의학석사 학위논문

Genomic evidence of reactive oxygen species elevation in papillary thyroid carcinoma with Hashimoto thyroiditis

하시모토 갑상선염이 동반된 갑상선 유두암에서 활성산소 증가에 대한 유전체적 증거

2016년 8월

서울대학교 대학원

의학과 외과학

이 진 욱

의학석사 학위논문

Genomic evidence of reactive oxygen species elevation in papillary thyroid carcinoma with Hashimoto thyroiditis

하시모토 갑상선염이 동반된 갑상선 유두암에서 활성산소 증가에 대한 유전체적 증거

2016년 8월

서울대학교 대학원

의학과 외과학

이 진 욱

A thesis of the Master's degree

하시모토 갑상선염이 동반된 갑상선 유두암에서 활성산소 증가에 대한 유전체적 증거

Genomic evidence of reactive oxygen species elevation in papillary thyroid carcinoma with Hashimoto thyroiditis

August 2016

The Department of Medicine

Seoul National University Graduate School

Jin Wook Yi

하시모토 갑상선염이 동반된 갑상선 유두암에서 활성산소 증가에 대한 유전체적 증거

지도 교수 이 규 언

이 논문을 의학석사 학위논문으로 제출함

2016년 8월

서울대학교 대학원
의학과 외과학
이 진 욱

이진욱의 의학석사 학위논문을 인준함
2016년 8월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)

Genomic evidence of reactive oxygen species elevation in papillary thyroid carcinoma with Hashimoto thyroiditis

by
Jin Wook Yi

A thesis submitted to the Department of Medicine
in partial fulfillment of the requirements for the
Degree of Master of Science in Medicine (Surgery) at
Seoul National University Graduate School

August 2016

Approved by Thesis Committee:

Professor _____ Chairman

Professor _____ Vice chairman

Professor _____

Abstract

Introduction: Elevated levels of reactive oxygen species (ROS) have been proposed as a risk factor for the development of papillary thyroid carcinoma (PTC) in patients with Hashimoto thyroiditis (HT). However, it has yet to be proven that the total levels of ROS are sufficiently increased to contribute to carcinogenesis. We hypothesized that if the ROS levels were increased in HT, ROS-related genes would also be differently expressed in PTC with HT.

Methods: To find differentially expressed genes (DEGs) we analyzed data from the Cancer Genomic Atlas, gene expression data from RNA sequencing: 33 from normal thyroid tissue, 232 from PTC without HT, and 60 from PTC with HT. We prepared 402 ROS-related genes from three gene sets by genomic database searching. We also analyzed a public microarray data to validate our results.

Results: Thirty-three ROS related genes were up-regulated in PTC with HT, whereas there were only nine genes in PTC without HT (*Chi-square p*-value < 0.001). Mean $\log_2$ fold changes of up-regulated genes was 0.562 in HT group and 0.252 in PTC without HT group (*t*-test *p*-value = 0.001). In microarray data analysis, 12 of 32 ROS-related genes showed the same differential expression pattern with statistical significance. In gene ontology analysis, up-regulated ROS-related genes were related with ROS metabolism and apoptosis. Immune function-related and carcinogenesis-related gene sets were enriched only in HT group in Gene Set Enrichment Analysis.

Conclusions: Our results suggested that ROS levels may be increased in PTC with HT. Increased levels of ROS may contribute to PTC development in

patients with HT.

Keywords: Reactive oxygen species, Thyroid cancer, Hashimoto thyroiditis, Gene expression, Bioinformatics

Student Number: 2011-23753

Contents

List of Tables

List of Figures

Introduction

The clinical association between Hashimoto thyroiditis (HT; chronic lymphocytic thyroiditis) and papillary thyroid carcinoma (PTC) has been extensively studied, but the precise relationship between these two diseases is still being debated. Several analyses of surgical resection specimens have shown that PTC is frequently accompanied by HT. Interestingly, PTC with HT is associated with a lower rate of neck lymph node metastasis and has a better prognosis than PTC without thyroiditis in spite of increased incidence correlation [1-10]. In contrast, other studies conducted by fine-needle aspiration biopsy specimens and ultrasonographic findings have found no correlation between the incidences of these two diseases, with researchers insisting that HT does not increase the risk of thyroid cancer [11-13]. To date these studies have consisted only of clinical retrospective reviews. To resolve the controversial relationship between these two diseases that occur in the same organ, additional biological and/or genomic evidence is needed.

One possible biological mechanism for a connection between these two diseases is the elevated levels of reactive oxygen species (ROS), expelled from the lymphocytes during the progression of HT [14, 15]. HT is a chronic autoimmune disease accompanied by extensive infiltration of lymphocytes. Lymphocytes produce ROS as a byproduct of cell-mediated immune processes. Although ROS production from the lymphocytes increases in HT, we cannot safely conclude that the total amount of intra-thyroidal ROS level also rises enough to induce the development of cancer because thyroid cells normally

produce huge amounts of hydrogen peroxide (H_2O_2), a typical form of ROS, during thyroid hormone biosynthesis. With the progression of HT, thyroid cells are destroyed by the lymphocytic invasion. Thyroid hormone production ceases and byproduct ROS production is also stopped as thyroid cell destruction proceeds [16]. In summary, ROS production from the lymphocytes was increased but ROS from normal thyroid cells was decreased. Therefore, it is difficult to estimate the changes in total intra-thyroidal ROS levels in thyroid gland with HT.

To circumvent these challenges, we took a bioinformatics approach to reveal the ROS level elevation in PTC with HT. We hypothesized that if the total intrathyroidal ROS level increases in HT due to the accumulation of lymphocytic ROS resulting in disruption of the redox balance, ROS-related genes and pathways would be differently expressed from the PTC with/without HT. To test this hypothesis, we analyzed gene expression data from RNA-sequencing in the Cancer Genome Atlas (TCGA, The Cancer Genome Atlas research network: http://cancergenome.nih.gov) and microarray data in Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/).

Materials and Methods

TCGA data processing and making subgroups

As of May 2014, TCGA group has released multigenomic analyses results from 502 PTC patients with partially open access. We downloaded clinical information and normalized RNA-sequencing mRNA read count results. Fig. 1 outlines the steps for creating subgroups. TCGA RNA sequencing data included 558 results: 58 from normal thyroid tissue, 492 from primary PTC tissue, and 8 from metastatic tissue. Each sample was reported as 20,531 rows (genes) and one column (read counts, in units of RSEM [RNA-Seq. by Expectation Maximization]), with upper-quartile normalization performed by the TCGA team. We integrated 58 normal thyroid tissue and 492 primary PTC results into a meta-table that comprised 20,531 rows (genes) and 550 columns (patients). Eight results from metastatic tissue were excluded. Next, we examined all clinical records and excluded samples with lack of pathologic information. Patients with a history of external radiation exposure were also excluded because the radiation could have degraded ROS-related systems and thus acted as a critical confounding factor in this ROS-related study.

