCTGTTTGTTTAAGTA-3' oligonucleotides were annealed and catenated with T4 DNA ligase (NEB, Ipswich, MA, USA) at 16ºC. The sequence of the putative FOXE1 response element (capital characters) was derived from previous work (34); overhangs (tcga) corresponding to *Xho*I sites and allowing subsequent ligation are shown in lower case. Catenated products were resolved in 1% TAE-agarose, fragments between 200 and 300bp were excised from the gel, purified using a FastGene Gel/PCR Extraction kit (Nippon Genetics, Kawaguchi City, Saitama, Japan), and ligated with *Xho*I-digested and shrimp alkaline phosphatase-treated (Takara Bio Inc., Otsu, Shiga, Japan) pGL4.23[luc2/minP]. Plasmids obtained from individual colonies were screened by PCR and subsequent sequencing in order to identify a clone containing the 10xFRE insert in the sense orientation upstream of the minimal promoter; this plasmid was further propagated, sequenced and used in downstream experiments.

Transfection and Dual luciferase assay

Assays were performed with the Dual Luciferase Reporter Assay System (Promega, Madison, WI, USA) according to the manufacturer's protocols. Cells were co-transfected by

electroporation using a Neon Transfection System (Invitrogen, Carlsbad, CA, USA) with 0.6 µg FOXE1 expression plasmid, 0.6 µg firefly and 6 ng renilla luciferase reporter vectors, maintained in 24-well plates and assayed for luciferase activity after 48h.

The transactivation of *TPO* or *FRE* promoters by different FOXE1 poly-Ala variants was determined as the ratio between firefly and renilla luciferase signals, relative to the ratio obtained in the cells co-transfected with the corresponding *FOXE1* expression plasmids and a non-modified pCMV6-AC-IRES-GFP as control. All experiments were performed in quadruplicates and reproduced several times.

Western blotting and quantitative real-time PCR

One day before transfection, $6x10^5$ Nthy-ori 3-1 cells were plated in a 10 cm dish in medium without antibiotics. The following day, cells were cotransfected with 4 µg of pCMV6-AC-IRES-GFP-FOXE1 expression plasmids or empty pCMV6-AC-IRES-GFP vector, 4 µg of pGL4.23-10xFRE and 80 ng of pGl4.74 luciferase reporter vectors using 20 µl of Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. After 3 hours of incubation with DNA-Lipofectamine complexes, the medium was replaced with fresh medium without antibiotics. After 48 hours, cells were scraped in ice-cold PBS for subsequent protein extraction for the dual luciferase assay, Western blotting and DNA isolation.

For the dual luciferase assay, approximately one-third of cells were treated with 1×Passive Lysis Buffer (Promega, Madison, U.S.A.), and the assay was performed as described above.

For Western blotting, approximately one-third of cells were lysed in a buffer containing 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 5% glycerol, 2 mM PMSF, 50 mM NaF, 10 mM sodium pyrophosphate, 1 mM sodium orthovanadate and 1x Complete Protease Inhibitor Cocktail (Roche Diagnostics K.K., Tokyo, Japan). After measuring protein concentration with a Pierce BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA), 40 µg of lysates were resolved in 4-15% gradient SDS-polyacrylamide gel (Mini-Protean TGX, Bio-Rad, Hercules, CA, USA), and blotted onto a PVDF membrane (Trans-Blot Turbo Transfer Pack mini, Bio-Rad, Hercules, CA, USA). After blocking with 10% skim milk in TBST for 1 h, the membrane was incubated overnight at 4°C with anti-TTF2 rabbit polyclonal antibodies (PA0200, Biopat, Perillo Sant'Angelo a Cupolo, Italy) diluted 1:1000 in 5% skim milk in TBST. After 2x washing for 5 min with TBST, secondary HRP-conjugated anti-rabbit antibodies (sc7074, Cell Signaling Technology, Tokyo, Japan) diluted 1:1000 in 5% skim milk in TBST were applied for 1 h at RT. After three wash steps for 5 min in TBST, the membrane was incubated in Pierce Western Blotting Substrate (Thermo Scientific, Rockford, IL, USA) for 1min at ambient temperature. Luminescence detection was performed in a LAS-4000 mini imaging system (Fujifilm, Tokyo, Japan). Next, after incubation with Western Blot Stripping Buffer

(Thermo Scientific, Rockford, IL, USA) for 45 min at 37ºC, the membrane was reprobed for 1 h at room temperature with a primary anti-β-actin mouse monoclonal antibody (sc-827, Santa Cruz Biotechnology, Santa Cruz, CA, USA) diluted 1:1000 in 5% skim milk in TBST and secondary HRP-conjugated anti-mouse antibody (sc7076, Cell Signaling Technology, Tokyo, Japan) diluted 1:1000 in 5% skim milk in TBST. β-actin signal was detected as described above. After densitometry, FOXE1 levels were normalized to the corresponding β-actin levels.

DNA was extracted from the remaining cells using a QIAamp DNA Mini Kit (QIAGEN, Tokyo, Japan). Quantitative real-time PCR was performed with the following primers: TG forward, GTGAGGGCACACATGCTTCAT and TG reverse, CGGAGCTTTGCTTCCTCACA (amplifying a 113 bp fragment of the human thyroglobulin (*TG)* gene, gene ID 7038), and FOXE1 forward, CGCCATGCTGCCGCTTAT and FOXE1 reverse, CTTATCGTCGTCATCCTTGTAATCCAG (amplifying a 126 bp region of plasmid-encoded FOXE1), and SYBR Premix Ex Taq II reagent (Takara Bio Inc., Otsu, Shiga, Japan). All reactions were performed in a Thermal Cycler Dice Real Time System II (Takara Bio Inc., Otsu, Shiga, Japan) under the same conditions: 95ºC for 30 sec, then 40 cycles of [95ºC for 5 sec and 60ºC for 30 sec] followed by dissociation curve analysis to ensure the signal from target amplicon. Plasmid DNA quantity was normalized for nuclear DNA.

Statistical analysis

Differences between case and control groups for each *FOXE1* polymorphism were examined using logistic regression analysis in the multiplicative model of inheritance. Association, linkage disequilibrium and haplotype analyses were performed using "gap" and "haplo.stats" R packages. Analysis of transactivation effects of different FOXE1 isoforms and correlation analysis was performed using IBM SPSS Statistics Version 21 (International Business Machines Corp., Armonk, NY, USA) and GraphPad InStat Version 3.10 (GraphPad Software, San Diego, CA, USA). *p*-values were 2-sided and considered significant if < 0.05.

Results

Genotyping

All individuals from Japan (501 patients with PTC and 748 population controls) were successfully genotyped for rs965513, rs1867277 and the *FOXE1* poly-Ala polymorphism. In the Belarusian series (660 patients with PTC and 820 population controls), rs965513 genotypes were obtained for all participants, rs1867277 genotypes were determined in 624 PTCs and 760 controls (0.945 and 0.927 call rates, respectively), and the *FOXE1* poly-Ala polymorphism was characterized in 635 PTCs and 777 controls (0.962 and 0.948 call rates, respectively). Genotyping results are shown in Table 1.

Information on *FOXE1* polyalanine tract length is extremely limited for Asian populations. We therefore present the prevalence and distribution of the corresponding alleles and genotypes in our series in some detail. In the Japanese cohort, the length of the poly-Ala tract varied from 14 to 17 repeats (except for 13 and 15) in healthy individuals, and from 11 to 16 (except for 13 and 15) in PTC patients. In the Belarusian cohort the range was from 12 to 19 repeats (except for 15 and 18) in healthy individuals, and from 12 to 19 (except for 13, 15 and

18) in PTC patients. The *FOXE1* Ala14 allele was the most prevalent in both cohorts (Fig. 1), accounting for 99.0% in healthy Japanese individuals and 97.5% in Japanese PTC patients, and for 59.7% and 51.5%, respectively, in the Belarusian cohort. The second prevalent allele was Ala16, which was observed with the frequencies of 0.9-1.2% in the Japanese series and 34.6-41.7% in the Belarusian cohort.

Homozygosity for the Ala14/14 genotype was the most frequent and found in 98.0% in healthy Japanese individuals and 95.4% in Japanese PTC patients (Supplementary Table 1). The second prevalent genotype in the Japanese cohort presented by the two most frequent alleles, heterozygosity for Ala14/16 was observed only in 1.9% of healthy individuals and 2.2% patients; no homozygous Ala16/16 genotype was found. In the Belarusian cohort, corresponding genotypes accounted for 36.0%, 40.9% and 12.1% in healthy individuals, and for 27.1%, 42.0% and 18.0% in PTC patients.

Association analysis of *FOXE1* polymorphisms with PTC

All three genetic variants displayed statistically significant individual association signals, which remained significant after Bonferroni correction for multiple testing in both Japanese and Belarusian cohorts (Table 1). The differences in effect sizes between the two ethnic groups were not statistically significant ($p > 0.08$-0.90, the Breslow-Day test). There was, however, a noticeable difference in the minor allele frequencies of all studied polymorphisms ranging from 0.01 to 0.15 in the Japanese cohort as compared to 0.35 to 0.48 in the Belarusian cohort. The frequency of *FOXE1* poly-Ala variants other than 14 repeats was particularly low in the Japanese individuals; this was the most likely reason for insufficient power to detect a significant association of the *FOXE1* poly-Ala16 allele with PTC ($p > 0.5$) clearly seen in the Belarusian cohort ($p = 1.219E-04$).

