Observational Study

Medicine®

OPEN

Expression quantitative trait loci for PI3K/AKT pathway

Dongchan Ryu, BS, Chaeyoung Lee, PhD*

Abstract

A genome-wide association study (GWAS) was conducted to identify expression quantitative trait loci (eQTLs) for the genes involved in phosphatidylinositol-3-kinase/v-akt murine thymoma viral oncogene homolog (PI3K/AKT) pathway.

Data on mRNA expression of 341 genes in lymphoblastoid cell lines of 373 Europeans recruited by the 1000 Genomes Project using Illumina HiSeq2000 were utilized. We used their genotypes at 5,941,815 nucleotide variants obtained by Genome Analyzer II and SOLiD.

The association analysis revealed 4166 nucleotide variants associated with expression of 85 genes ($P<5\times10^{-8}$). A total of 73 eQTLs were identified as association signals for the expression of multiple genes. They included 9 eQTLs for both of the genes encoding *collagen type I alpha 1* (*COL1A1*) and *integrin alpha 11* (*ITGA11*), which synthesize a major complex of plasma membrane. They also included eQTLs for type IV collagen molecules; 13 eQTLs for both *collagen type IV alpha 1* (*COL4A1*) and *collagen type IV alpha 2* (*COL4A2*) and 18 eQTLs for both *collagen type IV alpha 5* (*COL4A5*) and *collagen type IV alpha 6* (*COL4A6*). Some genes expressed by the eQTLs might induce expression of the genes encoding type IV collagen. One eQTL (rs16871986) was located in the promoter of *palladin* (*PALLD*) gene which might synthesize collagen by activating fibroblasts through the PI3K/AKT pathway. Another eQTL (rs34845474) was located in an enhancer of *cadherin related family member 3* (*CDHR3*) gene which can mediate cell adhesion.

This study showed a profile of eQTLs for the genes involved in the PI3K/AKT pathway using a healthy population, revealing 73 eQTLs associated with expression of multiple genes. They might be candidates of common variants in predicting genetic susceptibility to cancer and in targeting cancer therapy. Further studies are required to examine their underlying mechanisms for regulating expression of the genes.

Abbreviations: ChIA-PET = chromatin interaction analysis with paired-end-tag sequencing, eQTL = expression quantitative trait locus, PI3K/AKT = phosphatidylinositol-3-kinase/v-akt murine thymoma viral oncogene homolog, SNP = single nucleotide polymorphism.

Keywords: collagen, expression quantitative trait locus, genome-wide association study, PI3K/AKT pathway, single nucleotide variant

1. Introduction

Phosphatidylinositol-3-kinase/v-akt murine thymoma viral oncogene homolog (PI3K/AKT) pathway plays an essential role in cellular functions including cell growth, cell survival, cell proliferation, and glucose metabolism.[1–3] Many components of the pathway which consists of multiple kinase cascades are implicated in a variety of types of human cancers and have greatly been concerned as attractive pharmacological targets.[4–6] Everolimus and temsirolimus as inhibitors of *mammalian target of rapamycin* (*mTOR*) have been approved for cancer treatment, and some inhibitors of its components such as *PI3K*, *AKT*, and/or *mTOR* (eg, buparlisib as a pan-*PI3K* inhibitor and alpelisib as a *PI3Kα*-selective inhibitor) are currently in clinical development.[7,8] Many studies on the PI3K/AKT pathway have focused on targeting cancer therapy, and its genetic studies were mostly to examine mutations and its contribution to tumorigenesis and tumor maintenance by hyper-activating the PI3K/AKT pathway.[9] Somatic mutations of *phosphatidylinositol-3-kinase, catalytic, polypeptide* (*PIK3CA*) have been frequently observed within its kinase domain (H1047R in exon 20) and helical domain (E542K and E545K in exon 9) in a variety of cancers.[10] Germline mutations in exons of *phosphatase and tensin homolog* (*PTEN*) are well known for causes of hamartoma tumor syndromes.[1,11,12]