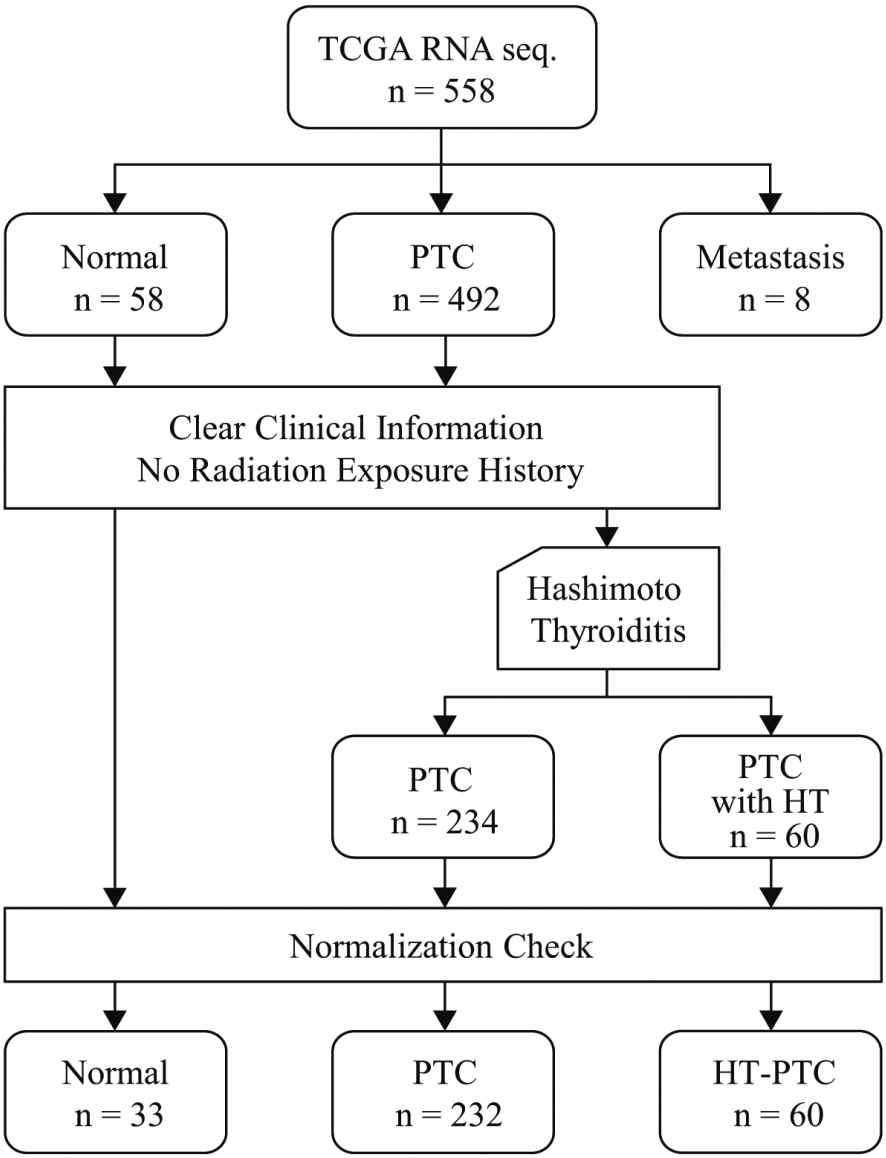

Fig 1. Flow chart for grouping TCGA RNA-sequencing results according to clinical information. The final groups used for our analyses were: (1) Normal, (2) PTC, and (3) HT-PTC.

We split the PTC samples into two groups according to pathologic status, i.e., PTC without HT and PTC with HT. Although the TCGA data had already been normalized, we verified the density plot to flag outlier samples ourselves and identified two outliers that we excluded from our analysis (Fig. 2). The final analysis included the following three groups: (1) Normal thyroid (n = 33, hereafter referred to as "Normal group", (2) PTC without HT (n = 232, "PTC group"), and (3) PTC with HT (n = 60, "HT-PTC group").

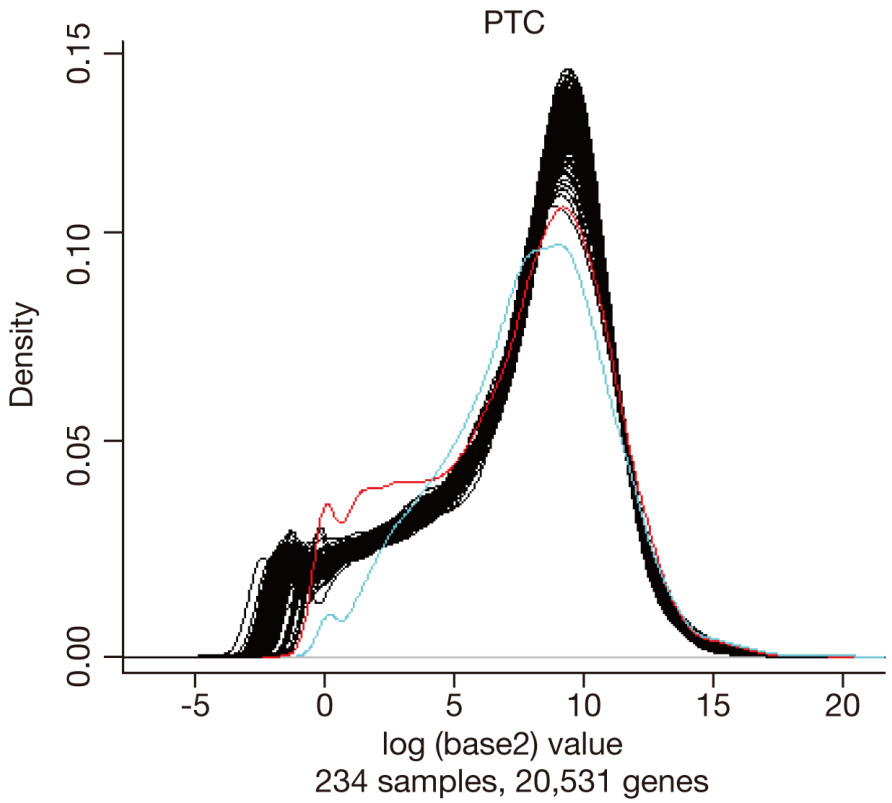

Fig 2. Density plot of the 234 initial PTC samples. Two outliers, drawn in red and sky blue, were excluded from the analysis.

Generating ROS-related gene sets

We created three ROS-related gene sets to use in our analysis. Gene Set 1 was prepared by searching the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) profile database (http://www.ncbi.nlm.nih.gov/geoprofiles) and the AmiGO2 gene ontology database (http://amigo.geneontology.org/amigo) [17]. As of May 2014, the NCBI GEO profile database provided 182,114 pertinent gene records and the AmiGO2 search engine returned 358 genes, in response to our query terms: "reactive oxygen species, homo sapiens". From these numerous records, we extracted unique gene names and IDs, including 364 unique genes that are known to be associated with the ROS-related system in the human system.

The information for Gene Sets 2 and 3 was obtained from the Gene Ontology Consortium (http://geneontology.org/), specifically, the gene ontology terms related to "reactive oxygen species". Gene Set 2 includes 146 genes annotated with Gene Ontology (GO):0000302, "response to ROS" and Gene Set 3 contains 133 genes annotated with GO:0072593, "ROS metabolic process". As a result we compiled a total of 402 ROS related genes when the three gene sets are combined.

Validation Study using public microarray data from GEO database

To validate our results from TCGA data, we analyzed a public microarray repository from the Gene Expression Omnibus (GEO): "GSE29315, compilation of expression profiles in distinct thyroid neoplasias", conducted by Tomas G *et al*.

