

# The Association of Long Noncoding RNA LOC105372577 with Endoplasmic Reticulum Protein 29 Expression: A Genome-wide Association Study

Soyeon Lee<sup>1</sup>, Kiang Kwon<sup>2</sup>, Younghwa Ko<sup>3</sup> and O-Yu Kwon<sup>3\*</sup>

<sup>1</sup>School of Systems Biomedical Science, College of Natural Sciences, Soongsil University, Seoul 06978, Korea

<sup>2</sup>Department of Clinical Laboratory Science, Wonkwang Health Science University, Iksan 54538, Korea

<sup>3</sup>Department of Anatomy & Cell Biology, College of Medicine, Chungnam National University, Daejeon 35015, Korea

Received April 27, 2021 / Revised April 29, 2021 / Accepted April 30, 2021

This study identified genomic factors associated with endoplasmic reticulum protein (ERp)29 gene expression in a genome-wide association study (GWAS) of genetic variants, including single-nucleotide polymorphisms (SNPs). In total, 373 European genes from the 1000 Genomes Project were analyzed. SNPs with an allelic frequency of less than or more than 5% were removed, resulting in 5,913,563 SNPs including in the analysis. The following expression quantitative trait loci (eQTL) from the long noncoding RNA LOC105372577 were strongly associated with ERp29 expression: rs6138266 ( $p < 4.172 \times 10^{-10}$ ), rs62193420 ( $p < 1.173 \times 10^{-10}$ ), and rs6138267 ( $p < 2.041 \times 10^{-10}$ ). These were strongly expressed in the testis and in the brain. The three eQTL were identified through a transcriptome-wide association study (TWAS) and showed a significant association with ERp29 and osteosarcoma amplified 9 (OS9) expression. Upstream sequences of rs6138266 were recognized by ChIP-seq data, while HaploReg was used to demonstrate how its regulatory DNA binds upstream of transcription factor 1 (USF1). There were no changes in the expression of OS9 or USF1 following ER stress.

**Key words** : Endoplasmic reticulum protein (ERp)29, expression quantitative trait loci (eQTL), genome-wide association study (GWAS), osteosarcoma amplified 9 (OS9), single-nucleotide polymorphisms (SNPs)

## Introduction

The Human Genome Project, which was completed in 2003, provided a plethora of genomic information. However, it remains difficult to experimentally determine which genes are involved in the myriad of genetic traits using only genomic information. To determine the biological association between genetic and natural mutations, a high-throughput genome analysis is required to analyze the exact sequence of the entire genome region. In other words, the position and allele of single-nucleotide polymorphisms (SNPs), such as insertions or deletions, related to specific genes need to be determined [13]. It is also necessary to study the correlation between specific gene regions and the phenotype of interest.

The genome-wide association studies (GWAS) technique, which analyzes genomes associated with specific diseases or traits, was first attempted by the Ozaki team in 2002[16]. Since GWAS analyzes the degree of association for all gene locations, a similar approach can be to analyze the structural variation of 50 base pairs (bp) or more of the noncoding genome and perform base sequencing as a screening method to find candidate genes that are primarily related to traits or diseases of interest. Therefore, even if a gene was found to associate with a disease or trait through GWAS, it may not necessarily be the causative gene. GWAS is not a causal search but a process of finding candidates for associated genes. The scientific justification of GWAS is based on the hypothesis of “common disease, a common variable” [5]. However, complex diseases are often not a one gene-one disease mechanism but are influenced by many genetic and environmental factors. Indeed, identifying a common variable at a frequency of approximately 1%-5% is typical of a phenotypic-genetic linkage in diseases such as diabetes and hypertension [12]. In 2010, two separate GWAS for type 2 diabetes demonstrated twelve and two biomarkers, respectively [22, 23]. Since gene recombination occurs *en masse* in germ cells, gene locational stents that are close to each other do not

---

### \*Corresponding author

Tel : +82-42-580-8206, Fax : +82-42-586-4800

E-mail : oykwon@cnu.ac.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

mix genotypes and therefore move together in mosaic patterns, resulting in linkage disequilibrium (LD) [1]. These instances are referred to as a LD block. For regions within the same LD block, a p-value for the correlation among expression quantitative trait loci (eQTL) are usually calculated [11]. Results from eQTL analyses are represented using the Manhattan plot; when the SNP density is high, so are the p-values for SNPs in the genetic region related to the trait of interest. However, since correlations do not imply causations, follow-up studies are required to gain mechanistic biological insight.