Moreover, common genetic variants in *PTEN*, *mTOR*, *AKT1*, and *AKT2* were found to be associated with clinical outcomes such as response to therapy, survival, and recurrence in esophageal cancer patients treated with chemoradiotherapy ($P<0.05$).[13] However, common genetic variants for regulating the pathway have been rarely investigated in healthy populations. Associations of common genetic variants with phosphorylation of *AKT1* or *RPS6KB1* were identified in lymphoblastoid cell lines of Centre d'Etude du Polymorphisme Humain ($6.35\times10^{-5}<P<4.20\times10^{-3}$).[14] In this study, we conducted a genome-wide association study to examine genetic associations of common variants with mRNA expression of the genes involved in the

Editor: Mirko Manchia.

Funding/support: This work was funded by the National Research Foundation of Korea, the Ministry of Education, Science, and Technology (Grant No. NRF-2012M3A9D1054705).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Department of Bioinformatics and Life Science, Soongsil University, Dongjak-gu, Seoul, Korea.

* Correspondence: Professor Chaeyoung Lee, Department of Bioinformatics and Life Science, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea (e-mail: clee@ssu.ac.kr).

Copyright © 2017 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medicine (2017) 96:1(e5817)

Received: 11 July 2016 / Received in final form: 8 December 2016 / Accepted: 13 December 2016

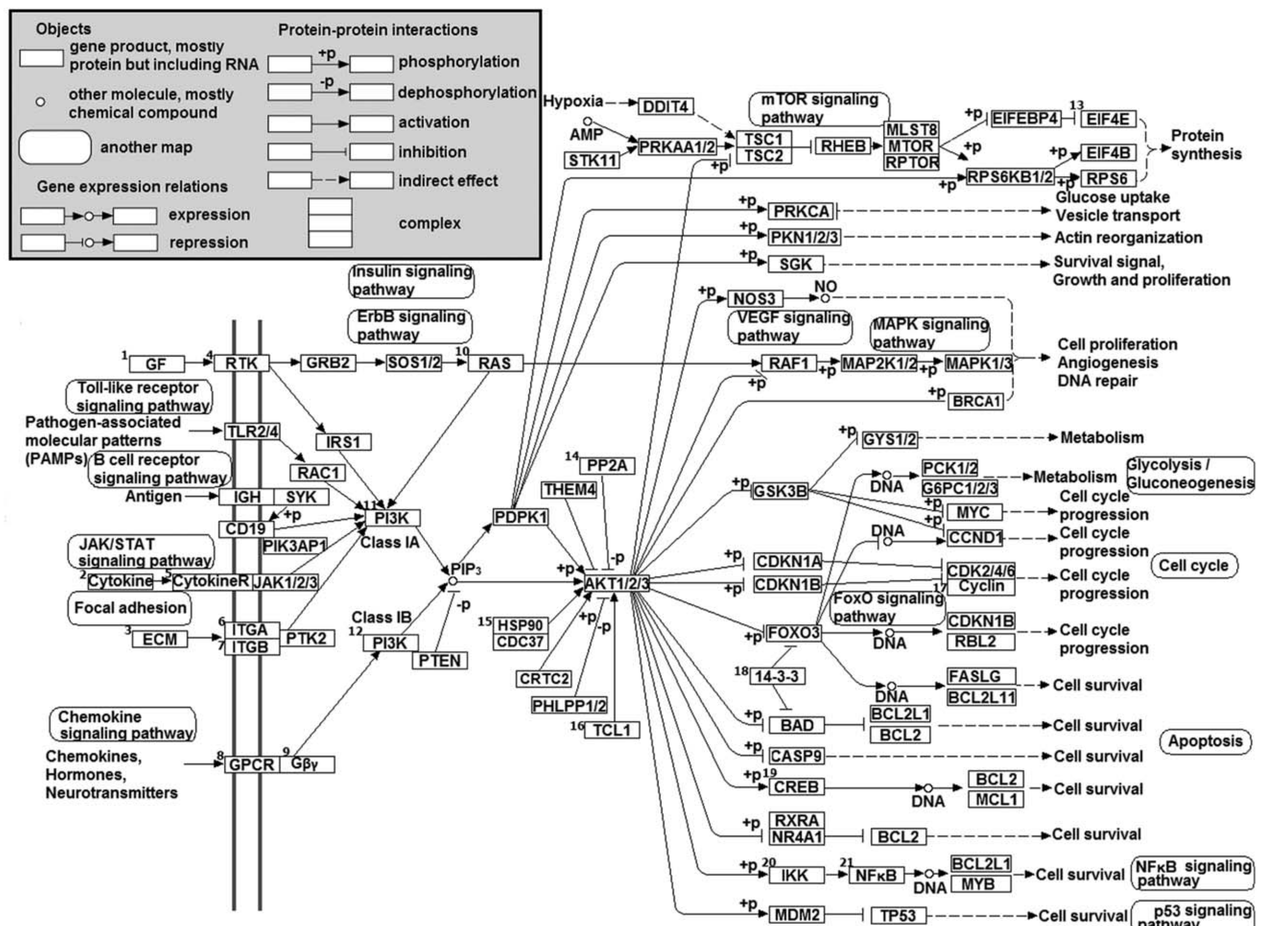
http://dx.doi.org/10.1097/MD.0000000000005817