The microarray platform used was the Affymetrix Human Genome U95

Version 2 Array. GSE29315 data contain 12,625 gene expression counts from various types of thyroid tissues: 17 follicular adenomas, 9 follicular thyroid carcinomas, 13 follicular variant PTCs, 9 PTCs, 6 HT (HT-PTCs), 8 thyroid hyperplasias, and 9 Hurthle cell adenomas. We downloaded the microarray results served as "*.CEL" files from six HT-PTC and nine PTC results. To analyze microarray data with background correction, normalization, and extract gene expressions (RMA, Robust Multi-array Average), we used "affy" package in Bioconductor (http://www.bioconductor.org/) [18].

Statistical analysis

All statistical analyses were performed in R.3.1.2 [19], with graphs generated in the same program. Fig. 3 shows the steps of our statistical analysis. To find differentially expressed genes (DEGs) among the three groups, we performed one way ANOVA test with all 20,531 genes. To correct false positive rates from multiple comparisons, the False Discovery Rate method was used, with adjusted q-values < 0.05 considered as statistically significant. To find specific DEGs between the PTC and HT-PTC groups that represented differential expression according to just HT, and not from the normal versus cancer, we applied Tukey's honestly significant difference (HSD) test on the significant genes obtained from initial ANOVA test. We filtered out 2,111 significant genes that show differential expression pattern in PTC versus HT-PTC group only using Tukey's HSD test. From these filtered genes, we selected ROS-related DEGs that intersected with the 402 ROS-related genes that we had prepared above. The fold-change of each DEG was calculated by the following equation: log2 (mean expression value in HT-PTC/mean expression value in PTC). To

validate ROS-related DEGs in microarray data, unpaired *t*-test was applied on gene expression counts (RMA) between six HT-PTC and nine PTC groups. A *p*-value < 0.05 was considered as a statistically significant differential expression between the two groups.

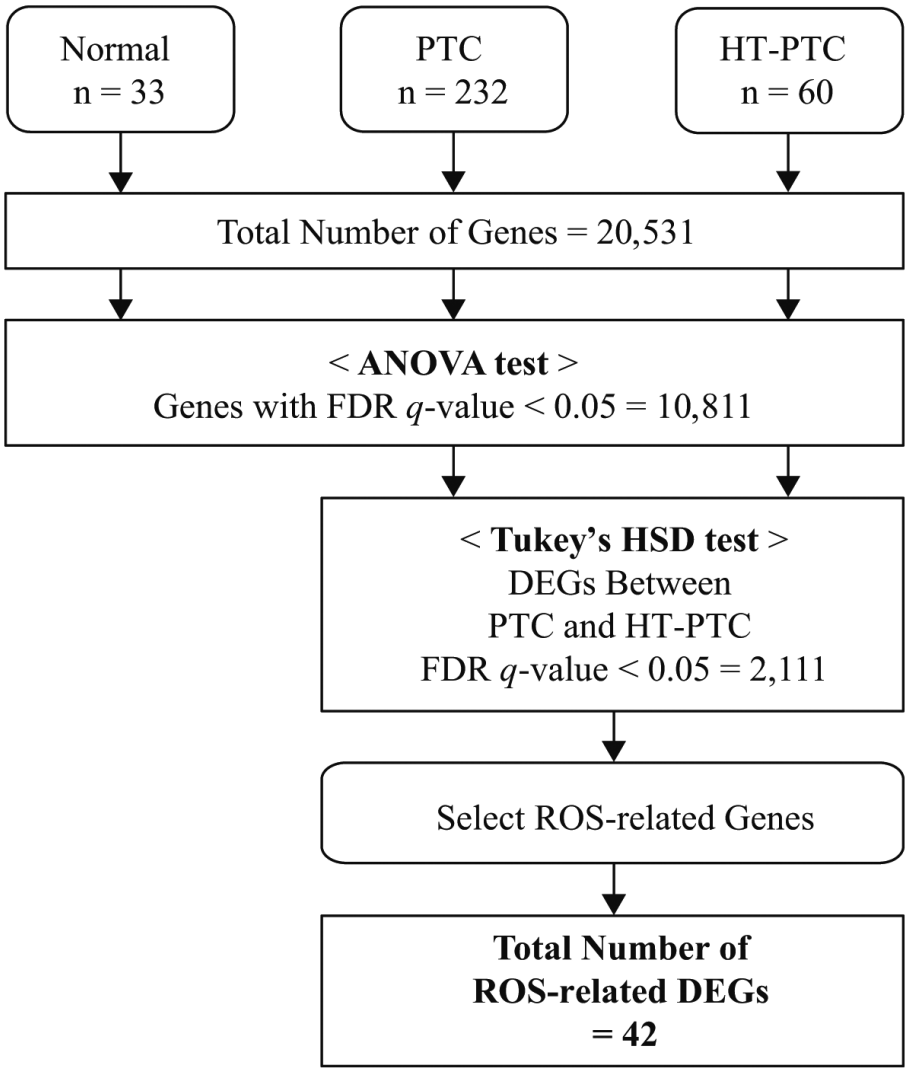

Fig 3. Steps in the statistical analysis used in this study. From the full set of 20,531 genes, we identified 42 DEGs between the PTC and HT-PTC groups.

Gene ontology (GO) and gene set enrichment analysis (GSEA)

After obtaining the ROS-related DEGs, we used the DAVID (Database for Annotation, Visualization and Integrated Discovery Bioinformatic Resources 6.7) program to check each DEG's dominantly-expressed tissue and its ROS-related biologic function such as ROS-producing or ROS-eliminating (scavenging) from microenvironment [20]. We also checked gene ontology enrichment using up-regulated genes in HT-PTC or PTC group with DAVID gene ontology analyzing tool. The Gene Set Enrichment Analysis (GSEA) program was used for pathway analysis [21]. We initially inputted all 20,531 gene expression counts from the RNA sequencing RSEM results to find which pathways were differently enriched between HT-PTC and PTC group. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to reference pathway database in GSEA program [22]. Additionally, we used our three ROS-related gene sets as another metric to determine whether these specific ROS-related gene sets were enriched or not in the HT-PTC group. In GSEA, an FDR q-value < 0.25 was considered to indicate statistically significant enrichment.

Results

DEG analysis

Table 1 shows the gene name, their description, dominantly-expressed tissue, and ROS-related function of 42 ROS-related DEGs between HT-PTC and PTC. Thirty-three genes were up-regulated in HT-PTC group and nine genes were up-regulated in PTC group. Among the 33 genes that were up-regulated in HT-PTC, 22 genes were dominantly expressed in immune function-related cells, such as lymphocytes, neutrophils, lymph nodes, and so on, and 11 genes were thyroid cells. Five genes were related to ROS production whereas 15 genes were correlated with the removal or elimination of ROS from tissue. In the PTC group, five genes were expressed in immune cells and three genes were in thyroid cells. No ROS-producing gene and only one ROS-eliminating gene was found in PTC group. DEG's tissue dominancy and ROS related functions were summarized in table 2. 11.5% of immune cell dominantly expressed genes are related with ROS production and 34.6% are related with ROS elimination. In the thyroid dominantly expressed genes, 7.1% are linked with ROS production whereas 57.1% are associated with ROS elimination.

Table 1. 42 DEGs between HT-PTC versus PTC group. Dominantly expressed tissue and ROS related function of DEGs were indicated by check sign.