The human endoplasmic reticulum protein (ERp) 29 gene encodes a protein consisting of 261 amino acids (MW 2.8993 kDa) that is localized to the lumen of the ER. The first ERp29 was cloned in 1997[7], and a stress-inducible ERp29 was reported a year later [14]. The genomic organization and crystal structure of the ERp29 protein were later reported in 2002 and 2009[2, 19], respectively. Many studies have demonstrated the association of ERp29 to cancer, as well as to thio-redoxin or thyrocyte secretion [10, 15, 17, 20]. Although many studies have been conducted on ERp29, it remains necessary to study the biological function of the ERp29 gene and protein to fully understand its physiological function. It will provide a new clue to the understanding of ERp29 function that identifying a candidate gene closely associated to ERp29 gene expression in human genome. This study investigated the chromosomal loci associated with ERp29 gene regulation through GWAS and identified a transcriptome-wide associated genes using total of 373 European genes were used in Utah (n=91), Finns (n=95), British (n=94), and Toscani (n=93).

## Materials and Methods

To screen for human genome eQTLs, ERp29 gene expression data generated through the Geuvadis RNA-sequencing project (<https://cordis.europa.eu/project/id/261123/> reporting) were used. The total number of transcripts of 373 European genes was calculated through the sum of reads per kilobase of transcript per million mapped reads (RPKM). Genotypic data were obtained from the 1000 Genomes Project (<http://www.1000genomes.org/>). Data with a minor allele frequency of less than 5% or a missing rate of more than 5% were removed, resulting in 5,913,563 SNPs that were used for the final analysis.

RPKM of a gene =

$$\frac{\text{Number of reads mapped to gene} \times 10^2 \times 10^6}{\text{Total number of mapped reads from given library} \times \text{gene length in bp}}$$

To identify ERp29 eQTLs, single marker regression was performed following a  $-\text{Log}_{10}(5 \times 10^{-8})$  transformation, and PLINK was used for statistical analysis (<http://pngu.mgh.harvard.edu/purcell/plink/>) [18]. LD blocks were analyzed using HaploView (<https://www.broadinstitute.org/haploview/downloads>) [3]. The transcription factor binding sites of eQTLs were identified using ChIP-seq data and subsequent analyses using HaploReg v4.1 of regulome DB (<https://www.encodeproject.org/software/regulomedb/> and <https://pubs.broadinstitute.org/mammals/hapls>) [6]. The transcriptome-wide association study (TWAS) of identified eQTLs was transformed to  $-\text{Log}_{10}(2 \times 10^{-5})$ , and the correlation between 9,981 genes and SNPs related to ERp29 expression was analyzed.

Each gene expression was mainly determined by RT-PCR as described below. RTPCR conditions included 30 cycles, each comprising of the following: 94°C for 30 sec, 58°C for 30 sec and 72°C for 1 min (10 min in the final cycle) using primers with *Taq* DNA polymerase (Solgent Co., Ltd. Daejeon, Korea). The RTPCR primers were supplied by Bioneer Corporation (Daejeon, Korea). All chemicals were purchased from SigmaAldrich; Merck KGaA. The RTPCR primers used in this study are as follows: ERp29 (5'-AAA GCA AGT TCG TCT TGG TGA-3', 5'-CGC CAT AGT CTG AGA TCC CCA-3'), OS9 (5'-CGC CAT ATC CAG CAG TAC CA-3' and 5'-CTT CTC TGG GCT TCC CAT TG-3'), USF1 (5'-AAA ACA GCC GAA ACC GAA GA-3' and 5'-TAG CCA CTG ATA GCG CCA GA-3'), and GAPDH (5'-ACA TCA AAT GGG GTG ATG CT-3' and 5'-AGG AGA CAA CCT GGT CCT CA-3').

## Results and Discussion

The ER molecular chaperone ERp29 has been widely studied at the molecular level since it was first reported in 1997 [19]. Specifically, studies that associate ERp29 with cancer and its function as a molecular chaperone has made great strides. However, to progress the clinical application and development of new drug treatments, studies related to ERp29 gene regulation are required. The ERp29 gene is located at chr12:112,013,340-112,023,449. In this study, we conducted a GWAS to determine SNPs that affect ERp29 expression.