PI3K/AKT pathway and to see whether the variants simultaneously influence multiple genes in the pathway.

2. Materials and methods

2.1. Genotype and gene expression data

All the genes mapped in the PI3K/AKT pathway of the Kyoto Encyclopedia of Genes and Genomes pathway database (http://www.genome.jp/dbget-bin/www_bget?pathway+hsa04151) were selected (Fig. 1). This study used their mRNA expression data on lymphoblastoid cell lines of 373 individuals from European populations of Centre d'Etude du Polymorphisme Humain, Finns, British, and Toscani recruited by the 1000 Genomes Project.[15] The expression levels were assessed through next generation sequencing using the Illumina HiSeq2000 platform with paired-end 75-bp mRNA-seq to reduce technical variation between laboratories than individuals (Mann–Whitney U $P < 2.2 \times 10^{-6}$).[16] The RNAseq data with low mapping quality (MAPQ < 150) were excluded in the current study. Then, mRNA reads were mapped with an average of 48.9 million reads per sample using the GEM mapper.[17] The expression of mRNA was quantified as the sum of all transcript reads per kilobase per million mapped reads per gene.[15] All the expression data were normalized by probabilistic estimation of expression residuals to remove technical variation.[18] Their genotypic data were obtained from the 1000 Genomes Project (phase 1) using Genome Analyser II and SOLiD.[19] They included genotypes