To determine whether the effects of the tree polymorphisms are independent, we performed a conditional analysis (Table 2). In both ethnic groups, rs965513 remained significant under any condition (i.e., after it was conditioned on either rs1867277, poly-Ala or both polymorphisms), indicating that this is the lead SNP with an independent signal (lowest p = 1.967E-04).

In the Japanese cohort, rs1867277 remained significant under all conditions, although with smaller effect size and weaker significance (lowest p = 1.386E-03) than for rs965513, strongly suggestive of its independent effect. In contrast, the *FOXE1* poly-Ala was weakly significant after conditioning on the distal rs965513 (p = 0.039), but lost the association signal after conditioning on the proximal rs1867277 or on both rs965513 and rs1867277 (p = 0.070 and 0.253, respectively). Although because of low frequency of variants other than poly-Ala14, this result needs to be interpreted with some caution, it suggests that the *FOXE1* poly-Ala may not have an independent effect. In the Belarusian cohort, rs1867277 lost its significance after conditioning on rs965513 or on rs965513 and poly-Ala combined (p = 0.475 and 0.338, respectively) but remained weakly significant after conditioning on the latter (p = 0.040). Thus, the effect of rs1867277 could not be detected upon comparison with that of rs965513, but could be distinguished from the poly-Ala in this ethnic group. An independent role of *FOXE1* poly-Ala was not observed after conditioning on either rs965513 or on both rs965513 and rs1867277 (lowest p = 0.589), corroborating these findings in the Japanese cohort.

Linkage disequilibrium analysis and *FOXE1* haplotype association with PTC

The above-mentioned two observations, namely the drastic differences in allelic frequencies of genetic variants between the Japanese and Belarusian cohorts, and a high collinearity between rs1867277 and *FOXE1* poly-Ala in the Belarusian cohort in the conditional

regression models (Table 2 footnote), prompted us to evaluate LD between the three polymorphisms and to examine haplotype associations with PTC in the two ethnic groups.

LD in the Japanese cohort was substantially weaker as compared to that in the Belarusian series (Fig. 2). In the Japanese cohort, only four haplotypes were identified, likely due to the low frequency of polymorphisms under analysis. All these haplotypes were associated with PTC (Table 3). The most prevalent haplotype was observed with about 80% prevalence, included all protective alleles (i.e., rs965513[G], rs1867277[G] and poly-Ala14), and negatively associated with PTC risk (OR = 0.547, p = 6.601E-09). Of note, two consequent haplotypes associating with an increased risk for PTC (OR = 1.548, p = 7.432E-04; and OR = 1.968, p = 1.088E-04, respectively) harbored either one rs1867277[A] or one rs965513[A] allele, but both contained the protective poly-Ala14. This observation again supports the limited contribution of the *FOXE1* poly-Ala to the risk of PTC, as compared to those of rs965513 and rs1867277.

In the Belarusian cohort, among seven haplotypes, only two were significantly associated with PTC (Table 3). Similarly to the results in the Japanese cohort, the most prevalent haplotype, accounting for about 50%, included all protective alleles and was negatively associated with PTC (OR = 0.688, p = 4.810E-07). The only haplotype conferring elevated risk for PTC contained risk alleles of all there polymorphisms (i.e., rs965513[A], rs1867277[A] and poly-Ala_non-14). Findings in the Belarusian series, again, do not enable distinguishing independent contributions of polymorphisms to the risk for developing PTC, most likely due to strong LD in this ethnic group.

In view of differences in allelic frequencies of the three polymorphisms under study, we additionally assessed whether other genetic variants in the 100 kb region of chromosome 9q22.33 encompassing rs965513 and the *FOXE1* gene may have similar frequencies in populations of Asian and European ancestry. In Asian populations (JPT and HCB), minor allele

444 frequencies of the common SNPs varied from 0.003 to 0.453 (mean 0.108, median 0.091)
445 (Supplementary Table 2). In European individuals the range was from 0 to 0.498 (mean 0.278,
446 median 0.337). The difference between allelic frequencies in the two ethnic groups was
447 statistically significant (p = 6.705E-8), demonstrating that polymorphic variants in this
448 chromosomal region are less frequent in Asian than in European populations.
449
450 **Transactivation effects of FOXE1 isoforms**
451 Five different *FOXE1* expression constructs (encoding 12, 14, 15, 16 and 19 poly-Ala
452 repeats) were functionally examined in dual luciferase assays for their ability to activate the
453 reporter expression driven by the minimal 32 bp promoter, or TPO or 10xFRE promoters in an
454 immortalized normal human thyroid cell line (Nthy-ori 3-1) and three differentiated human
455 thyroid cancer cell lines (TPC1, KTC-1 and WRO) which have relatively low level of
456 endogenous FOXE1 protein (Supplementary Fig. 2).
457 A strong decrease in transactivation potential of FOXE1 isoforms with increasing poly-
458 Ala tract length was observed for all cell lines and promoters (Supplementary Fig. 3 and
459 Supplementary Table 3). The negative correlation was also seen for data aggregated by cell- or
460 promoter-type (Fig. 3, Table 4 univariate analysis). Of note, the FOXE1 poly-Ala tract length-
461 dependent activation of the minimal promoter was also significant (Fig. 3 and Table 4 univariate
462 analysis) and deserves special attention since the minimal promoter is a constituent part of TPO-
463 and 10xFRE-driven reporter vectors. When these activities were controlled for minimal promoter
464 activation, the effect of the FOXE1 poly-Ala tract length was no longer observed in regression
465 models (Table 4 multivariate analysis).
466 To address the reason for the decline in reporter signal with increasing FOXE1 poly-Ala
467 tract length, we performed Western blotting, dual luciferase assay and real-time PCR analysis of

corresponding materials from the same transfection experiments in Nthy-ori 3-1 cells. The declining pattern of reporter signal was reproduced as in previous functional assays (Fig. 4). Surprisingly, we found that despite equal amounts of different *FOXE1* poly-Ala isoform-encoding plasmids were used for transfections, FOXE1 protein levels and the levels of cell-associated *FOXE1* expression vectors also changed as a function of the poly-Ala tract length. Strong positive correlations between the three endpoints of these assays were confirmed statistically. A plausible explanation could be that plasmids encoding FOXE1 isoforms of increasing length may display declining transfection efficacies or a declining stability inside cells after transfection (but not due to vector degradation before transfection, Supplementary Fig. 4) eventually affecting the plasmid-encoded FOXE1 protein level. Further experiments would be necessary to distinguish between these scenarios and shed light on the underlying mechanism.

Discussion

In this study, we aimed at answering whether three functional polymorphisms on chromosome 9q22.33, which have been reported in association with thyroid cancer, may play independent roles. It is noteworthy that the associations of genetic variants in the *FOXE1* locus have been initially reported in genome-wide studies for both adult sporadic and radiation-related PTC in young patients with similar effect sizes in the cohorts of Caucasian origin (1, 3). The associations were later replicated in Japanese and Chinese studies of adult sporadic thyroid cancer, also identifying similar effect sizes (4, 7). Thus, there is no evidence for either age-dependent, etiological or ethnical correlations for inherited genetic variants at this locus that would cause potential bias once association analyses are performed separately within the ethnic groups.

Several converging lines of evidence indicated that two SNPs, rs965513 and rs1867277, located 60 kb upstream or immediately in the *FOXE1* gene, respectively, are likely to have independent signals. In contrast, an independent role for *FOXE1* poly-Ala could not be demonstrated.

First, the association signals of rs965513 and rs1867277 were replicated in the Japanese and Belarusian populations with effect sizes very similar to those reported before, i.e. with an OR of 1.6-1.9 for rs965513 and of 1.5-2.0 for rs1867277 (1-14, 25, 26). An association of the *FOXE1* poly-Ala tract with PTC was also confirmed in both populations with an allelic OR comparable to previously published values of 1.3-2.5 in different ethnicities (5, 14, 26). In the Belarusian population, the association signal of the *FOXE1* poly-Ala tract was seen in the poly-Ala14/other and the poly-Ala other/16 models confirming the protective effect of the poly-Ala14 allele and the risk-conferring role of poly-Ala16, which is in line with earlier reports. In the Japanese cohort, a significant association signal was revealed only in the poly-Ala14/other model. The reason is that the *FOXE1* poly-Ala variant has a very low degree of variability in the Japanese population with a minor allele frequency of about 1-2%. It, however, did not hamper statistical analysis demonstrating a significant association with risk for PTC for non-poly-Ala14 alleles. In contrast, the association could not be demonstrated for the poly-Ala16 allele in the Japanese cohort due to its low frequency. A statistical power estimate of case-control sample size indicates the study should have enrolled about 38,000 participants to detect an effect of this allele at OR = 1.3.

The frequency of the *FOXE1* poly-Ala14 homozygotes among healthy Japanese individuals was 98.0%, in good agreement with 96.4% found in the Chinese population (32). Importantly, among Japanese patients with PTC, homozygous carriers of the protective *FOXE1* poly-Ala14 allele accounted for 95.4% (allelic frequency 97.5%). Given the apparent rarity of

the risk-associated non-poly-Ala14 alleles in this group, it would be difficult to assign them a causative role in conferring predisposition to thyroid cancer at the population level although rare variants may well be used for the identification of disease-associated genes or chromosomal regions (35).