Gene Name	Description	Dominant Tissue		ROS Related Function	
		Immune cell*	Thyroid	Produce	Eliminate
SFTPC	Surfactant protein C				
CD38	CD38 molecule	✔			
TPO	Thyroid peroxidase	✔	✔		✔
IPCEF1	Interaction protein for cytohesin exchange factors 1		✔		✔
LCK	LCK proto-oncogene, Src family tyrosine kinase	✔			
NCF1	Neutrophil cytosolic factor 1	✔		✔	
BTK	Bruton agammaglobulinemia tyrosine kinase	✔			
PPARGC1A	Peroxisome proliferator-activated receptor gamma, coactivator 1 alpha				
TRPM2	Transient receptor potential cation channel, subfamily M, member 2				✔
CYBB	Cytochrome b-245, beta polypeptide	✔		✔	✔
BCO2	Beta-carotene oxygenase 2				✔
RAC2	Ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	✔			
PPIF	Peptidylprolyl isomerase F	✔	✔		
PDK1	Pyruvate dehydrogenase				

	kinase, isozyme 1				
SLC8A1	Solute carrier family 8 (sodium/calcium exchanger), member 1				✔
IL18BP	Interleukin 18 binding protein	✔			
NCF2	Neutrophil cytosolic factor 2			✔	
SFTPD	Surfactant protein D		✔		
MICB	MHC class I polypeptide-related sequence B	✔			
GPX3	Glutathione peroxidase 3 (plasma)		✔		✔
SOD2	Superoxide dismutase 2, mitochondrial	✔	✔	✔	✔
PTGS1	Prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	✔			✔
BAK1	BCL2-antagonist/killer 1	✔			
JAK2	Janus kinase 2			✔	
GCLC	Glutamate-cysteine ligase, catalytic subunit	✔			✔
PREX1	Phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 1	✔			
MSRB2	Methionine sulfoxide reductase B2	✔			✔
PRDX1	Peroxiredoxin 1		✔		✔
ETFDH	Electron-transferring-flavoprotein dehydrogenase	✔	✔		✔
C12orf5	Chromosome 12 open reading frame 5	✔			
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S	✔	✔		✔

	protein 1, 75kDa (NADH-coenzyme Q reductase)				
GRB2	Growth factor receptor-bound protein 2	✔	✔		
PRDX3	Peroxiredoxin 3	✔	✔		✔
ERCC3	Excision repair cross-complementation group 3	✔			
HDAC6	Histone deacetylase 6				✔
LIG1	Ligase I, DNA, ATP-dependent	✔			
PON2	Paraoxonase 2				
RBM11	RNA binding motif protein 11				
OXR1	Oxidation resistance 1	✔	✔		
ANXA1	Annexin A1	✔	✔		
PLK3	Polo-like kinase 3				
KLF4	Kruppel-like factor 4 (gut)	✔	✔		

* Lymphocytes, neutrophil and lymph nodes.

Table 2. Tissue dominancy and ROS related function in ROS related DEGs.

Dominant tissue	ROS related function	Numbers	Percentage
Immune cell	Production	3	11.5%
(n = 26)	Elimination	9	34.6%
Thyroid cell	Production	1	7.1%
(n = 14)	Elimination	8	57.1%

Fold changes of 42 ROS-related DEGs are shown in Fig. 4. Positive fold changes indicate that the genes had higher expression in the HT-PTC whereas negative fold changes indicate that the genes had higher expression in the PTC group. The proportions of up-regulated genes were significantly different between the two groups according to the Chi-square test ($p < 0.001$). Mean $\log_2$ fold change was 0.562 in the HT-PTC group and 0.252 in the PTC group. These values were also significantly different (t-test $p = 0.001$). In the HT-PTC group, 13 DEGs had $\log_2$ fold-change values greater than 0.5. However, no DEGs in the PTC group had $\log_2$ fold-change values greater than 0.5.

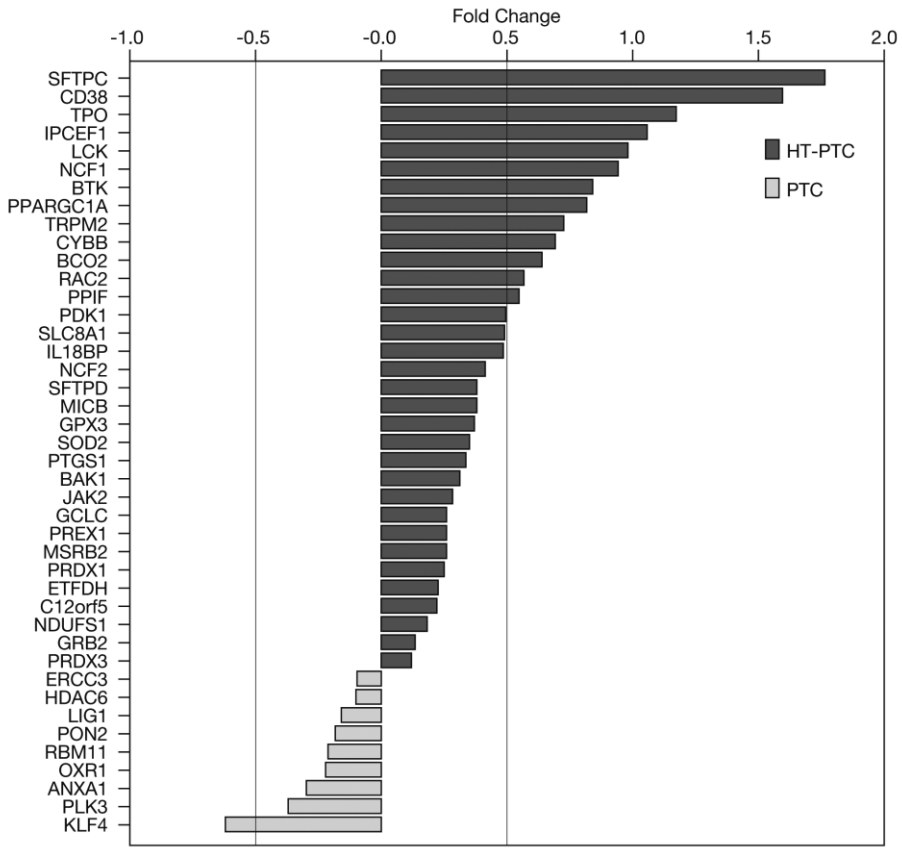

Fig 4. Log_2 fold changes of ROS-related DEGs ordered by fold-change values.

After obtaining DEGs from TCGA RNA sequencing data, we validated our results with another dataset from the microarray experiment. Among the 42 R OS related DEGs, 32 genes were represented in the microarray gene list. Twelve genes showed significant differential expression pattern under the *t*-test *p*-value < 0.05 between HT-PTC and PTC groups. Eleven genes were up-regulated in HT-PTC and one gene was up-regulated in PTC (Fig. 5). The differential expression patterns of these 12 genes were the same as the TCGA fold change results.