[1]EGF, FGF1, FGF2, FGF3, FGF4, FGF17, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF16, FGF5, FGF18, FGF19, FGF20, FGF21, FGF22, FGF23, NGF, INS, IGF1, PDGFA, PDGFB, PDGFC, PDGFD, CSF1, KITLG, VEGFA, VEGFB, PGF, VEGFC, FIGF, HGF, ANGPT1, ANGPT2, ANGPT4, EFNA1, EFNA2, EFNA3, EFNA4, and EFNA5; [2] CSF3, CSH1, CSH2, EPO, GH1, GH2, IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA7, IFNA8, IFNA10, IFNA13, IFNA14, IFNA16, IFNA17, IFNA21, IFNB1, IL2, IL3, IL4, IL6, IL7, OSM, and PRL; [3]COL1A1, COL1A2, COL2A1, COL4A2, COL4A4, COL4A6, COL4A1, COL4A5, COL4A3, COL6A1, COL6A2, COL6A3, COL6A6, COL6A5, COL9A1, COL9A2, COL9A3, LAMA1, LAMA2, LAMA3, LAMA5, LAMA4, LAMB1, LAMB2, LAMB3, LAMB4, LAMC1, LAMC2, LAMC3, CHAD, RELN, THBS1, COMP, THBS2, THBS3, THBS4, FN1, SPP1, VTN, TNC, TNN, TNR, TNXB, VWF, and IBSP; [4]CSF1R, EGFR, EPHA2, FGFR1, FGFR3, FGFR2, FGFR4, FLT1, FLT4, IGF1R, INSR, KDR, KIT, MET, NGFR, PDGFRA, PDGFRB, and TEK; [5]CSF3R, EPOR, GHR, IFNAR1, IFNAR2, IL2RA, IL2RB, IL2RG, IL3RA, IL4R, IL6R, IL7R, PRLR, and OSMR; [6]ITGA1,ITGA2, ITGA2B, ITGA3, ITGA4, ITGA5, ITGA6, ITGA7, ITGA8, ITGA8, ITGA9, ITGA10, ITGA11, and ITGAV; [7]ITGB1, ITGB3, ITGB4, ITGB5, ITGB6, ITGB7, and ITGB8; [8]F2R, CHRM1, CHRM2, LPAR1, LPAR2, LPAR3, LPAR4, LPAR5, and LPAR6; [9]GNB1, GNB2, GNB3, GNB4, GNB5, GNG2, GNG3, GNG4, GNG5, GNG7, GNG8, GNG10, GNG11, GNG12, GNG13, GNGT1, and GNGT2; [10]HRAS, KRAS, and NRAS; [11]PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, and PIK3R3; [12]PIK3CG and PIK3R5; [13]EIF4E, EIF4E1B, and EIF4E2; [14]PPP2CA, PPP2CB, PPP2R1B, PPP2R1A, PPP2R2A, PPP2R2B, PPP2R2C, PPP2R2D, PPP2R3B, PPP2R3C, PPP2R3A, PPP2R5B, PPP2R5C, PPP2R5D, PPP2R5E, and PPP2R5A; [15]HSP90AA1, HSP90AB1, and HSP90B1; [16]MTCP1, TCL1A, and TCL1B; [17]CCND1, CCND2, CCND3, CCNE1, and CCNE2; [18]YWHAB, YWHAE, YWHAG, YWHAH, YWHAQ, and YWHAZ; [19]CREB1, CREB3, CREB5, CREB3L1, CREB3L2, CREB3L3, CREB3L4, ATF2, ATF4, and ATF6B; [20]CHUK, IKBKB, and IKBKG; [21]NFKB1 and RELA.

Figure 1. Diagram of genes in the PI3K/AKT signaling pathway. The genes were obtained from the KEGG pathway database and used in this study. KEGG = kyoto encyclopedia of genes and genomes, PI3K/AKT = phosphatidylinositol-3-kinase/v-akt murine thymoma viral oncogene homolog.

called using Beagle[20] and imputed using and MaCH/Thunder.[21] Variants with high confidence calls (phred-scaled quality score = 100) and high imputation quality (MaCH RSQ < 0.3) were used in this study. Depth of coverage was unavailable and thus could not be considered in filtering variants. Nucleotide variants with genotypic call rate <95%, with minor allele frequency <0.05, or in Hardy–Weinberg disequilibrium (P < 0.001) were removed, and a total of 5,941,815 nucleotide variants were used in the analysis. Ethical approval was not necessary because we dealt with publically available data.

2.2. Statistical analyses

The genome-wide association analysis for expression of 341 genes involved in the PI3K/AKT pathway was conducted using a mixed model.[22] The analytical model employed a genetic relationship matrix for random polygenic effects to avoid population stratification as follows:

$$y = \mu\mathbf{1} + \mathrm{x}\beta + \mathrm{g} + \varepsilon$$

where $\mathbf{y}$ is the vector of gene expressions, μ the overall mean, $\mathbf{1}$ the vector of 1's, β the fixed effect for the minor allele of the candidate variant, and $\mathbf{x}$ is the vector with elements of 0, 1, and 2 for the homozygote of the minor allele, heterozygote, and homozygote of the major allele, $\mathbf{g}$ is the vector of random polygenic effects $\left(g \sim \mathrm{N}(\mathbf{0}, A\sigma_{\mathrm{g}}^2)\right)$ where $\mathbf{A}$ is the genomic relationship matrix with elements of pairwise relationship coefficients and σ_g^2 is the polygenic variance component, $\boldsymbol{\varepsilon}$ is the vector of residual effects $\left(\varepsilon \sim \mathrm{N}(0, I\sigma_{\varepsilon}^2)\right)$ where σ_{ε}^2 is the residual variance component, and $\mathbf{I}$ is the identity matrix. The polygenic and residual variance components were estimated using restricted maximum likelihood. The fixed effect was then solved with the estimated variance components under the mixed model equations. The association analysis was conducted using Genome-wide Complex Trait Analysis version 1.26.[23] Multiple testing was applied with a significance threshold of $P < 5 \times 10^{-8}$.[24] Linkage disequilibrium blocks were mapped to exclude dependent associations produced by linkage. The blocks were determined based on the algorithm of Gabriel et al[25] with 95% confidence interval of D′ value using Haploview.[26] Potential deleteriousness of identified variants were examined using the combined annotation dependent depletion.[27] Potential regulatory variants were also determined based on RegulomeDB.[28] We further examined whether the variants associated with gene expression showed functional evidence for an enhancer or a promoter using data resulted from a chromatin interaction analysis with paired-end-tag sequencing (ChIA-PET) study.[29] The ChIA-PET study was conducted with RNA polymerase II using MCF7 cells (human breast cancer cell line) and K562 cells (human erythroleukemic cell line). Correlation between gene expressions was assessed using SPSS 23.0 software for Windows (SPSS Inc., Chicago, IL). Epistasis was analyzed to examine interaction between expression quantitative loci (eQTLs) simultaneously identified for specific genes using linear regression with the function of "epistasis" implemented in PLINK version 1.9.[30] Significance threshold was adjusted for the multiple testing by the Bonferroni correction. We conducted a functional enrichment analysis with the genes containing the eQTLs using ToppGene.[31]

3. Results

The genome-wide association analysis revealed 4166 single nucleotide polymorphisms (SNPs) associated with expression of 85 genes involved in the PI3K/AKT pathway ($P < 5 \times 10^{-8}$, Supplementary Table S1, http://links.lww.com/MD/B493). Among them, 244 SNPs were associated with expression of multiple genes. The SNPs were contained in 73 linkage disequilibrium blocks (Supplementary Table S2; Supplementary Figs. S1 and S2, http://links.lww.com/MD/B493). This indicates that 73 association signals were identified in the current study. The signals included 3 cis-eQTLs and 70 trans-eQTLs. There were one signals identified for association with 4 genes, 3 signals for association with 3 genes, and 69 signals for association with 2 genes. Most of the gene groups had multiple signals associated simultaneously with their expression. In particular, 13 association signals were associated with expression of both *collagen type IV alpha 1* (*COL4A1*) and *collagen type IV alpha 2* (*COL4A2*; Table 1), and 18 signals were associated with expression of both *collagen type IV alpha 5* (*COL4A5*) and *collagen type IV alpha 6* (*COL4A6*; Table 2).

The combined annotation dependent depletion scores of the identified variants showed that rs78851618 and rs117766049 might have the 10% most deleterious substitution. The functional annotation analysis with RegulomeDB revealed variants likely to have regulatory functions (Supplementary Tables S3 and S4, http://links.lww.com/MD/B493). Especially, both rs7165046 and rs145692667 had annotation for transcription factor binding, any motif, DNase footprint and DNase peak. These nucleotide variants were identified as an enhancer or a promoter identified by ChIA-PET (Tables 1 and 2).

Expression of *COL4A1* was positively correlated with that of *COL4A2* in lymphoblastoid cell lines ($r^2 = 0.878$; Fig. 2A). There were 2 individuals with a large expression (>2 SD) of *COL4A2*. One of them had all homozygotes of minor alleles at single-nucleotide variants identified in Table 1. Similarly, expression of *COL4A5* was also positively correlated with that of *COL4A2* ($r^2 = 0.876$; Fig. 2B). Two individuals were observed with a large expression (>2 SD) of *COL4A5*, and one of them had all homozygotes of minor alleles at single-nucleotide variants identified in Table 2. After removing individuals who had a large expression (>2 SD), positive correlations were still obtained ($P < 0.05$; $r^2 = 0.771$ between *COL4A1* and *COL4A2*, $r^2 = 0.674$ between *COL4A5* and *COL4A6*).