Second, despite several studies having simultaneously genotyped more than one polymorphism in the *FOXE1* locus with or without the poly-Ala (5, 6, 12, 14, 26), only one examined their independent associations with thyroid cancer. The work by Jones et al. reported that rs965513 and rs1867277 are independent risk alleles based on the analysis of a large series of patients of Caucasian origin and controls from the United Kingdom (6). We performed a conditional analysis, which unambiguously demonstrated that rs965513 is a lead SNP with an independent association signal in both Japanese and Belarusian ethnic groups (Table 2). With regard to rs1867277 in the *FOXE1* 5'-UTR, the results were different between the Japanese and Belarusian cohorts. While in the Japanese series rs1867277 remained significant under all conditions, indicative of its independent signal, the association was lost in the Belarusian group after conditioning on rs965513, or on rs965513 and poly-Ala together, likely due to a strong LD between rs1867277 and rs965513. This finding in the Belarusian group is at variance with the report by Jones et al., and may stem from the difference in sample size or different LD relationships in the populations enrolled in the two works. However, our analysis in the Japanese population is in line with the report by Jones et al., and supports the independent effect of rs1867277.

In the Belarusian cohort, rs1867277 remained weakly significant after conditioning on *FOXE1* poly-Ala, supportive of the independent role of the former. In contrast, the *FOXE1* poly-Ala signal lost significance after conditioning on either the proximal rs1867277 in the 5'-UTR of the gene, or on the combined distal rs965513 and rs1867277 in either population. This finding

strongly suggests that the *FOXE1* poly-Ala is unlikely to have an independent effect, especially from rs1867277.

Third, LD relationships corresponded well with the results of conditional analysis. In the Japanese cohort, weak LD between the three polymorphisms allowed the intragenic rs1867277 and poly-Ala to remain significant after their conditioning on distal rs965513. However, the somewhat stronger LD between rs1867277 and the *FOXE1* poly-Ala results in a non-significant effect of *FOXE1* poly-Ala after conditioning on rs1867277, or on rs965513 and rs1867277 combined. In the Belarusian cohort, LD between the three polymorphisms was rather strong, in line with findings in the Portuguese population (5). It is probably for this reason that both rs1867277 and the *FOXE1* poly-Ala repeat lost significance after conditioning on the lead rs965513 SNP making their independent associations with PTC undetectable in this ethnic group. Also, the stronger association signal of rs1867277 and strong LD with *FOXE1* poly-Ala rendered the latter non-significant.

Assessment of haplotype associations with PTC in the two populations were in a good agreement with the results of our conditional regression analysis. Data from the Japanese group demonstrated a limited, if any, contribution of non-poly-Ala14 alleles to the risk haplotype(s), while the demonstration of independent roles of intragenic *FOXE1* polymorphisms in the Belarusian series has proved difficult, likely because of strong LD. Note that with regard to *FOXE1* poly-Ala tract, the results of haplotype analysis pertain to its length only, which was the focus of our study. Since a detailed sequence analysis of poly-Ala tract was not performed, it is difficult, for example, to determine whether poly-Ala14 alleles abundant in the Japanese population are identical-by-decent or not. The low frequency of other common genetic variants in the *FOXE1* locus in this ethnic group, however, suggests that our haplotype analysis is rather adequate, although not free of some potential bias.

Finally, although FOXE1 polyalanine tract length-dependent transactivation of the reporter expression was observed in our functional analyses, i.e. a decrease in transactivation with increasing FOXE1 polyalanine tract length, as previously reported for the poly-Ala14 and poly-Ala16 isoforms (26), a similar effect on the reporter vectors regulated by the minimal promoter only was also noticed. The reason for the FOXE1-dependent activation of the minimal promoter remains unclear and it is not known whether FOXE1 interacts directly with these sequences. Nevertheless, transactivation of the minimal promoter by FOXE1 should not be dismissed to avoid technical misinterpretation of the results of functional studies, which indicate that different transactivation capacities of FOXE1 isoforms with different poly-Ala tract length could not be accurately demonstrated in conventional reporter assays employing the vectors containing a particular minimal promoter. In our supportive experiments we also observed that the descending pattern of promoter activation with increasing *FOXE1* polyalanine tract length could likely be attributed to the decline in transfection efficacies or a declining stability of corresponding *FOXE1* expression vectors inside the cells after transfection seen as poly-Ala tract length-related changes in the vector DNA levels. These changes would be expected to affect corresponding FOXE1 protein levels, which were confirmed, and which may be the reason for the difference in the reporter signal intensities, thus masking the potential difference in transactivation activity of different FOXE1 isoforms, if exists. Special experiments, using redesigned promoter-driven reporter vector and controlling for intracellular transgenic FOXE1 levels would be necessary to demonstrate functional difference of FOXE1 isoforms.

Taken together, our findings show that rs965513 and rs1867277 SNPs independently associate with risk for thyroid cancer while the multinucleotide *FOXE1* poly-Ala polymorphism does not. It should be emphasized that on single-track association analysis *FOXE1* poly-Ala was nominally associated with PTC in our study, in both populations, in full agreement with previous

588 reports (5, 12, 14, 26). However, conditional analysis demonstrated the loss of association when
589 the effects of other SNPs were taken into consideration. Our LD analysis showed that the *FOXE1*
590 poly-Ala was in strong relationship with rs965513 and rs1867277 in the Belarusian cohort,
591 particularly with the latter. Since rs1867277 is significantly associated with the risk for thyroid
592 cancer, significant *FOXE1* poly-Ala association signal could be expected too. The likeliest
593 reason for this, however, is the strong LD of *FOXE1* poly-Ala with *bona fide* risk variant(s)
594 rather than own effect.

595 It is worth noting that both rs965513 (23, 24) and rs1867277 (2) are functionally
596 involved in the transcriptional regulation of *FOXE1* and/or *PTCSC2*. The risk allele of rs965513
597 was associated with decreased expression of *FOXE1*, unspliced *PTCSC2* and *TSHR* (thyroid
598 stimulating hormone receptor, gene ID: 7253) in normal thyroid tissue (24). Interestingly, our
599 recent study demonstrated that overexpression of *FOXE1* in the thyroids of transgenic mice
600 restrained the proliferation of follicular cells (36), in support of the functional effect of rs965513.
601 In our earlier study we also observed a correlation between immunohistochemical expression of
602 FOXE1 in PTC tissue and the rs1867277 genotype (37). Ectopic expression of *PTCSC2* in a
603 papillary carcinoma cell line resulted in altered expression of a subset of genes implicated in cell
604 cycle and cancer (24). Despite whether rs1867277 may regulate *PTCSC2* and the precise roles of
605 *FOXE1* and/or *PTCSC2* in thyroid cancer remain to be established in detail, the growing body of
606 evidence implicates namely the *FOXE1* and *PTCSC2* expression levels, which are at least in part
607 regulated by the functional SNPs, in predisposition to PTC.

608 While it seems reasonable to hypothesize that the poly-Ala polymorphism in the coding
609 region of *FOXE1* may also contribute to inherited risk for thyroid cancer, the results of our study
610 favor the notion that the associations with PTC of functional SNPs rs965513 and rs1867277 but

not of *FOXE1* poly-Ala polymorphism are independent. These findings provide a better understanding of the role of these genetic factors in predisposition to thyroid cancer.

Acknowledgments

We thank Ms. C. Hayashida (Nagasaki University) for technical assist. This work was supported in part by KAKENHI Grant Number 16H02774, 15K09438 and 26293142 from the Japan Society for the Promotion of Science (JSPS).

Author Disclosure Statement

No competing financial interests exist.

Corresponding Author

Vladimir Saenko, Ph.D.

Department of Health Risk Control, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7122; Fax: +81-95-819-7169

References

1. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir, M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T, Hjartarsson H, Mayordomo JI, de la Chapelle A, Hrafnkelsson J, Thorsteinsdottir U, Kong A, Stefansson K 2009 Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nat Genet **41**: 460-464.
2. Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, Pita G, Milne R, Maravall J, Ramos I, Andia V, Rodriguez-Poyo P, Jara-Albarran A, Meoro A, del Peso C, Arribas L, Iglesias P, Caballero J, Serrano J, Pico A, Pomares F, Gimenez G, Lopez-Mondejar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodriguez-Antona C, Gonzalez-Neira A, Matias-Guiu X, Santisteban P, Robledo M 2009 The variant rs1867277 in *FOXE1* gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet **5**:e1000637.
3. Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, Akulevich NM, Ratanajaraya C, Mitsutake N, Takamura N, Danilova LI, Lushchik ML, Demidchik YE, Heath S, Yamada R, Lathrop M, Matsuda F, Yamashita S 2010 The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. Hum Mol Genet **19**:2516-2523.
4. Matsuse M, Takahashi M, Mitsutake N, Nishihara E, Hirokawa M, Kawaguchi T, Rogounovitch T, Saenko V, Bychkov A, Suzuki K, Matsuo K, Tajima K, Miyauchi A, Yamada R, Matsuda F, Yamashita S 2011 The *FOXE1* and *NKX2-1* loci are associated with

susceptibility to papillary thyroid carcinoma in the Japanese population. J Med Genet **48**:645-648.

5. Tomaz RA, Sousa I, Silva JG, Santos C, Teixeira MR, Leite V, Cavaco BM 2012 *FOXE1* polymorphisms are associated with familial and sporadic nonmedullary thyroid cancer susceptibility. Clin Endocrinol (Oxf) **77**:926-933.

6. Jones AM, Howarth KM, Martin L, Gorman M, Mihai R, Moss L, Auton A, Lemon C, Mehanna H, Mohan H, Clarke SE, Wadsley J, Macias E, Coatesworth A, Beasley M, Roques T, Martin C, Ryan P, Gerrard G, Power D, Bremmer C, Tomlinson I, Carvajal-Carmona LG 2012 Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. J Med Genet **49**:158-163.