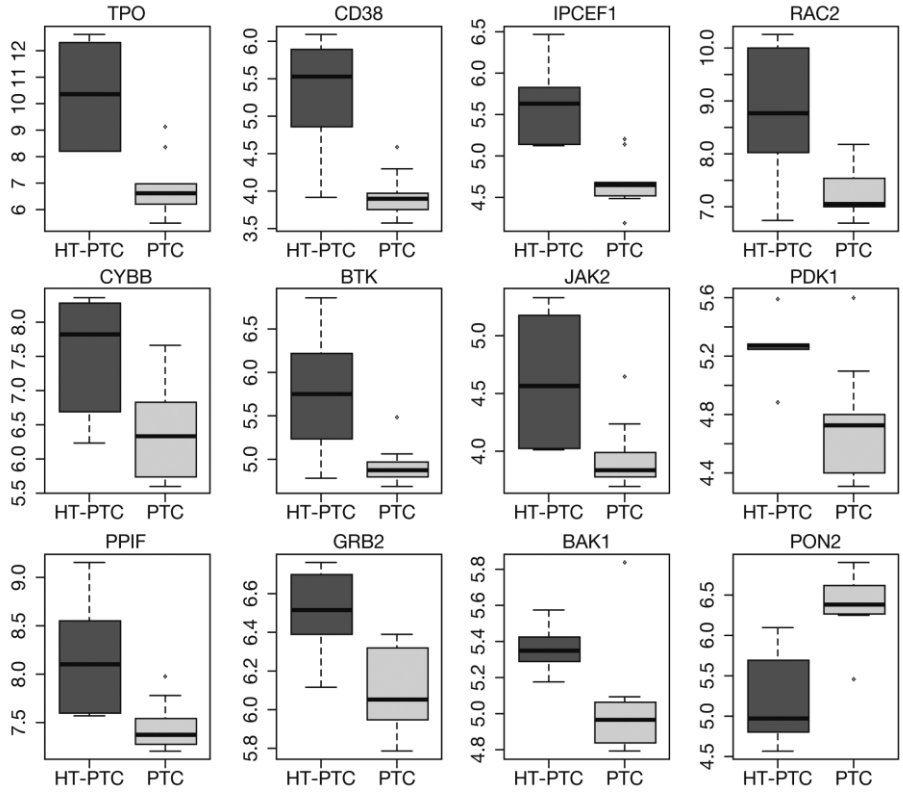

Fig 5. Twelve genes that have differential expression pattern under the *t*-test *p*-value < 0.05 between HT-PTC and PTC groups. Eleven genes were up-regulated in HT-PTC and one gene was up-regulated in PTC. Y-axis means the gene expression value, unit is RMA (Robust Multi-array Average).

To determine possible biased effects from clinical variables such as age, sex, or disease stage that may influence gene expression, we reclassified the PTC and HT-PTC groups according to clinical variables and again performed the *t*-test to compare the different clinical subgroups. No significant variables were identified that could have confounded our DEG results.

Gene ontology (GO) and gene set enrichment analysis

Table 3 shows the significantly enriched GOs in biologic process using up-regulated ROS-related DEGs as input to DAVID analysis. Total 24 GOs were significantly enriched. Twelve GOs were related to ROS metabolism and oxidative stress. Six GOs contained the term "homeostasis" and three GOs were related to cell apoptosis. There was no significantly enriched pathway using up-regulated DEGs in PTC group.

Table 3. Significantly enriched (FDR *q*-value < 0.05) gene ontologies (Biologic Process) in the HT-PTC group.

GO Number	Term	Count	Fold Enrichment	FDR
GO:0006979	Response to oxidative stress	14	36.088	<0.001
GO:0042592	Homeostatic process	13	7.318	<0.001
GO:0055114	Oxidation reduction	12	7.939	<0.001
GO:0006800	Oxygen and reactive oxygen species metabolic process	11	69.407	<0.001
GO:0019725	Cellular homeostasis	11	9.979	<0.001
GO:0048878	Chemical homeostasis	11	9.083	<0.001
GO:0042981	Regulation of apoptosis	11	5.784	0.011
GO:0043067	Regulation of programmed cell death	11	5.727	0.012
GO:0010941	Regulation of cell death	11	5.706	0.013
GO:0055082	Cellular chemical homeostasis	10	11.125	0.000
GO:0006873	Cellular ion homeostasis	9	10.173	0.002
GO:0050801	Ion homeostasis	9	9.303	0.004
GO:0010035	Response to inorganic substance	7	14.435	0.009
GO:0006801	Superoxide metabolic process	6	101.460	0.000
GO:0042743	Hydrogen peroxide metabolic process	6	101.460	0.000
GO:0042542	Response to hydrogen peroxide	6	45.295	0.000
GO:0000302	Response to reactive oxygen species	6	33.820	0.001
GO:0034614	Cellular response to reactive oxygen species	5	70.458	0.001
GO:0034599	Cellular response to oxidative stress	5	49.157	0.004
GO:0033194	Response to hydroperoxide	4	281.833	0.000
GO:0051881	Regulation of mitochondrial membrane potential	4	153.727	0.003
GO:0042554	Superoxide anion generation	4	130.077	0.005
GO:0042744	Hydrogen peroxide catabolic process	4	99.471	0.011
GO:0070301	Cellular response to hydrogen peroxide	4	93.944	0.013

In the GSEA with all 20,531 genes, 53 pathways were significantly enriched in HT-PTC group but no pathway was enriched in PTC group (Table 4). Ten pathways were related to immune process, and only one pathway, the chemokine signaling pathway, was correlated with carcinogenesis. The enrichment plots from GSEA using our three ROS-related gene sets are shown in Fig. 6. Gene Sets 1 and 3 were significantly enriched toward the HT-PTC group, using FDR q-value < 0.25. Gene Set 2 was also enriched toward HT-PTC side but had no significant q-value. This result suggests that the ROS-related gene sets are highly activated in the HT-PTC group.

Table 4. Significantly enriched pathways in HT-PTC group from the GSEA analysis.

Pathways in KEGG database	**FDR q-value**
Antigen processing and presentation *	0.012
Autoimmune thyroid disease *	0.026
Natural killer cell mediated cytotoxicity *	0.050
Type I diabetes mellitus	0.052
Intestinal immune network for IgA production	0.052
Systemic lupus erythematosus	0.058
Hematopoietic cell lineage	0.061
Asthma	0.062
T cell receptor signaling pathway *	0.063
Glycolysis gluconeogenesis	0.065
Graft versus host disease	0.068
Allograft rejection	0.069
Primary immunodeficiency	0.074
Ascorbate and aldarate metabolism	0.075
Pyruvate metabolism	0.076
Beta alanine metabolism	0.077
Chemokine signaling pathway *	0.079
Leishmania infection	0.080
Tryptophan metabolism	0.080
Butanoate metabolism	0.080
B cell receptor signaling pathway *	0.082
Propanoate metabolism	0.082
Valine leucine and isoleucine degradation	0.084
Proteasome	0.085
Viral myocarditis	0.092
Fatty acid metabolism	0.101
Terpenoid backbone biosynthesis	0.111
Porphyrin and chlorophyll metabolism	0.116
Pantothenate and coa biosynthesis	0.118
Citrate cycle TCA cycle	0.119
Peroxisome	0.119
Cell adhesion molecules cams *	0.121
Oxidative phosphorylation	0.122

Cytokine cytokine receptor interaction *	0.123
Cytosolic DNA sensing pathway	0.124
Toll like receptor signaling pathway *	0.126
Fc gamma R mediated phagocytosis *	0.145
Huntingtons disease	0.163
Parkinson's disease	0.166
Fructose and mannose metabolism	0.168
Nicotinate and nicotinamide metabolism	0.192
Fc epsilon RI signaling pathway *	0.197
Alzheimers disease	0.200
Amino sugar and nucleotide sugar metabolism	0.202
Lysine degradation	0.208
Pentose and glucuronate interconversions	0.209
Histidine metabolism	0.211
Protein export	0.216
RNA degradation	0.216
Arginine and proline metabolism	0.246
Vibrio cholerae infection	0.246
Snare interactions in vesicular transport	0.247
Glutathione metabolism	0.249