As shown in Fig. 1A, eQTL analysis of ERp29 gene expression revealed that there were three SNPs on chromosome 20 (rs6138266, rs62193420, and rs6138267). A LD map of the three SNPs is presented in Fig. 1B, which was obtained from the 1000 Genomes Project. A high similarity (97% - 98%) was observed in the LD patterns of the three identified SNPs (<http://genome.ucsc.edu>). These three SNPs were identified as intron variants of an uncharacterized gene (20p11.21) encoding a long noncoding (lnc) RNA LOC 105372577, which shows the strongest expression in the testis, followed by the brain (Fig. 1C). A long noncoding RNA (lncRNA) is a noncoding RNA (ncRNA) with a length of 200 nucleotides and does not translate into a protein. lncRNAs are abundant in mammalian transcripts and have been reported to greatly associate with diseases such as cancer and Alzheimer's. Considering the studies relating ERp29 to carcinogenesis, we were intrigued with the potential role

of LOC105372577 in this process. All association signals with  $-\text{Log}_{10}(5 \times 10^{-8})$  (three SNPs) are summarized in Table 1A.

The recent trend of gene discovery involves TWAS, which maps genomic information to a large scale in functional units, and is a useful method for predicting the expression of SNP genes obtained from GWAS. As shown in Table 1B, TWAS demonstrated that the three identified eQTLs (rs6138266, rs62193420, and rs6138267) are significantly associated with ERp29 and osteosarcoma amplified 9 (OS9) expression [4]. In particular, the association between rs6138266 and ERp29 has a p-value of  $4.2 \times 10^{-9}$  while that of rs6138266 and OS9 has a p-value of  $7.1 \times 10^{-6}$ . OS9 is a lectin protein that is deeply involved in ER quality control and ER-associated degradation of glycoproteins in the ER lumen [9, 21]. In this respect, it is thought to have a functional relationship with ERp29 since ERp29 also plays a role in the folding and assembly of secretory proteins in ER lumen.