7. Wang YL, Feng SH, Guo SC, Wei WJ, Li DS, Wang Y, Wang X, Wang ZY, Ma YY, Jin L, Ji QH, Wang JC 2013 Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. J Med Genet **50**:689-695.

8. Kohler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, Cipollini M, Cristaudo A, Bambi F, Hoffmann P, Herms S, Kalemba M, Kula D, Harris S, Broderick P, Houlston R, Pastor S, Marcos R, Velazquez A, Jarzab B, Hemminki K, Landi S, Forsti A 2013 Genome-wide association study on differentiated thyroid cancer. J Clin Endocrinol Metab **98**:E1674-1681.

9. Damiola F, Byrnes G, Moissonnier M, Pertesi M, Deltour I, Fillon A, Le Calvez-Kelm F, Tenet V, McKay-Chopin S, McKay JD, Malakhova I, Masyakin V, Cardis E, Lesueur F, Kesminiene A 2014 Contribution of *ATM* and *FOXE1* (*TTF2*) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. Int J Cancer **134**:1659-1668.

683 **10.** Bonora E, Rizzato C, Diquigiovanni C, Oudot-Mellakh T, Campa D, Vargiolu M, Guedj M,
684 McKay JD, Romeo G, Canzian F, Lesueur F 2014 The *FOXE1* locus is a major genetic
685 determinant for familial nonmedullary thyroid carcinoma. Int J Cancer **134**:2098-2107.
686 **11.** Penna-Martinez M, Epp F, Kahles H, Ramos-Lopez E, Hinsch N, Hansmann ML, Selkinski
687 I, Grunwald F, Holzer K, Bechstein WO, Zeuzem S, Vorlander C, Badenhoop K 2014
688 *FOXE1* association with differentiated thyroid cancer and its progression. Thyroid **24**:845-
689 851.
690 **12.** Maillard S, Damiola F, Clero E, Pertesi M, Robinot N, Rachedi F, Boissin JL, Sebbag J,
691 Shan L, Bost-Bezeaud F, Petitdidier P, Doyon F, Xhaard C, Rubino C, Blanche H,
692 Drozdovitch V, Lesueur F, de Vathaire F 2015 Common variants at 9q22.33, 14q13.3, and
693 *ATM* loci, and risk of differentiated thyroid cancer in the French Polynesian population.
694 PLoS One **10**:e0123700.
695 **13.** Rogounovitch TI, Bychkov A, Takahashi M, Mitsutake N, Nakashima M, Nikitski AV,
696 Hayashi T, Hirokawa M, Ishigaki K, Shigematsu K, Bogdanova T, Matsuse M, Nishihara E,
697 Minami S, Yamanouchi K, Ito M, Kawaguchi T, Kondo H, Takamura N, Ito Y, Miyauchi A,
698 Matsuda F, Yamashita S, Saenko VA 2015 The common genetic variant rs944289 on
699 chromosome 14q13.3 associates with risk of both malignant and benign thyroid tumors in
700 the Japanese population. Thyroid **25**:333-340.
701 **14.** Pereda CM, Lesueur F, Pertesi M, Robinot N, Lence-Anta JJ, Turcios S, Velasco M,
702 Chappe M, Infante I, Bustillo M, Garcia A, Clero E, Xhaard C, Ren Y, Maillard S, Damiola
703 F, Rubino C, Salazar S, Rodriguez R, Ortiz RM, de Vathaire F 2015 Common variants at
704 the 9q22.33, 14q13.3 and *ATM* loci, and risk of differentiated thyroid cancer in the Cuban
705 population. BMC Genet **16**:22.

706 **15.** Zhu H, Xi Q, Liu L, Wang J, Gu M 2014 Quantitative assessment of common genetic
707 variants on FOXE1 and differentiated thyroid cancer risk. PLoS One **9**:e87332.
708 **16.** Zhuang Y, Wu W, Liu H, Shen W 2014 Common genetic variants on *FOXE1* contributes to
709 thyroid cancer susceptibility: evidence based on 16 studies. Tumour Biol **35**:6159-6166.
710 **17.** Kang J, Deng XZ, Fan YB, Wu B 2014 Relationships of *FOXE1* and *ATM* genetic
711 polymorphisms with papillary thyroid carcinoma risk: a meta-analysis. Tumour Biol
712 **35**:7085-7096.
713 **18.** Geng P, Ou J, Li J, Liao Y, Wang N, Xie G, Sa R, Liu C, Xiang L, Liang H 2015 *TITF1* and
714 *TITF2* loci variants indicate significant associations with thyroid cancer. Endocrine
715 doi:10.1007/s12020-015-0664-0.
716 **19.** Chadwick BP, Obermayr F, Frischauf AM 1997 *FKHL15*, a new human member of the
717 forkhead gene family located on chromosome 9q22. Genomics **41**:390-396.
718 **20.** Zannini M, Avantaggiato V, Biffali E, Arnone MI, Sato K, Pischetola M, Taylor BA,
719 Phillips SJ, Simeone A, Di Lauro R 1997 TTF-2, a new forkhead protein, shows a temporal
720 expression in the developing thyroid which is consistent with a role in controlling the onset
721 of differentiation. EMBO J **16**:3185-3197.
722 **21.** Trueba SS, Auge J, Mattei G, Etchevers H, Martinovic J, Czernichow P, Vekemans M,
723 Polak M, Attie-Bitach T 2005 *PAX8, TITF1,* and *FOXE1* gene expression patterns during
724 human development: new insights into human thyroid development and thyroid dysgenesis-
725 associated malformations. J Clin Endocrinol Metab **90**:455-462.
726 **22.** Fernandez LP, Lopez-Marquez A, Santisteban P 2015 Thyroid transcription factors in
727 development, differentiation and disease. Nat Rev Endocrinol **11**:29-42.
728 **23.** He H, Li W, Liyanarachchi S, Srinivas M, Wang Y, Akagi K, Wu D, Wang Q, Jin V, Symer
729 DE, Shen R, Phay J, Nagy R, de la Chapelle A 2015 Multiple functional variants in long-

730 range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. Proc Natl
731 Acad Sci U S A **112**:6128-6133.
732 **24.** He H, Li W, Liyanarachchi S, Jendrzejewski J, Srinivas M, Davuluri RV, Nagy R, de la
733 Chapelle A 2015 Genetic predisposition to papillary thyroid carcinoma: involvement of
734 *FOXE1*, *TSHR*, and a novel lincRNA gene, *PTCSC2*. J Clin. Endocrinol Metab **100**:E164-
735 172.
736 **25.**Kallel R, Belguith-Maalej S, Akdi A, Mnif M, Charfeddine I, Galofre P, Ghorbel A, Abid M,
737 Marcos R, Ayadi H, Velazquez A, Hadj Kacem H 2010 Genetic investigation of *FOXE1*
738 polyalanine tract in thyroid diseases: new insight on the role of *FOXE1* in thyroid carcinoma.
739 Cancer Biomark **8**:43-51.
740 **26.** Bullock M, Duncan EL, O'Neill C, Tacon L, Sywak M, Sidhu S, Delbridge L, Learoyd D,
741 Robinson BG, Ludgate M, Clifton-Bligh RJ 2004 Association of *FOXE1* polyalanine repeat
742 region with papillary thyroid cancer. J Clin Endocrinol Metab **97**:E1814-1819.
743 **27.** Brown LY, Brown SA 2004 Alanine tracts: the expanding story of human illness and
744 trinucleotide repeats. Trends Genet **20**:51-58.
745 **28.**Albrecht A, Mundlos S 2005 The other trinucleotide repeat: polyalanine expansion disorders.
746 Curr Opin Genet Dev **15**:285-293.
747 **29.** Hishinuma A, Ohyama Y, Kuribayashi T, Nagakubo N, Namatame T, Shibayama K,
748 Arisaka O, Matsuura N, Ieiri T 2001 Polymorphism of the polyalanine tract of thyroid
749 transcription factor-2 gene in patients with thyroid dysgenesis. Eur J Endocrinol **145**:385-
750 389.
751 **30.**Santarpia L, Valenzise M, Di Pasquale G, Arrigo T, San Martino G, Ciccio MP, Trimarchi F,
752 De Luca F, Benvenga S 2007 *TTF-2*/*FOXE1* gene polymorphisms in Sicilian patients with
753 permanent primary congenital hypothyroidism. J Endocrinol Invest **30**:13-19.