* : Immune process and cancer related pathways

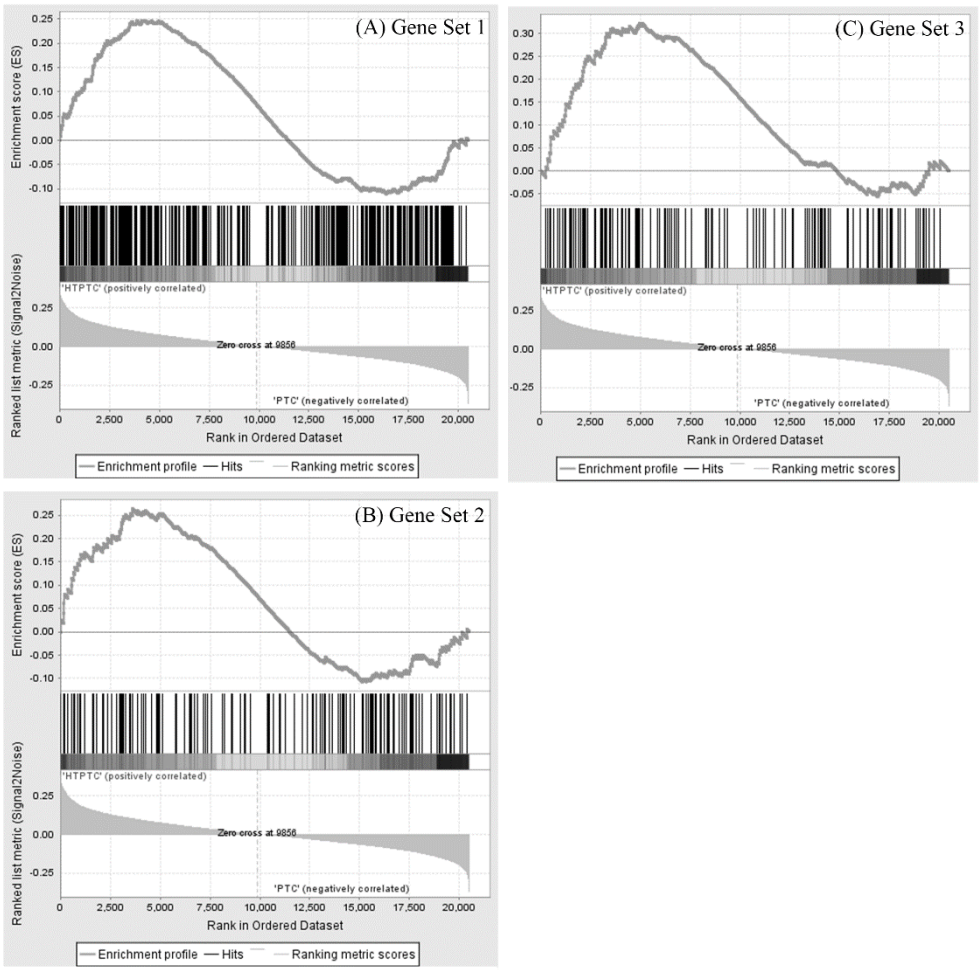

Fig 6. Enrichment plots of three ROS-related gene sets. (A) Gene Set 1, FDR *q*-value = 0.155. (B) Gene Set 2, FDR *q*-value = 0.252. (C) Gene Set 3, FDR *q*-value = 0.093. FDR *q*-value < 0.25 was considered to indicate statistical significance.

Discussion

Increased levels of ROS and oxidative stress are strongly associated with age-dependent disease, chronic inflammation, and increased risk of cancer development [23]. As discussed in a recent review article, many studies have shown that the loss of ROS balance is related to cancer, such as cancer of the breast, prostate, liver, lung, pancreas and so on. Furthermore, known cancer-related pathways, such as chemokine, c-myc, EGF, DNA repair mechanism, JAK/STAT, JNK, MAPK, mTOR, PI3K/Akt, p53, and PTEN, are linked with altered ROS status [24]. A recent clinical experimental study with radical prostatectomy tissue has shown that Nox-5 derived ROS exerts a carcinogenic effect by depletion of protein kinase C zeta and c-Jun N-terminal kinase (JNK) [25].

In the same context, oxidative stress and elevated ROS levels have been proposed to be a risk factor for the development of thyroid cancer [26-28]. Various experimental methods have been developed to detect the levels of ROS in specific tissues. However, these techniques tend to be extremely demanding for a variety of reasons. For instance, ROS have extremely short half-lives, generally on the order of nanoseconds to seconds. Moreover, their steady-state concentrations are in the picomolar to very low nanomolar range. ROS detection techniques also have demanding experimental requirements, such as fresh blood or tissue, expensive detection kits, and specialized techniques [29]. However, estimating the ROS level in thyroid or thyroid cancer tissue is more challenging than in other organs because thyroid cells themselves normally

produce substantial amounts of hydrogen peroxide. To protect itself from the harmful effects of hydrogen peroxide production, the thyroid gland contains many anti-ROS defense mechanisms, such as antioxidants and free radical-scavengers [16]. This active ROS-production and elimination system in normal thyroid tissue makes it difficult to conduct ROS-related experimental study in thyroid cancer or HT.

Several groups have developed promising methods for accurately quantifying the levels of ROS in thyroid tissue. Although these methods were shown to effectively estimate ROS levels in the thyroid, a biological or genomic relationship has not been clearly demonstrated between ROS and PTC using these methods [29]. Only a limited number of genomic studies focusing on ROS-related genes in thyroid cancer or HT have been performed. Moreover, the genomic studies that have been performed have focused on specific genes, including *DUOX1/2*, *NOX*, *TPO*, *TG*, *GSH* and *GPX* [30-34]. In 2013, Martinez et al. performed a purely bioinformatics-based analysis of TCGA data and found that the oxidative response pathway is frequently down-regulated in PTC via genomic and epigenetic mechanisms [35]. However, only one protein complex, the KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex, was considered in the activation of the pathway. Furthermore, samples were not grouped according to their clinical information such as thyroiditis.

According to our TCGA RNA-sequencing data analysis, a number of ROS-related genes were significantly up-regulated in the HT-PTC group. Compared with the PTC group, more genes were up-regulated in the HT-PTC group including those that showed higher fold changes. Among the up-regulated genes, *NCF1*, *CYBB*, *SOD2* and *NCF2* are dominantly expressed in immune

cells and its role in the ROS system is correlated with ROS production. In contrast, genes that are dominantly expressed in thyroid tissue, *TPO*, *IPCEF1*, *GPX3*, *SOD2*, *PRDX1*, *ETFDH*, *NDUFS1* and *PRDX3*, are correlated with ROS elimination (Table 1). From this result, we can infer that ROS from immune cells are increased, and that the protective mechanism from thyroid cells was enhanced against the increased ROS from immune cells in the HT-PTC microenvironment.

In the validation study from microarray data, part of the ROS-related DEGs from TCGA had the same expression pattern between the two PTC groups. Moreover, genes that showed higher fold change value over 1.0 in TCGA analysis, such as *TPO*, *CD38* and *IPCEF1*, have similarly large differential expressions between the two groups. Although some genes show a reverse expression direction in microarray dataset, they have no statistical significance in *t*-test. With this validation analysis, we can assert that up-regulated genes in HT-PTC from RNA sequencing have similar expression patterns in different analysis platforms such as microarray chip analysis.