Further analysis showed interactions between the eQTLs significant for the genes encoding type IV collagen in Supplementary Tables S5 and S6, http://links.lww.com/MD/B493. All the pairwise interactions were significant ($P < 2.16 \times 10^{-4}$).

Functional enrichment analysis showed that the genes containing the eQTLs were significantly enriched in many biological processes, especially in cellular functions such as cell differentiation, cell development, cell morphogenesis, and cell adhesion (FDR < 0.05; Fig. 3).

4. Discussion

The current study revealed 73 eQTLs associated with 2 or more genes involved in the PI3K/AKT pathway ($P < 5 \times 10^{-8}$). The gene groups were mostly associated with multiple eQTLs. The genes were simultaneously upregulated by specific allele of each eQTL. These results suggested that the genes in each group might be needed for incorporative functions. For example, 9 signals were identified for association with mRNA expression of both *collagen type I alpha 1* (*COL1A1*) and *integrin alpha 11* (*ITGA11*). Dimer of *ITGA11/integrin beta 1* (*ITGB1*) is well known as a receptor for *COL1A1*, and the complex of *ITGA11/ITGB1/COL1A1* is a major part of plasma membrane.[32] In

Table 1

Nucleotide variants associated with gene expression of *COL4A1* and *COL4A2* in lymphoblastoid cell lines of Europeans by a genome-wide association study[*].

					COL4A1		COL4A2			
SNV	**Position[†]**	**Gene**	**Allele[‡]**	**MAF**	β	***P***	β	***P***	**E/P[§]**	**CADD score[**]**
rs116248210	2 103,945,833	Intergenic	C/T	0.054	0.124	1.49×10^{-10}	0.345	8.79×10^{-10}		0.290
rs116506968	2 183,551,471	LOC101929976 (Intron 1)	G/A	0.080	0.108	1.13×10^{-10}	0.264	2.96×10^{-9}		1.173
rs34845474[‖]	4 169,751,339	PALLD (Intron 10)	C/T	0.054	0.119	7.82×10^{-9}	0.346	9.32×10^{-10}	P(*PALLD*)	0.249
rs33951557[‖]	6 81,393,482	Intergenic	C/A	0.055	0.116	2.79×10^{-8}	0.338	1.51×10^{-9}		5.383
rs55869773	7 4,498,947	Intergenic	G/A	0.054	0.113	4.27×10^{-8}	0.325	2.84×10^{-9}		1.921
rs76598348[‖]	9 7,568,139	Intergenic	G/A	0.076	0.106	3.95×10^{-9}	0.267	2.37×10^{-8}		1.328
rs1987249[‖]	9 108,683,134	Intergenic	C/A	0.060	0.115	1.26×10^{-8}	0.328	2.07×10^{-9}		2.760
rs4276755	9 124,377,411	DAB2IP (Intron 1)	A/G	0.056	0.109	3.10×10^{-8}	0.331	2.99×10^{-10}		5.354
rs116674084	12 61,420,772	Intergenic	G/A	0.070	0.102	4.36×10^{-8}	0.270	4.18×10^{-8}		2.798
rs182060879	12 85,650,851	LRRIQ1 (Intron 27)	A/C	0.055	0.117	5.21×10^{-9}	0.340	1.11×10^{-10}		0.241
rs11867392	17 1,921,816	RTN4RL1 (Intron 1)	C/T	0.068	0.103	6.45×10^{-9}	0.281	2.14×10^{-9}		2.584
rs6002762[‖]	22 22,910,191	Intergenic[¶]	T/C	0.051	0.126	9.86×10^{-10}	0.354	7.68×10^{-11}	P (*PRAME, LOC648691*)[#]	4.302
rs74617949	22 27,681,626	Intergenic	T/C	0.068	0.110	4.09×10^{-9}	0.297	2.14×10^{-9}		1.396