754 **31.** Carre A, Castanet M, Sura-Trueba S, Szinnai G, Van Vliet G, Trochet D, Amiel J, Leger J,
755 Czernichow P, Scotet V, Polak M 2007 Polymorphic length of *FOXE1* alanine stretch:
756 evidence for genetic susceptibility to thyroid dysgenesis. Hum Genet **122**:467-476.
757 **32.** Qin CR, Yao JL, Zhu WJ, Wu WQ, Xie JS 2011 *FOXE1* polyalanine tract length screening
758 by MLPA in idiopathic premature ovarian failure. Reprod Biol Endocrinol **9**:158.
759 **33.** Au AY, McBride C, Wilhelm KG Jr, Koenig RJ, Speller B, Cheung L, Messina M,
760 Wentworth J, Tasevski V, Learoyd D, Robinson BG, Clifton-Bligh RJ 2006 PAX8-
761 peroxisome proliferator-activated receptor gamma (PPARgamma) disrupts normal PAX8 or
762 PPARgamma transcriptional function and stimulates follicular thyroid cell growth.
763 Endocrinology **147**:367-376.
764 **34.** Aza-Blanc P, Di Lauro R, Santisteban P 1993 Identification of a cis-regulatory element and
765 a thyroid-specific nuclear factor mediating the hormonal regulation of rat thyroid peroxidase
766 promoter activity. Mol Endocrinol. **7**:1297-1306.
767 **35.** Carvajal-Carmona LG 2010 Challenges in the identification and use of rare disease-
768 associated predisposition variants. Curr Opin Genet Dev doi:S0959-437X(10)00085-7.
769 **36.** Nikitski A, Saenko V, Shimamura M, Nakashima M, Matsuse M, Suzuki K, Rogounovitch
770 T, Bogdanova T, Shibusawa N, Yamada M, Nagayama Y, Yamashita S, Mitsutake N 2016
771 Targeted Foxe1 overexpression in mouse thyroid causes the development of multinodular
772 goiter but does not promote carcinogenesis. Endocrinology **157**:182-195.
773 **37.** Bychkov A, Saenko V, Nakashima M, Mitsutake N, Rogounovitch T, Nikitski A, Orim F,
774 Yamashita S 2013 Patterns of FOXE1 expression in papillary thyroid carcinoma by
775 immunohistochemistry. Thyroid **23**:817-828.
776
777

778 **Figure legends**

779

780 Figure 1. Relative frequency of *FOXE1* poly-Ala alleles in the Japanese and Belarusian cohorts.
781 Frequencies are shown for the groups of (a) healthy individuals and (b) patients with PTCs in
782 each population.

783

784 Figure 2. Schematic representation of three polymorphisms in the *FOXE1* locus at chromosome
785 9q22.33 and corresponding LD measures. The *FOXE1* gene is shown as a rectangle with coding
786 region of a single exon shaded; unshaded regions represent the 5'- and 3'-UTR. Linear distances
787 between polymorphic sites are indicated (59.8 kb and 0.8 kb). For *D*' and r^2, the intensity of box
788 shading is proportional to the corresponding measures (black and white colors represent the
789 strong or weak LD, respectively).

790

791 Figure 3. Effect of different FOXE1 isoforms on activation of the reporter expression. (a):
792 Transactivation effect in normal thyroid and thyroid cancer cells for all types of promoters. (b):
793 Overall transactivation effect by different promoter type. Shown fold change values are log-
794 transformed; error bars represent the 95% CI. Statistical comparisons were performed with
795 Kruskal-Wallis test followed by Dunn's post-test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

796

797 Figure 4. Transactivation activity of different FOXE1 isoforms (on the 10xFRE promoter),
798 FOXE1 protein levels and real-time quantification of cell-associated FOXE1 isoform expression
799 vector levels in Nthy-ori 3-1 cells. Protein extracts for dual luciferase assays and Western
800 blotting, and DNA for real-time PCR assays were obtained from the portions of cells collected
801 from the same dishes 48 hours after transfection. The coefficient of determination (r^2) and

802 statistical significance of Pearson correlation coefficient are indicated for each pair of endpoints.
803 Shown are the results of a representative experiment. All experiments were reproduced three
804 times with a similar result.

Table 1. Association analysis of the three polymorphisms in the *FOXE1* locus in the Japanese and Belarusian series

	Genotypes in Controls					Genotypes in PTC					Allelic association	
Polymorphism	11	12	22	HWE[a]	MAF[b]	11	12	22	HWE	MAF	OR (95% CI)	*p*
Japan												
rs965513 G/A*	677	69	2	0.863	0.049	418	75	8	0.036	0.091	1.884 (1.377-2.579)	7.603E-05
rs1867277 G/A*	600	139	9	0.767	0.105	359	129	13	0.730	0.155	1.552 (1.223-1.968)	2.927E-04
poly-Ala14/other*	733	15	0	0.782	0.010	478	21	2	0.002	0.025	2.375 (1.263-4.464)	7.267E-03
poly-Ala other/16*	734	14	0	0.796	0.009	489	12	0	0.786	0.012	1.287 (0.590-2.805)	5.263E-01
Belarus												
rs965513 G/A*	323	398	99	0.160	0.363	185	325	150	0.751	0.473	1.590 (1.367-1.850)	1.274E-09
rs1867277 G/A*	286	356	118	0.679	0.389	170	304	150	0.513	0.484	1.459 (1.254-1.698)	8.443E-07
poly-Ala14/other*	280	367	130	0.601	0.403	172	310	153	0.567	0.485	1.383 (1.192-1.605)	1.820E-05
poly-Ala other/16*	333	350	94	0.890	0.346	219	302	114	0.577	0.417	1.347 (1.157-1.569)	1.219E-04

[a] Compliance with Hardy-Weinberger equilibrium, chi-square test

[b] Minor allele frequency

* Risk allele

Table 2. Conditional analysis of associations with PTC of the three polymorphisms in the *FOXE1* locus in the Japanese and Belarusian series

Polymorphism	Unconditioned	Conditional on rs965513	Conditional on rs1867277	Conditional on poly-Ala14[a]	Conditional on rs1867277 and poly-Ala14	Conditional on rs965513 and poly- Ala14	Conditional on rs965513 and rs1867277
Japan							
rs965513							
OR (95% CI)	1.884 (1.377-2.579)	NA[b]	1.896 (1.383-2.598)	1.888 (1.366-2.611)	1.838 (1.334-2.532)	NA	NA
p	7.603E-05		6.983E-05	1.202E-05	1.967E-04		
rs1867277							
OR (95% CI)	1.552 (1.223-1.968)	1.560 (1.228-1.981)	NA	1.524 (1.185-1.959)	NA	1.500 (1.170-1.923)	NA
p	2.927E-04	2.685E-04		1.015E-03		1.386E-03	
poly-Ala14							
OR (95% CI)	2.375 (1.263-4.464)	1.957 (1.035-3.704)	1.825 (0.952-3.497)	NA	NA	NA	1.471 (0.759-2.849)
p	7.267E-03	0.039	0.070				0.253
Belarus							
rs965513							
OR (95% CI)	1.590 (1.367-1.850)	NA	1.511 (1.198-1.906)	1.478 (1.185-1.844)	1.460 (1.150-1.855)	NA	NA
p	1.274E-09		4.991E-04	5.354E-04	1.911E-03		
rs1867277							
OR (95% CI)	1.459 (1.254-1.698)	1.085 (0.867-1.359)	NA	1.426 (1.016-2.000)	NA	1.192 (0.832-1.708)	NA
p	8.443E-07	0.475		0.040*		0.338*	
poly-Ala14							
OR (95% CI)	1.383 (1.192-1.605)	1.056 (0.853-1.307)	1.008 (0.719-1.412)	NA	NA	NA	1.100 (0.778-1.556)
p	1.820E-05	0.620	0.962*				0.589*

[a] The *FOXE1* poly-Ala14/other model

[b] Not applicable

* Variance inflation factor (VIF) $\geq$ 5 was observed in the model suggestive of a considerable collinearity between predictors

Table 3. Haplotype association with PTC in the Japanese and Belarusian series

	rs965513	rs1867277	poly-Ala	Frequency			OR (95% CI)	P
				Controls	Cases	Overall		
Japan								
1	G	G	14	0.852	0.759	0.814	0.547 (0.447 - 0.671)	6.601E-09
2	G	A	14	0.094	0.138	0.112	1.548 (1.206 - 1.986)	7.432E-04
3	A	G	14	0.043	0.081	0.059	1.968 (1.403 - 2.760)	1.088E-04
rare	*	*	*	0.011	0.021	0.014	1.980 (1.028 - 3.814)	0.043
Belarus								
1	G	G	14	0.525	0.432	0.483	0.688 (0.594 - 0.796)	4.810E-07
2	A	A	non-14	0.309	0.415	0.356	1.581 (1.359 - 1.840)	3.321E-09
3	G	A	non-14	0.058	0.051	0.055	0.870 (0.630 - 1.199)	0.417
4	A	G	14	0.047	0.056	0.051	1.206 (0.869 - 1.673)	0.275
5	G	G	non-14	0.033	0.025	0.030	0.753 (0.485 - 1.169)	0.229
6	G	A	14	0.020	0.019	0.020	0.940 (0.556 - 1.589)	0.894
rare	*	*	*	0.007	0.003	0.005	0.413 (0.133 - 1.283)	0.135

* Any nucleotide

824 Table 4. Regression analysis of the joint effect of FOXE1 poly-Ala isoforms and minimal promoter on reporter activity[a]

Cell type Factors	R^2	Anova p[b]	B[c]	SE(B)[d]	Beta[e]	p[f]
			Univariate analysis[g]			
Normal thyroid cells						
poly-Ala	0.894	1.03E-07	-0.365	0.035	-0.946	1.03E-07
Minimal promoter	0.951	6.87E-10	0.975	0.055	0.975	6.87E-10
Thyroid cancer cells						
poly-Ala	0.443	5.97E-07	-0.257	0.044	-0.666	5.97E-07
Minimal promoter	0.765	4.23E-15	0.874	0.074	0.874	4.23E-15
			Multivariate analysis[h]			
Normal thyroid cells	0.960	4.23E-09				
poly-Ala			-0.103	0.063	-0.267	0.130
Minimal promoter			0.725	0.164	0.725	0.001
Thyroid cancer cells	0.767	5.22E-14				
poly-Ala			0.030	0.047	0.077	0.534
Minimal promoter			0.936	0.123	0.936	1.83E-09

825

826 [a] Transcriptional activity of *FOXE1* isoforms with 12, 14, 15, 16 or 19 alanine repeats were tested; mean fold change in luciferase activity
827 compared to an empty pCMV6-AC-IRES-GFP vector served as an outcome variable; all variables were ln-transformed and standardized
828 by promoter type