Our study proposed to obtain genomic evidence of ROS elevation in PTC with HT by RNA sequencing and microarray, without resorting to difficult ROS detection experiments. Increased ROS from immune cells in HT status can destroy the normal thyroid tissue by its cytotoxic effect which can then lead to thyroid cancer development. In the GSEA result in Table 3, the HT-PTC group was more enriched than the PTC group in immune function-related pathways and one cancer related pathway, the “chemokine signaling pathway”, that contains cancer-related sub-pathways, such as the Jak-STAT signaling pathway, MAPK signaling pathway, and PI3K/Akt pathway. Moreover, ROS-related

Gene Sets 1, 2, and 3 were enriched toward the HT-PTC group. Increased enrichment of immune function related pathways may be derived from chronic inflammation under HT. However, chemokine signaling pathway enrichment could provide a molecular background for the carcinogenic effect of elevated ROS levels in PTC with HT. To date, there is yet no study that provides clear evidence of biologic mechanisms for prognostic advantage and low metastatic potential in PTC with HT patients [36]. Enhanced immune function may be able to suppress the metastatic potential of thyroid cancer, leading to more favorable clinical prognoses in patients with PTC with HT. However, this hypothesis requires solid verification because there is no published report of any particular immune response pathway that contributes to either progression or suppression of carcinogenesis.

This study has another meaning for clinical practitioners. Several clinical studies have reported that antioxidant therapy provides protective effects against various thyroid diseases, including improvements in chronic thyroiditis, amelioration of thyroid disease in pregnant women, and relief of ophthalmopathic symptoms in Graves' disease [37-39]. Our genomic background for ROS elevation in HT-PTC could support the potential benefits of antioxidant therapy in HT patients, not only in treating thyroiditis symptoms, but also in delaying or preventing the development of thyroid cancer.

Our study has some limitations. Although we could estimate the ROS level elevation in HT-PTC group from the gene expression study using RNA sequencing and microarray data, classic experimental validation such as quantitative RT-PTC is still necessary to provide biological confirmation of our findings. Also, there is no experimental or published evidence that the

differential expression of ROS-related genes is significantly correlated with actual ROS levels in thyroid tissue. To address these issues, future studies will need to focus on ROS level estimation as well as the expression levels of ROS-related genes.

From the oncologic standpoint, TCGA clinical data do not provide a clear description of the obligate order of PTC and HT. Therefore, we cannot safely conclude that the increased ROS levels in patients with longstanding HT contribute to the development of thyroid cancer because the obligate order in which these diseases develop in each patient has yet to be established. Prospective studies should be performed in patients with HT with sequencing or microarray before and after they are diagnosed with PTC. Such studies would enable a more meaningful evaluation of the real carcinogenic effect of longstanding HT.

The final limitation is the shortage of pathologic description in TCGA, especially for the tissue preparations used in the RNA-sequencing analyses. In the TCGA pathologic reports, which were provided with *.pdf files, the amounts of lymphocyte infiltration and the proportions of normal tissue in the PTC samples were not described. If some of the HT-PTC samples contained extremely large numbers of lymphocytes or only small amounts of cancer tissue, the expression of ROS-related genes could have been profoundly influenced by the presence of the lymphocytes, which themselves express high levels of ROS-related genes. This phenomenon could have significantly biased the RNA sequencing results. Similarly, all the PTC samples, from patients both with and without HT, contained remnants of normal thyroid tissue. Samples with higher proportions of normal tissue would presumably exhibit different gene

expression patterns compared with samples with higher proportions of cancer tissue. Ideally, such variations in pathologic factors would be eliminated before beginning the analysis. However, the TCGA pathologic reports did not contain sufficient information for this purpose. Thus, it would be advantageous for future RNA-sequencing studies to exclusively use quality-controlled PTC samples.

In summary, we obtained several lines of evidence that ROS levels are altered in HT-PTC using recent public next-generation sequencing and microarray data. First, more DEGs were observed in the HT-PTC group compared with the PTC group (33 versus 9), and the functions of the up-regulated genes in the HT-PTC group were associated with ROS production in immune cell-related genes and ROS elimination in thyroid cell related genes. Second, part of the ROS-related DEGs showed a similar expression pattern in HT in microarray data analysis. Third, the fold changes were higher in the HT-PTC group, as demonstrated by the difference in mean log fold change. Thirteen genes in the HT-PTC group had log fold changes greater than 0.5, whereas only one gene in the PTC group exhibited this degree of change. Fourth, immune function-related pathways and one cancer-related pathway were significantly associated with HT-PTC group only. Finally, the ROS-related gene sets were enriched only in the HT-PTC group. Taken together, ROS levels may be elevated to sufficiently high levels that it contributes to the progression of carcinogenesis in HT-PTC.

References

1. Lee JH, Kim Y, Choi JW, Kim YS (2013) The association between papillary thyroid carcinoma and histologically proven Hashimoto's thyroiditis: a meta-analysis. Eur J Endocrinol 168: 343-349.

2. Jankovic B, Le KT, Hershman JM (2013) Clinical Review: Hashimoto's thyroiditis and papillary thyroid carcinoma: is there a correlation? J Clin Endocrinol Metab 98: 474-482.

3. Dvorkin S, Robenshtok E, Hirsch D, Strenov Y, Shimon I, et al. (2013) Differentiated thyroid cancer is associated with less aggressive disease and better outcome in patients with coexisting Hashimotos thyroiditis. J Clin Endocrinol Metab 98: 2409-2414.

4. Zhang L, Li H, Ji QH, Zhu YX, Wang ZY, et al. (2012) The clinical features of papillary thyroid cancer in Hashimoto's thyroiditis patients from an area with a high prevalence of Hashimoto's disease. BMC Cancer 12: 610.

5. Yoon YH, Kim HJ, Lee JW, Kim JM, Koo BS (2012) The clinicopathologic differences in papillary thyroid carcinoma with or without co-existing chronic lymphocytic thyroiditis. Eur Arch Otorhinolaryngol 269: 1013-1017.

6. Paulson LM, Shindo ML, Schuff KG (2012) Role of chronic lymphocytic thyroiditis in central node metastasis of papillary thyroid carcinoma. Otolaryngol Head Neck Surg 147: 444-449.

7. Jeong JS, Kim HK, Lee CR, Park S, Park JH, et al. (2012) Coexistence of chronic lymphocytic thyroiditis with papillary thyroid carcinoma: clinical manifestation and prognostic outcome. J Korean Med Sci 27: 883-889.

8. Kim KW, Park YJ, Kim EH, Park SY, Park do J, et al. (2011) Elevated risk of papillary thyroid cancer in Korean patients with Hashimoto's thyroiditis. Head Neck 33: 691-695.

9. Ahn D, Heo SJ, Park JH, Kim JH, Sohn JH, et al. (2011) Clinical relationship between Hashimoto's thyroiditis and papillary thyroid cancer. Acta Oncol 50: 1228-1234.

10. Repplinger D, Bargren A, Zhang YW, Adler JT, Haymart M, et al. (2008) Is Hashimoto's thyroiditis a risk factor for papillary thyroid cancer? J Surg Res 150: 49-52.

11. Grani G, Calvanese A, Carbotta G, D'Alessandri M, Nesca A, et al. (2015) Thyroid autoimmunity and risk of malignancy in thyroid nodules submitted to fine-needle aspiration cytology. Head Neck 37: 260-264.