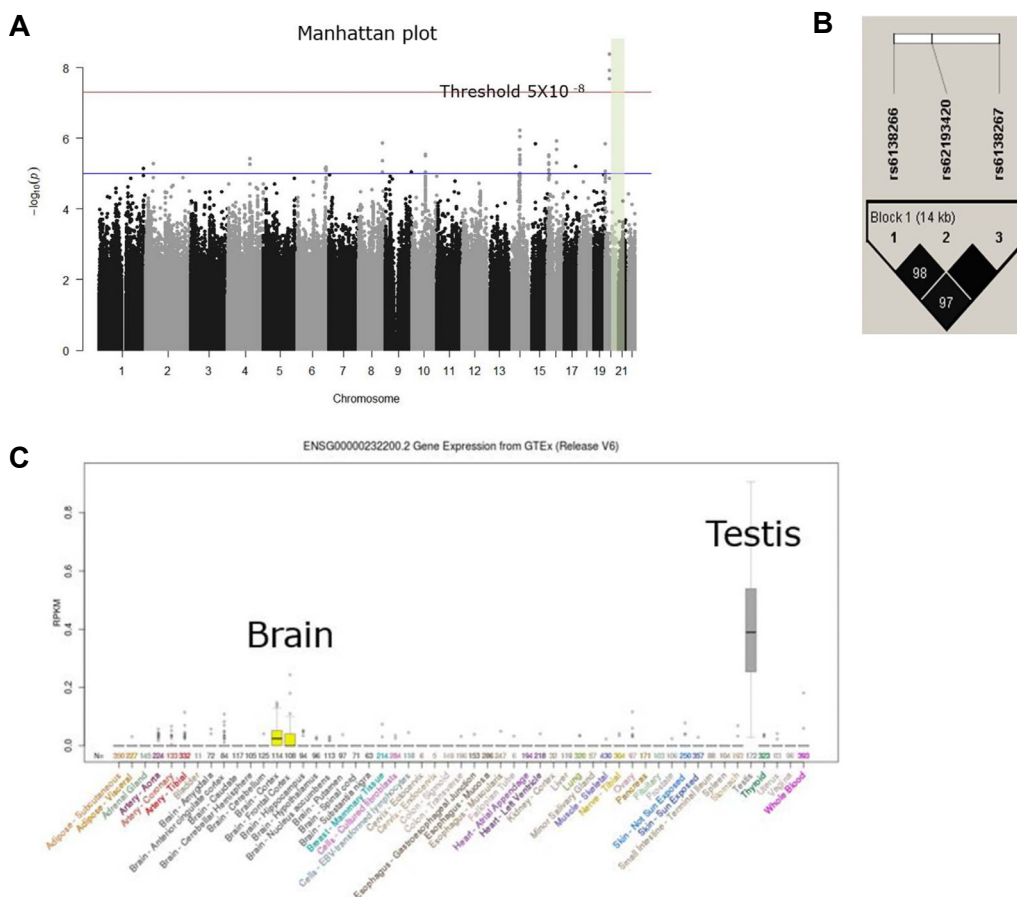


Fig. 1. Expression quantitative trait loci (eQTL) analysis of ERp29 gene expression by genome-wide association studies (GWAS). (A) Manhattan plot for genome-wide association study between eQTLs and mRNA expression of ERp29. The red line indicates a genome-wide significance threshold ( $P=5 \times 10^{-8}$ ), and the blue line is suggestive line ( $P=10^{-5}$ ). LD block and haplotype structure of the SORBS2 SNPs in this stud. (B) Linkage disequilibrium blocks for signals identified to have an association with the expression of ERp29. (C) mRNA expression of LOC105372577(RPKM).

Table 1. GWAS and TWAS of genes with eQTLs identified for ERp29

A. All 3 SNPs with  $-\text{Log}_{10}(5 \times 10^{-8})$  according to GWAS analysis

SNP	Position	Gene (NCBI dbSNP)	Allele (minor/major)	MAF	BETA	P
rs6138266	Chr20: 24,299,534	LOC105372577 : Intron Variant	A, C / G	A=0.0825	9.354	$4.172 \times 10^{-9}$
rs62193420	Chr20: 24,304,884		A / G	A=0.0829	9.126	$1.173 \times 10^{-8}$
rs6138267	Chr20: 24,314,057		A / G	A=0.0877	8.945	$2.041 \times 10^{-8}$

B. TWAS of genes with eQTLs identified for ERp29

SNP	CHR	BP	A1	TEST	NMISS	BETA	STAT	P	Gene
20	rs6138266	24299534	A	ADD	373	5.845	4.556	$7.1 \times 10^{-6}$	OS9
						7.054	4.336	$1.9 \times 10^{-5}$	MRPL51
						86.05	4.160	$4E \times 10^{-5}$	TMSB10
						1.315	4.113	$4.8 \times 10^{-5}$	ALKBH2
						9.354	6.021	$4.2 \times 10^{-9}$	ERP29
20	rs6138267	24314057	A	ADD	373	5.585	4.343	$1.8 \times 10^{-5}$	OS9
						86.78	4.197	$3.4 \times 10^{-5}$	TMSB10
						1.337	4.182	$3.6 \times 10^{-5}$	ALKBH2
						6.789	4.165	$3.9 \times 10^{-5}$	MRPL51
						0.347	4.106	$5.0 \times 10^{-5}$	AIFM3
8.945	5.733	$2.0 \times 10^{-8}$	ERP29						
20	rs62193420	24304884	A	ADD	373	5.663	4.388	$1.5 \times 10^{-5}$	OS9
						7.057	4.320	$2.0 \times 10^{-5}$	MRPL51
						1.363	4.251	$2.7 \times 10^{-5}$	ALKBH2
						86.70	4.175	$3.7 \times 10^{-5}$	TMSB10
						0.7684	4.172	$3.8 \times 10^{-5}$	CCDC30
9.126	5.835	$1.2 \times 10^{-8}$	ERP29						

SNP, single-nucleotide polymorphism; Chromosome positions are based on NCBI build 37. Gene is defined as the gene containing the SNP or the closest genes up to 100 kb up/downstream of the SNP. MAF, minor allele frequency; BETA, effect size estimate, TWAS, transcriptome-wide association studies; eQTLs, expression quantitative trait loci; p-value, as would be calculated by any linear regression software.

Next, we investigated the candidate proteins that bind to the enhancer and promoter regions of ERp29 and can regulate the three SNPs. While no candidate was found at both rs6138267 and rs62193420, upstream transcription factor 1 (USF1) was found to bind to rs6138266. USF1 is a transcription factor with bHLH-ZIP domains that play an important role in DNA binding [8]. Although the three SNPs were located in the same LD, thus sharing the same enhancer, USF1 was found to be a trans-promoter since it only bound to rs6138266. Furthermore, we investigated how ERp29, OS9, and USF1 responded to ER stresses such as  $\text{Ca}^{++}$ -ionophore (A23187) and tunicamycin (blocking N-linked glycosylation) (Fig. 2). While both BiP and ERp29 were transcriptionally upregulated by ER stress, OS9 and USF1 did not exhibit significant upregulation. Although GWAS analysis showed a close relationship between ERp29 and OS9, these genes did not have the same response to

ER stress.