829 [b] Statistical significance of the regression model for each type of cells

830 [c] Regression coefficient of the factor

831 [d] Standard error of the regression coefficient

832 [e] Standardized regression coefficient

833 [f] Statistical significance of the regression coefficient

834 [g] Effects of FOXE1 poly-Ala isoforms and of minimal promoter tested independently

835 [h] Effects of FOXE1 poly-Ala isoforms and of minimal promoter tested simultaneously

836

837 Supplementary Table 1. Distribution of *FOXE1* poly-Ala genotypes in the Japanese and Belarusian series

Allele 1	Allele 2	Japan controls, N (%)	Japan PTC, N (%)	Belarus controls, N (%)	Belarus PTC, N (%)
11	14	0	2 (0.4)	0	0
12	12	0	1 (0.2)	0	1 (0.2)
12	14	0	7 (1.4)	9 (1.2)	12 (1.9)
12	16	0	1 (0.2)	6 (0.8)	6 (0.9)
12	17	0	0	0	1 (0.2)
12	19	0	0	1 (0.1)	0
13	14	0	0	1 (0.1)	0
14	14	733 (98.0)	478 (95.4)	280 (36.0)	172 (27.1)
14	15	0	1 (0.2)	0	0
14	16	14 (1.9)	11 (2.2)	318 (40.9)	267 (42.0)
14	17	1 (0.1)	0	13 (1.7)	16 (2.5)
14	19	0	0	26 (3.3)	15 (2.4)
16	16	0	0	94 (12.1)	114 (18.0)
16	17	0	0	9 (1.2)	14 (2.2)
16	19	0	0	17 (2.2)	15 (2.4)
17	17	0	0	0	1 (0.2)
17	19	0	0	2 (0.3)	0
19	19	0	0	1 (0.1)	1 (0.2)

838

Supplementary Table 2. Statistical analysis of transactivation activity of different FOXE1 polyAla isoforms[a] in different cell lines by promoter type[b]

Cells	Promoter[c]	Anova p^d	p_{het}[e]	Slope[f]	p_{trend}[e]
		Normal thyroid cells			
Nthy-ori 3-1	Minimal	0.012	0.008	-0.250	<0.0001
	TPO	0.036	0.520	-0.156	0.0003
	FRE	0.004	0.374	-0.336	<0.0001
		Thyroid cancer cells			
TPC1	Minimal	0.002	0.246	-0.371	<0.0001
	TPO	0.002	0.364	-0.190	<0.0001
	FRE	0.004	0.260	-0.415	<0.0001
KTC-1	Minimal	0.095	0.814	-0.077	0.048
	TPO	0.361	0.910	0.049	0.253
	FRE	0.0005	0.431	-0.194	<0.0001
WRO	Minimal	0.002	0.804	-0.161	<0.0001
	TPO	0.002	0.641	-0.160	<0.0001
	FRE	0.001	0.896	-0.323	<0.0001

[a] *FOXE1* inserts with 12, 14, 15, 16 and 19 alanine repeats cloned into pCMV6-AC-IRES-GFP vector were tested

[b] Ln-transformed fold change in luciferase activity compared to an empty pCMV6-AC-IRES-GFP vector were analyzed

[c] Promoter regulating *luc2* expression in the pGL4.23 reporter vector

[d] Non-parametric Anova, the Kruskal-Wallis test

[e] Bartlett's test for equal variances; $p > 0.05$ indicates the absence of significant heterogeneity between data

[f] Slope of the linear trend

[e] Statistical significance of the linear trend

Supplementary Table 3. Correlation between luciferase activity in the reporter assays and *FOXE1*-carrying vector content (DNA) and expression levels (cDNA) of the *FOXE1* and *Neo* plasmid genes in transfected Nthy-Ori 3.1 cells

		DNA		**cDNA**	
		Neo	*FOXE1*	*Neo*	*FOXE1*
Luciferase	*r*	0.456	0.479	0.231	0.562
	p	0.087	0.071	0.408	0.029
Neo DNA	*r*	1.000	0.954	0.425	0.607
	p		3.8E-8	0.114	0.016
FOXE1 DNA	*r*		1.000	0.350	0.536
	p			0.201	0.040
Neo cDNA	*r*			1.000	0.804
	p				3.1E-4

Luciferase activity, as expected, correlated significantly ($p = 0.029$) with the *FOXE1* expressed from the plasmid. In turn, *FOXE1* cDNA level correlated significantly with both *FOXE1*- and *Neo*-pCMV6-AC-IRES-GFP plasmid DNA content ($p = 0.040$ and $p = 0.016$, respectively) in transfected cells. The latter can only be observed if patterns of cDNA level and of corresponding plasmid DNA content change by *FOXE1* isoform in a similar manner. Note that luciferase activity also tended to correlate with the plasmid DNA content ($p = 0.087$ and $p = 0.071$ for *Neo* and *FOXE1*, respectively). These data strongly suggest that different *FOXE1* isoform-encoding plasmids may either display different transfection efficacies or, alternatively, different stability inside cells after transfection with equal amounts of each plasmid. The absence

865 of plasmid DNA degradation before transfection was confirmed by resolving intact or
866 MluI/XhoI-digested *FOXE1* expression vectors in agarose gel (Supplementary Fig. 4).
867 Nthy-Ori 3.1 cells in 6-well plates were transfected in duplicates with each *FOXE1*
868 isoform-carrying (or empty) pCMV6-AC-IRES-GFP/promoter-luciferase/renilla-luciferase
869 plasmid cocktail as described in the Transfection and Dual luciferase assay subsection. Protein
870 extract from one well was used for routine dual luciferase assay, and DNA and RNA were
871 extracted from the cells from the replica well. RNA was cleaned-up with RNase-Free DNase Set
872 (QIAGEN, Tokyo, Japan), purified using Isogen (Nippon Genetics, Kawaguchi City, Saitama,
873 Japan), precipitated with isopropanol and reverse transcribed with SuperScript III First-Strand
874 Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA). Quantitative real-time PCR was
875 performed using DNA or cDNA as template in triplicate with primers specified in
876 Supplementary Table 4 using SYBR Premix EX Taq II reagent (Takara Bio Inc., Otsu, Shiga,
877 Japan). All reactions were performed in a Thermal Cycler Dice Real Time System II (Takara Bio
878 Inc., Otsu, Shiga, Japan) under the same conditions: 95ºC for 30 sec, then 40 cycles of [95ºC for
879 5 sec and 60ºC for 30 sec] followed by dissociation curve analysis to ensure the signal from
880 target amplicon. Plasmid DNA quantity was normalized for the nuclear DNA (TG primers;
881 Thyroglobulin, gene ID 7038); cDNA levels of the *Neo* (encodes NeoR in the plasmid) and
882 *FOXE1* genes were normalized for the expression of a housekeeping gene (EMC7 primers; ER
883 membrane protein complex subunit 7, gene ID 56851; reference (S1)). Correlation analysis (for

884 Spearman's *r* and *p*-value) was performed using IBM SPSS Statistics Version 21 (International
885 Business Machines Corp., Armonk, NY, USA).
886
887 **Supplementary reference**
888 **S1.** Eisenberg E, Levanon EY 2013 Human housekeeping genes, revisited. Trends in Genetics
889 **29**:569–574.
890

Supplementary Table 4. Primers for quantitative real-time PCR

Primer	Sequence, 5'-3'	Genomic Localization, GRCh38/hg38 assembly	Amplicon size for DNA, bp	Amplicon size for cDNA, bp	Application
TG forward	GTGAGGGCACACATGCTTCAT	Chr8: 132893608 - 132893628	113	NA[1]	DNA
TG reverse	CGGAGCTTTGCTTCCTCACA	Chr8: 132893701- 132893720			DNA
EMC7 forward[2]	CGGGATCAAATCTGTAAGCTG	Chr15: 34095929 - 34095949	5696	107	cDNA
EMC7 reverse[2]	AGAGCACGTCGGTTTCCTTA	Chr15: 34101605 - 34101624			cDNA
Neo forward	ACCTTGCTCCTGCCGAGAAA	plasmid	125	125	DNA, cDNA
Neo reverse	CCGAGTACGTGCTCGCTCGAT	plasmid			DNA, cDNA
FOXE1 forward	CGCCATGCTGCCGCTTAT	plasmid	126	126	DNA, cDNA
FOXE1 reverse[3]	CTTATCGTCGTCATCCTTGTAATCCAG	plasmid			DNA, cDNA

[1] Not applicable

[2] Primer sequences were downloaded from qPrimerDepot, a quantitative real time PCR primer database (http://primerdepot.nci.nih.gov/; accessed October 2015)

[3] This primer anneals to the transcribed part of the pCMV6-AC-IRES-GFP backbone which is localized at the 3' end of the *FOXE1* insert impeding amplification of endogenous *FOXE1* message