12. Anil C, Goksel S, Gursoy A (2010) Hashimoto's thyroiditis is not associated with increased risk of thyroid cancer in patients with thyroid nodules: a single-center prospective study. Thyroid 20: 601-606.

13. Del Rio P, Cataldo S, Sommaruga L, Concione L, Arcuri MF, et al. (2008) The association between papillary carcinoma and chronic lymphocytic thyroiditis: does it modify the prognosis of cancer? Minerva Endocrinol 33: 1-5.

14. Akinci M, Kosova F, Cetin B, Sepici A, Altan N, et al. (2008) Oxidant/antioxidant balance in patients with thyroid cancer. Acta Cir Bras 23: 551-554.

15. Sadani GR, Nadkarni GD (1996) Role of tissue antioxidant defence in thyroid cancers. Cancer Lett 109: 231-235.

16. Karbownik-Lewinska M, Kokoszko-Bilska A (2012) Oxidative damage to macromolecules in the thyroid -experimental evidence. Thyroid Res 5: 25.

17. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, et al. (2009) AmiGO: online access to ontology and annotation data. Bioinformatics 25: 288-289.

18. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20: 307-315.

19. Team RC (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

20. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.

21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genomewide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550.

22. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30.

23. Brieger K, Schiavone S, Miller FJ, Jr., Krause KH (2012) Reactive oxygen species: from health to disease. Swiss Med Wkly 142: w13659.

24. Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB (2010) Oxidative stress, inflammation, and cancer: how are they linked? Free Radic Biol Med 49: 1603-1616.

25. Holl M, Koziel R, Schafer G, Pircher H, Pauck A, et al. (2015) ROS signaling by NADPH oxidase 5 modulates the proliferation and survival of prostate carcinoma cells. Mol Carcinog doi: 10.1002/mc.22255.

26. Xing M (2012) Oxidative stress: a new risk factor for thyroid cancer. Endocr Relat Cancer 19: C7-11.

27. Wang D, Feng JF, Zeng P, Yang YH, Luo J, et al. (2011) Total oxidant/antioxidant status in sera of patients with thyroid cancers. Endocr Relat Cancer 18: 773-782.

28. Metere A, Chiesa C, Di Cosimo C, Fierro G, Giacomelli L, et al. (2012) A novel approach to study oxidative stress in thyroid diseases: a preliminary study. Eur Rev Med Pharmacol Sci 16: 646-652.

29. Dikalov SI, Harrison DG (2014) Methods for detection of mitochondrial and cellular reactive oxygen species. Antioxid Redox Signal 20: 372-382.

30. Weyemi U, Caillou B, Talbot M, Ameziane-El-Hassani R, Lacroix L, et al. (2010) Intracellular expression of reactive oxygen species-generating NADPH oxidase NOX4 in normal and cancer thyroid tissues. Endocr Relat Cancer 17: 27-37.

31. Song Y, Ruf J, Lothaire P, Dequanter D, Andry G, et al. (2010) Association of duoxes with thyroid peroxidase and its regulation in thyrocytes. J Clin Endocrinol Metab 95: 375-382.

32. Ohye H, Sugawara M (2010) Dual oxidase, hydrogen peroxide and thyroid diseases. Exp Biol Med (Maywood) 235: 424-433.

33. Rigutto S, Hoste C, Grasberger H, Milenkovic M, Communi D, et al. (2009) Activation of dual oxidases Duox1 and Duox2: differential regulation mediated

by camp-dependent protein kinase and protein kinase C-dependent phosphorylation. J Biol Chem 284: 6725-6734.

34. Rostami R, Aghasi MR, Mohammadi A, Nourooz-Zadeh J (2013) Enhanced oxidative stress in Hashimoto's thyroiditis: inter-relationships to biomarkers of thyroid function. Clin Biochem 46: 308-312.

35. Martinez VD, Vucic EA, Pikor LA, Thu KL, Hubaux R, et al. (2013) Frequent concerted genetic mechanisms disrupt multiple components of the NRF2 inhibitor KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex in thyroid cancer. Mol Cancer 12: 124.

36. Lun Y, Wu X, Xia Q, Han Y, Zhang X, et al. (2013) Hashimoto's thyroiditis as a risk factor of papillary thyroid cancer may improve cancer prognosis. Otolaryngol Head Neck Surg 148: 396-402.

37. Dharmasena A (2014) Selenium supplementation in thyroid associated ophthalmopathy: an update. Int J Ophthalmol 7: 365-375.

38. Drutel A, Archambeaud F, Caron P (2013) Selenium and the thyroid gland: more good news for clinicians. Clin Endocrinol (Oxf) 78: 155-164.

39. Kohrle J, Gartner R (2009) Selenium and thyroid. Best Pract Res Clin Endocrinol Metab 23: 815-827.

요약 (국문초록)

서론: 하시모토 갑상선염 환자에서, 갑상선 내부의 활성 산소 (Reactive oxygen species, ROS)의 증가는 갑상선 유두암 (Papillary thyroid carcinoma, PTC) 을 일으키는 위험인자로 제기되어 왔다. 하지만, 활성산소의 총량이 발암 기전에 기여할 만큼 충분히 증가되는지에 대해서는 밝혀진 바가 없다. 우리는 만약 하시모토 갑상선염에서 활성산소가 증가한다면, 하시모토 갑상선염이 동반된 갑상선암 조직에서는 활성산소와 관련된 유전자들이 다른 양상으로 발현되어 있을 것이라고 가정하였다.

방법: 차별발현 유전자 (Differentially Expressed Genes, DEG)를 찾기 위해, 유전체 지도 (The Cancer Genomic Atlas, TCGA) 데이터 중 RNA 서열분석에서 계산된 유전자 발현 데이터를 분석하였다. 33개의 정상 갑상선조직, 232개의 하시모토 갑상선염이 동반되지 않은 갑상선 유두암조직, 60 개의 하시모토 갑상선염이 동반된 갑상선 유두암 조직이 분석에 포함되었다. 활성 산소와 관련된 402개의 유전자들이 데이터베이스 검색을 통해 준비되었다. 공개된 마이크로어레이 데이터를 분석하여 TCGA 분석 결과를 뒷받침하였다.

결과: 33개의 활성산소 관련 유전자들이 하시모토 갑상선염이 있는 갑상선 유두암 조직에서 증가된 발현을 보였고, 9개의 유전자가 하시모토 갑상선염이 없는 갑상선 유두암조직에서 증가된 발현을 보였다. 평균 로그 변환 발현량 차이는 0.562 대 0.252 였다. 마이크로어레이 분석에서는, 32개의 활성산소 관련 유전자 중 12개의 유전자가 TCGA 분석 결과와 같은 양상의 차별 발현 양상을 가지고 있었다. 증가된 활성산소 관련 유전자들은 활성산소 대사 및 세포자멸사와 관련되어 있었고, 경로 분석에서는 면역 기능 및 발암과 관련된 유전자 세트의 발현이 하시모토 갑상선염이 동반된 갑상선암

조직에서 증가되어 있었다.

결론: 하시모토 갑상선염이 동반된 갑상선 유두암에서는 활성 산소의 총량이 증가되어 있을 것이고, 증가된 활성산소는 하시모토 갑상선염 환자에서 갑상선 유두암 발병 인자로 작용할 수 있을 것이다.

주요어: 활성산소, 갑상선암, 하시모토 갑상선염, 유전자 발현, 바이오인포매틱스

학 번: 2011-23753