CADD = combined-annotation dependent depletion, MAF = minor allele frequency, SNV = single-nucleotide variant.
[*] Only SNVs with significant association by multiple testing ($P<5\times10^{-8}$) are presented.
[†] The 1st column shows chromosome no., and the 2nd column shows chromosomal position (bp from GRCh37).
[‡] Major/minor allele.
[§] SNV was identified as an enhancer (E)/a promoter (P) of the gene(s) in parenthesis.
[‖] The association signals were identified by multiple SNVs, and the most significant SNV at each signal is presented.
[¶] The most significant SNV at the signal was an intergenic variant, but the variants at the association signal were located in the genes (*PRAME*, *LOC648691*).
[#] The nucleotide sequences identified as a promoter did not include rs6002762, but the variants linked to it at the association signal (rs3747098, rs8405, rs2073725, rs5996147, rs145692667, rs35317486, rs57306165, rs62225994, rs34575077, rs2266991, rs2266993, and rs2266994).
[**] CADD score greater than 10 means variant with the 10% most deleterious substitution.

Table 2

Nucleotide variants associated with gene expression of *COL4A5* and *COL4A6* in lymphoblastoid cell lines of Europeans by a genome-wide association study[*].

					COL4A5		COL4A6			
SNV	**Position[†]**	**Gene**	**Allele[‡]**	**MAF**	β	***P***	β	***P***	**E/P[§]**	**CADD score[‖]**
rs12759318	1 22,308,435	CELA3B (Intron 4)	A/G	0.064	0.346	4.20×10^{-10}	0.108	2.30×10^{-12}		0.37
rs2506992	1 48,097,805	Intergenic	G/C	0.054	0.378	1.10×10^{-10}	0.092	4.03×10^{-8}		3.10
rs17019582[¶]	2 4,702,434	LINC01249 (Intron 1)	T/C	0.051	0.423	3.02×10^{-9}	0.096	7.69×10^{-9}		1.73
rs72992248	2 231,197,652	SP140L (Intron 2)	C/G	0.051	0.361	1.25×10^{-12}	0.092	3.86×10^{-8}		4.09
rs72766225	5 63,042,657	Intergenic	T/C	0.051	0.354	3.33×10^{-10}	0.092	1.58×10^{-8}		4.64
rs17545761[¶]	5 146,647,488	STK32A (Intron 2)	C/A	0.066	0.368	5.14×10^{-11}	0.109	7.59×10^{-13}		3.94
rs78851618	6 147,047,296	ADGB (Exon 19)	T/C	0.052[#]	0.368	3.04×10^{-9}	0.093	4.95×10^{-8}		12.86
rs16871986	7 105,399,650	ATXN7L1 (Intron 18)	C/A	0.082	0.265	1.93×10^{-8}	0.076	2.53×10^{-8}	E (ATXN7L1, CDHR3)	6.18
rs117766049[¶,**]	7 129,684,199	ZC3HC1 (Intron 8)	C/T	0.052	0.411	2.87×10^{-12}	0.122	3.90×10^{-14}		10.95
rs117417471[**]	10 56,973,794	Intergenic	A/G	0.104	0.335	2.44×10^{-8}	0.111	5.95×10^{-11}		3.59
rs7121626	11 5,379,516	OR51B5 (Intron 2)	C/T	0.054	0.368	2.30×10^{-9}	0.100	3.39×10^{-9}		0.47
rs7964807[¶,††]	12 74,484,387	Intergenic	C/G	0.058	0.331	2.03×10^{-8}	0.088	3.91×10^{-8}	E (LOC100507377)[‡‡]	1.57
rs189351960[¶,††]	12 74,560,964	LOC100507377 (Intron 4)	G/T	0.055	0.339	2.88×10^{-8}	0.102	1.09×10^{-9}	E (LOC100507377)[‡‡]	0.88
rs117219682	14 59,558,654	Intergenic	T/C	0.051	0.399	8.17×10^{-11}	0.115	2.13×10^{-11}		5.06
rs72721031	14 68,012,909	PLEKHH1 (Intron 2)	C/T	0.062	0.327	3.44×10^{-9}	0.089	5.12×10^{-9}		0.25
rs7165046	15 97,778,882	Intergenic	A/C	0.059	0.318	4.18×10^{-8}	0.088	4.72×10^{-8}	E (NR2F2)	0.45
rs3784843[¶]	16 62,003,873	CDH8 (Intron 2)	G/A	0.070	0.302	7.84×10^{-9}	0.088	1.01×10^{-9}		3.36
rs7226068	17 17,231,431	NT5M (Intron 3)	G/A	0.058	0.351	1.00×10^{-9}	0.096	1.68×10^{-9}		2.55