Supplementary figure legends

Supplementary Figure 1. An example of a chromatogram of *FOXE1* poly-Ala tract analysis. The read peaks correspond to size standards, the blue peaks – to the PCR products under analysis. Analysis was performed by PCR amplification of the *FOXE1* fragment with the primers flaking the poly-Ala encoding region: forward 5'-CCCCAACGCGGAGGAC-3' and reverse 5'-CCGCTCAGGAACCAGGC-3'. Amplicon size achievable with this primer pair is 301 bp if the length of FOXE1 poly-Ala tract is 16 repeats (i.e., encoded by 48 nucleotides) corresponding to the NCBI Reference Sequence NG_011979.1. Each PCR reaction contained 10 pM of the 5'FAM-labeled forward primer, 95 pM of unlabeled forward primer, 100 pM of unlabeled reverse primer (all primers from FASMAC Co., Ltd., Atsugi, Japan), 200 pM of each dNTP, 25mU of ExTaq HS polymerase (Takara Bio Inc., Otsu, Japan), 10% v/v of dimethylsulfoxide (Sigma-Aldrich, St. Louis, MO, USA) and 25 ng of template DNA in a total volume of 10 µl. Reactions were done using the following thermal settings: 94°C for 1 min and 35 cycles of [94°C for 30 sec / 56°C for 5 sec / 72°C for 20 sec] in a C1000 Touch Thermal Cycler (BioRad, Indianapolis, USA). PCR products (0.5 µl) were then diluted 1:30 with 14.5 µl of formamide (Roche, Indianapolis, IN, USA) containing 0.1 µl of Genescan 400HD ROX Standard (Applied Biosystems, Foster City, CA, USA). The mixture was denatured at 95°C for 1 min, immediately chilled on ice and loaded to an ABI Prism 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA). Data acquisition was performed in GeneScan mode; analysis was performed with GeneMapper version 3.7 software (Applied Biosystems, Foster City, CA, USA). Direct sequencing of PCR products obtained with unlabeled primers was used for fine adjustment of GeneScan data. For these purpose, 2 µl of PCR product were treated with ExoSap reagent (Affymetrix, Santa Clara, CA, USA) and sequenced with the forward primer in the presence of

BigDye Terminator v3.1 Cycle Sequencing Kit reagents (Applied Biosystems, Foster City, CA, USA). Reaction products were resolved in an ABI PRISM 3730xl genetic analyzer.

Supplementary Figure 2. Western blot analysis of endogenous FOXE1 levels in Nthy-ori 3-1, TPC1, KTC-1 and WRO cell lines used in functional assays in comparison with normal rat thyroid PCCL3 cells. Nthy-ori 3-1, TPC1, KTC-1 and WRO cells were cultured as described in Materials and Methods. PCCL3 cells were cultured in H4 complete medium consisting of Coon's medium/F12 high zinc supplemented with 0.3 mg/ml L-glutamine, 1 mIU/ml TSH, 10 μg/ml insulin, 5 μg/ml apo-transferrin, 10 nM hydrocortisone, 5% fetal bovine serum and 1% penicillin/streptomycin. Cells were lysed in a buffer containing 20mM Tris-HCl, 150mM NaCl, 1mM EDTA, 0.5% Triton X-100, 5% glycerol, 2mM PMSF, 50mM NaF, 10mM sodium pyrophosphate, 1mM sodium orthovanadate and 1X cOmplete protease inhibitor cocktail (Roche Diagnostics K.K., Tokyo, Japan). After measuring protein concentration with Pierce BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA), 30μg of proteins were resolved in 10% SDS-polyacrylamide gel, and blotted onto Immobilon-P PVDF membranes (Merck Millipore, Darmstadt, Germany). After blocking with 10% skim milk in TBS/0.05% Tween 20 for 1h, the membranes were incubated with anti-TTF2 rabbit polyclonal antibodies (PA0200, Biopat, Perrillo Sant'Angelo a Cupolo, Italy) diluted 1:1000 overnight at 4°C. To ensure equivalent loading, membranes were stripped and reprobed with anti-β-actin mouse monoclonal antibody (sc-827, Santa Cruz Biotechnology, Santa Cruz, CA, USA). The signals were visualized with HRP-conjugated anti-rabbit or anti-mouse IgG secondary antibody (Cell Signaling Technology, Tokyo, Japan) and Pierce Western Blotting Substrate (Thermo Scientific, Rockford, IL, USA). Detection was performed in a LAS-4000 mini imaging system (Fujifilm, Tokyo, Japan).

Supplementary Figure 3. Functional analysis of transactivation potential of different FOXE1 isoforms by dual luciferase assay. Cells in 24-well plates were co-transfected with FOXE1 expression plasmid, and firefly and renilla luciferase reporter vectors, and assayed for luciferase activities after 48h. The activation of TPO- or 10xFRE-driven promoters by different FOXE1 poly-Ala isoforms was determined as the ratio between firefly and renilla luciferase signals, relative to the ratio obtained in the cells co-transfected with the corresponding expression plasmids and a non-modified pGL4.23[luc2/minP]. Data are shown for each cell line for the three types of promoters (Minimal, TPO and 10xFRE). Statistical comparisons were performed with Kruskal-Wallis test followed by Dunn's post-test. * $p < 0.05$; ** $p < 0.01$. Note an apparent inverse correlation between the reporter signal and the length of FOXE1 poly-Ala tract.

Supplementary Figure 4. Intactness of FOXE1 expression vectors before transfection. Non-modified pCMV6-AC-IRES-GFP plasmid or plasmids containing inserts of *FOXE1* with different poly-Ala tract lengths were treated or not with MluI and XhoI enzymes in BSA-supplemented NEB2 buffer for 2 h, heat-inactivated at 65ºC for 20 min, resolved in 1% agarose-TAE gel and visualized with ethidium bromide. M, 1 kb DNA ladder (NEB, Ipswich, MA, USA); the lower band corresponds to 500 bp. Predicted size of the *FOXE1* poly-Ala12 insert between MluI/XhoI cloning sites is 1117 bp including the overhangs of the restriction sites.