In summary, three LDs (rs6138266, rs62193420, and rs6138267) associated with ERp29 expression were identified

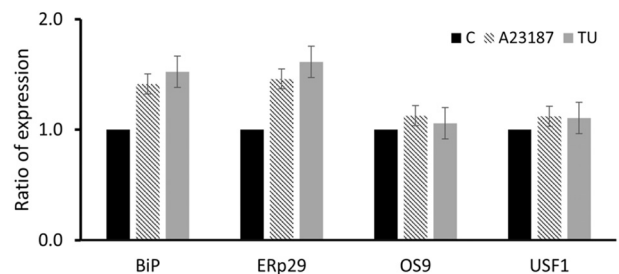


Fig. 2. Gene expression of OS9 and USF1 by ER stress. PC12 cells were treated with tunicamycin (2  $\mu\text{g}/\text{ml}$ ) and 10  $\mu\text{M}$   $\text{Ca}^{++}$ -ionophore (A23187) for 24 hr, followed by RT-PCR. DNA bands were quantified using the ImageJ program (NIH, Bethesda, MD, USA). Experiments were performed in triplicate and the results represent an average.

in the intron of lncRNA LOC105372577. TWAS analysis further demonstrated that these three LDs commonly associate with ERp29 and OS9. These results support the functional similarity between ERp29 and OS9 with regards to the ER signaling pathway. Our findings suggest that ERp29 expression can be significantly regulated by the expression of an uncharacterized lncRNA as yet a gene.

### Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1F1A1049836).

### The Conflict of Interest Statement

The authors declare that they have no conflicts of interest with the contents of this article.

### References

- Aissani, B. 2014. Confounding by linkage disequilibrium. *J. Hum. Genet.* **59**, 110-115.
- Barak, N. N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Miroslav, M., Heike Reichardt, Fischer, G., Stubbs, M. T. and Ferrari, D. D. 2009. Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29. *J. Mol. Biol.* **385**, 1630-1642.
- Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. 2005. Haplotype: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.
- Bernasconi, R., Pertel, T., Luban, J. and Molinari, M. 2008. A dual task for the Xbp1-responsive OS-9 variants in the mammalian endoplasmic reticulum: Inhibiting secretion of misfolded protein conformers and enhancing their disposal. *J. Biol. Chem.* **283**, 16446-16454.
- Bogaert, D. J., Dullaers, M., Lambrecht, B. N., Vermaelen, K. Y., De Baere, E. and Haerynck, F. 2016. Genes associated with common variable immunodeficiency: One diagnosis to rule them all? *J. Med. Genet.* **53**, 575-590.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., Cherry, J. M. and Snyder, M. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790-1797.
- Demmer, J., Zhou, C. and Hubbard, M. J. 1997. Molecular cloning of ERp29, a novel and widely expressed resident of the endoplasmic reticulum. *FEBS Lett.* **402**, 145-150.
- Fan, Y. M., Hernesniemi, J., Oksala, N., Levula, M., Raitoharju, E., Collings, A., Hutri-Kähönen, N., Juonala, M., Marniemi, J., Lyytikäinen, L-P., Seppälä, I., Mennander, A., Tarkka, M., Kangas, A. J., Soininen, P., Salenius, J. P., Klopp, N., Illig, T., Tomi, T., Ala-Korpela, M., Laaksonen, R., Viikari, J., Kähönen, M., Raitakari, O. T. and Lehtimäki, T. 2014. Upstream Transcription Factor 1 (USF1) allelic variants regulate lipoprotein metabolism in women and USF1 expression in atherosclerotic plaque. *Sci. Rep.* **4**, 4650.
- Hosokawa, N. 2011. OS-9 and XTP3-B: Lectins that regulate endoplasmic reticulum-associated degradation (ERAD). *Seikagaku* **83**, 26-31.
- Kwon, O. Y., Park, S., Lee, W., You, K. H., Kim, H. and Shong, M. 2000. TSH regulates a gene expression encoding ERp29, an endoplasmic reticulum stress protein, in the thyrocytes of FRTL-5 cells. *FEBS Lett.* **475**, 27-30.
- Li, Q., Seo, J. H., Stranger, B., McKenna, A., Pe'er, I., Laframboise, T., Brown, M., Tyekucheva, S. and Freedman, M. L. 2013. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633-641.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A. and Hunter, D. J., et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.
- Matsuda, K. 2017. PCR-based detection methods for single-nucleotide polymorphism or mutation: real-time PCR and its substantial contribution toward technological refinement. *Adv. Clin. Chem.* **80**, 45-72.
- Mkrtchian, S., Fang, C., Hellman, U. and Ingelman-Sundberg, M. 1998. A stress-inducible rat liver endoplasmic reticulum protein, ERp29. *Eur. J. Biochem.* **251**, 304-313.
- Mkrtchian, S. and Sandalova, T. 2006. ERp29, an unusual redox-inactive member of the thioredoxin family. *Antioxid. Redox Signal.* **8**, 325-337.
- Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y. and Tanaka, T. 2002. Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650-654.
- Park, S., You, K. H., Shong, M., Goo, T. W., Yun, E. Y., Kang, S. W. and Kwon, O. Y. 2005. Overexpression of ERp29 in the thyrocytes of FRTL-5 cells. *Mol. Biol. Rep.* **32**, 7-13.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Daly, M. J. and Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575.
- Sargsyan, E., Baryshev, M., Backlund, M., Sharipo, A. and Mkrtchian, S. 2002. Genomic organization and promoter characterization of the gene encoding a putative endoplasmic reticulum chaperone, ERp29. *Gene* **285**, 127-139.
- Sargsyan, E., Baryshev, M., Szekely, L., Sharipo, A. and Mkrtchian, S. 2002. Identification of ERp29, an endoplasmic reticulum luminal protein, as a new member of the thyroglobulin folding complex. *J. Biol. Chem.* **277**, 17009-17015.
- Seayfan, E., Defontaine, N., Demaretz, S., Zaarour, N. and Laghmani, K. 2016. OS9 protein interacts with Na-K-2Cl Co-transporter (NKCC2) and targets its immature form for the endoplasmic reticulum-associated degradation pathway. *J. Biol. Chem.* **291**, 4487-4502.
- Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P.,