Figure 1

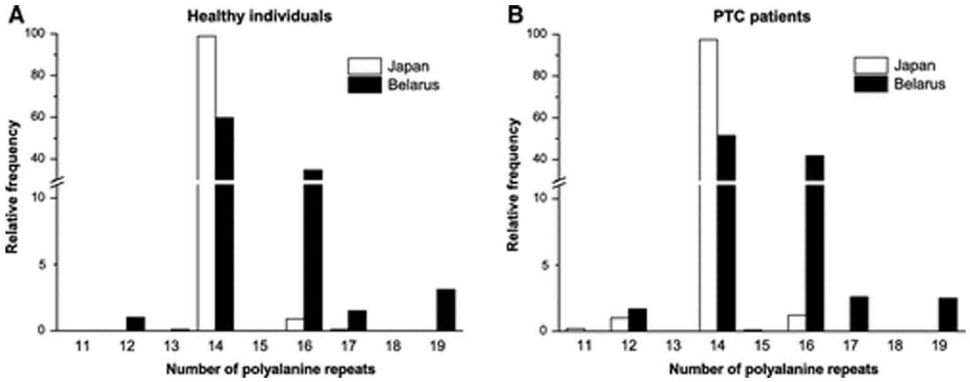


Figure 2

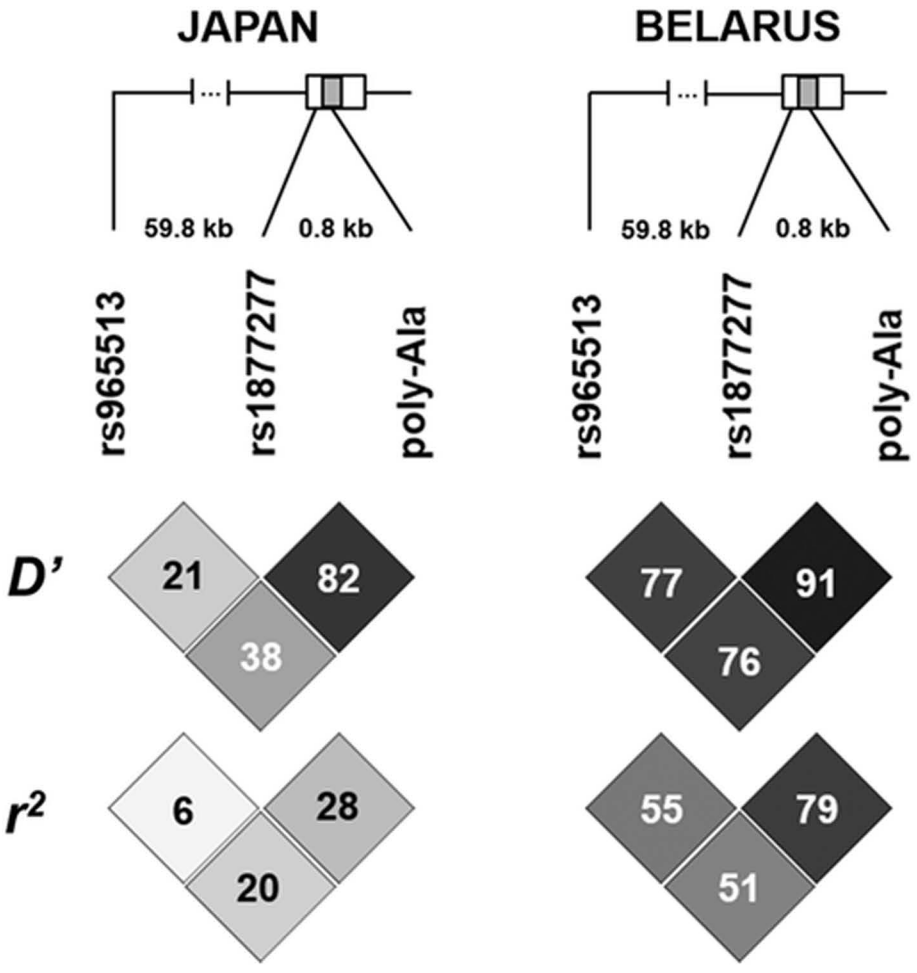


Figure 3

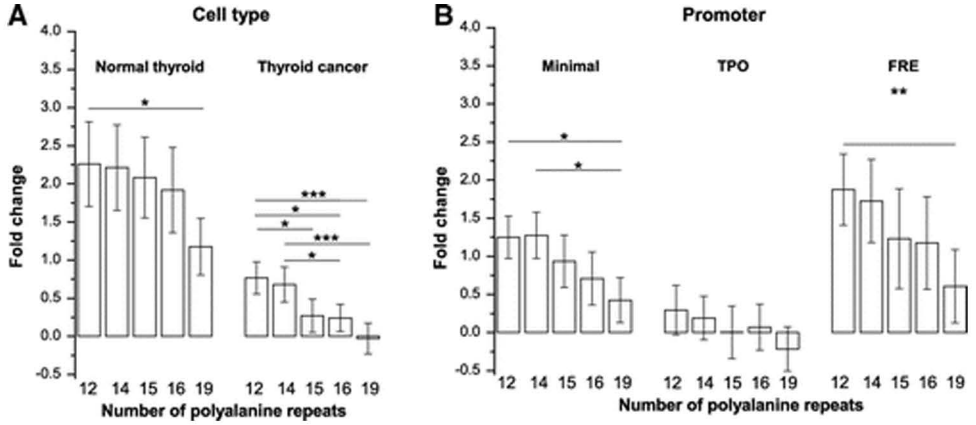
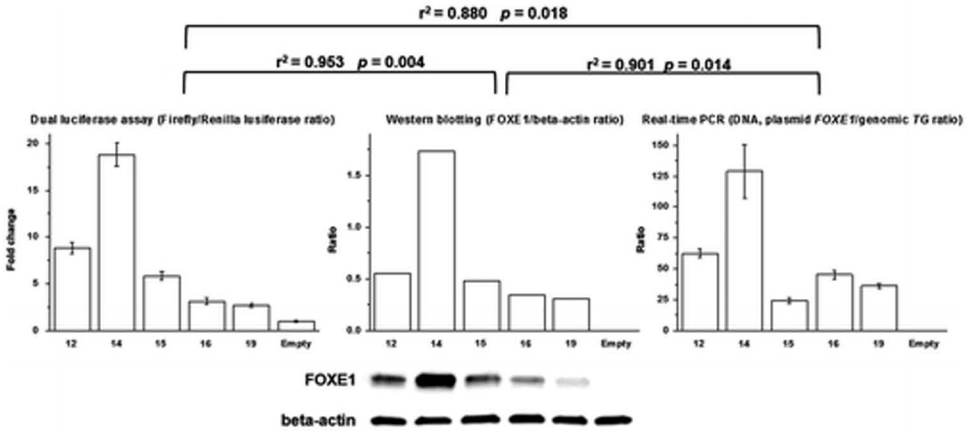
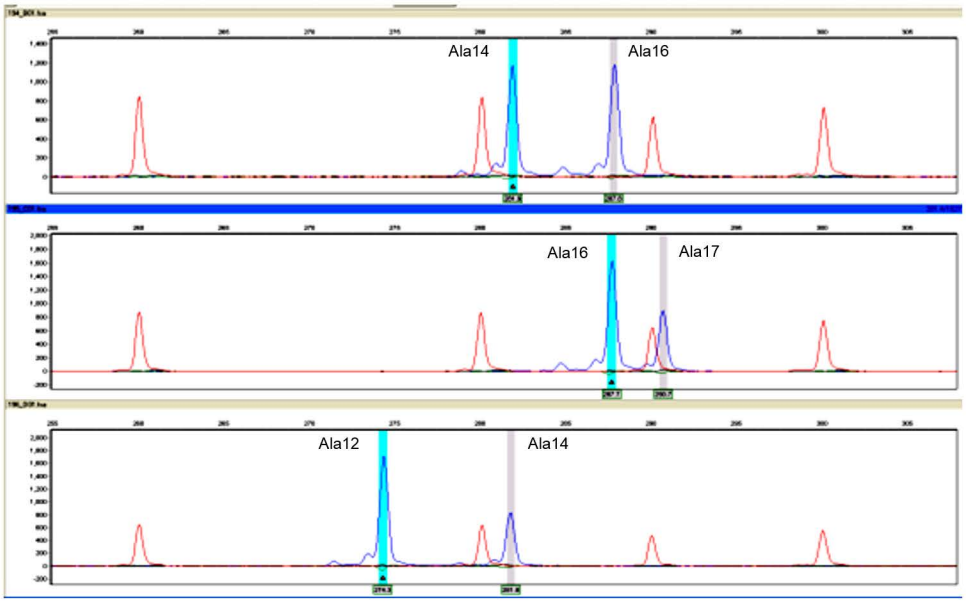


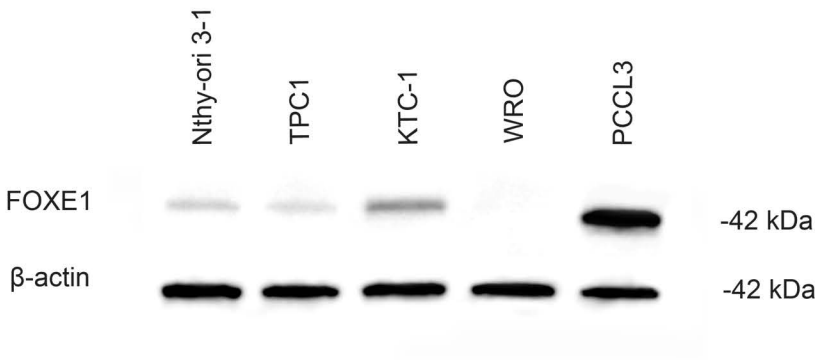
Figure 4



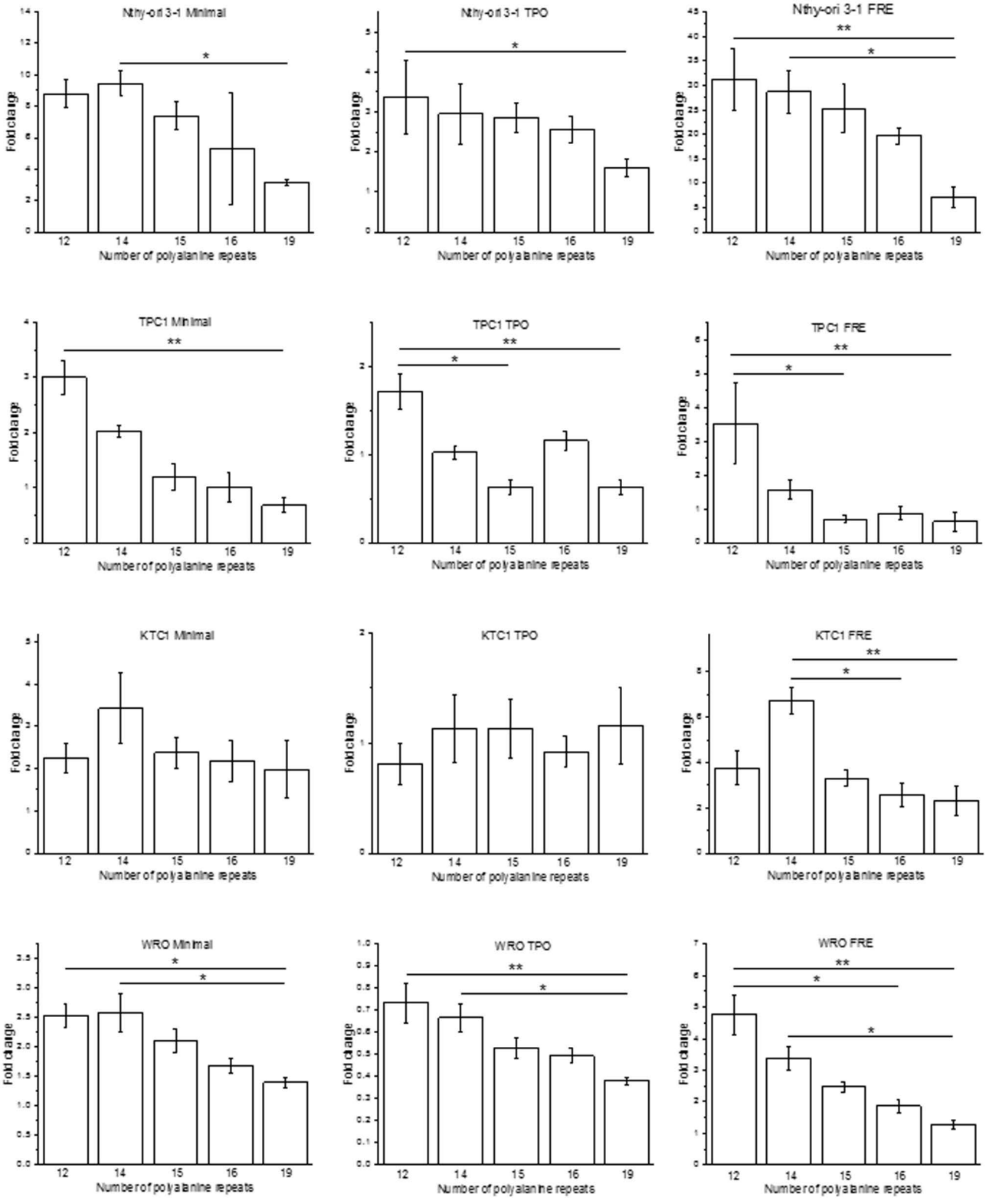
Supplementary Fig. 1



Supplementary Fig. 2



Supplementary Fig. 3



Supplementary Fig. 4