Dina, C., Welch, R. P., Zeggini, E. and Huth, C., *et al.* 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579-589.

23. Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi,

A., Horikoshi, M., Nakamura, M. and Fujita, H., *et al.* 2010. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat. Genet.* **42**, 864-868.

## 초록 : ERp29 유전자 발현과 관련된 long noncoding RNA LOC105372577의 전장 유전체 연관성 분석

이소연<sup>1</sup> · 권기상<sup>2</sup> · 고영화<sup>3</sup> · 권오유<sup>3\*</sup>

(<sup>1</sup>승실대학교 자연과학대학 의생명시스템학부, <sup>2</sup>원광보건대학교 임상병리학과, <sup>3</sup>충남대학교 의과대학 해부학교실)

본 연구는 전장 유전체 연관성 분석(genome-wide association study, GWAS)을 통해 ERp29의 mRNA 발현과 관련된 유전자좌위(expression quantitative trait loci, eQTL)을 식별하는 것을 목표로 하였다. 대상 유전자는 ERp29이다. ERp29는 소포체(ER)의 lumen에 단백질의 folding & assembly 기능을 가진 분자 chaperone 단백질로서 소포체 스트레스에 의해 발현량이 증가하며, 분비 단백질의 생합성에 관여한다. 최근 연구 결과 발암과 연관성이 알려지면서 주목을 받고 있다. 총 373명의 유럽인의 genome을 대상으로 GWAS 분석 결과, ERp29 유전자 발현은 정소와 뇌에서 강하게 발현하는 long noncoding RNA (LncRNA) LOC105372577과 관계가 있었다. 즉, 3개의 eQTL: rs6138266 ( $p < 4.172 \times 10^{-9}$ ), rs62193420 ( $p < 1.173 \times 10^{-8}$ ), rs6138267 ( $p < 2.041 \times 10^{-8}$ )와 연관성이 깊은 것으로 밝혀졌다. ERp29의 발현과 연관이 있는 것으로 확인된 3개의 eQTL을 사용한 transcriptome-wide association study (TWAS) 결과 osteosarcoma amplified 9 (OS9) 발현과 유의한 연관성을 보이며 OS9 유전자의 up-stream에 upstream of transcription factor 1 (USF1)이 결합할 수 있는 것을 알았다.