CADD = combined annotation dependent depletion, MAF = minor allele frequency, SNV = single-nucleotide variant.
[*] Only SNVs with significant association by multiple testing ($P<5\times10^{-8}$) are presented.
[†] The 1st line shows chromosome no. and the 2nd line shows chromosomal position (bp from GRCh37).
[‡] Major/minor allele.
[§] SNV was identified as an enhancer (E)/a promoter (P) of the gene in parenthesis.
[‖] CADD score greater than 10 means variant with the 10% most deleterious substitution.
[¶] The association signals were identified by multiple SNVs, and the most significant SNV at each signal is presented.
[#] The Exome Aggregation Consortium database showed 0.063 as an MAF estimate of the exonic nucleotide variant in Europeans.
[**] The variants were additionally identified for other genes (rs117766049 associated with expression of *KITLG* with $P=1.23\times10^{-8}$; rs117417471 associated with expression of *F2R* with $P=4.91\times10^{-8}$).
[††] The 2 variants (rs7964807 and rs189351960) were identified at 1 association signal. At the signal, rs7964807 was the most significant variant associated with expression of *COL4A5*, and rs189351960 was the most significant variant associated with expression of *COL4A6*.
[‡‡] The nucleotide sequences identified as an enhancer did not include rs7964807 or rs189351960, but include variants (rs7960444 and rs2049205) linked to them at the association signal.

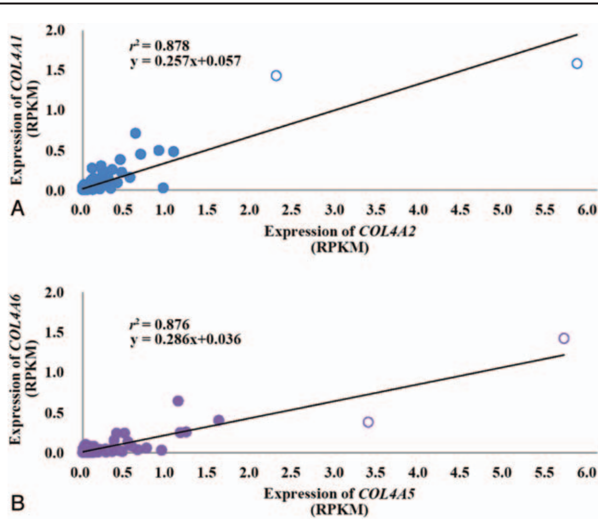


Figure 2. Correlation of expression between *COL4A1* and *COL4A2* (A), between *COL4A5* and *COL4A6* (B) in European lymphoblastoid cell lines. Unfilled circles indicate individuals with expression greater than 2 standard deviations of reads per kilobase per million mapped reads (RPKM).

addition, 2 gene pairs are encoding type IV collagen molecules. The genes of *COL4A1* and *COL4A2* were associated with 13 eQTLs, and the genes of *COL4A5* and *COL4A6* were associated with 18 eQTLs. The collagen molecules are most abundant in basement membranes and are involved in ensuring structural integrity in tissues and in modulating cell differentiation, cell growth, and cell adhesion.[33,34] The genes of *COL4A1* and *COL4A2* have a unique head-to-head orientation, and the genes of *COL4A5* and *COL4A6* also have it. They can initiate transcription from opposite DNA strands by sharing a bidirectional promoter.[35,36] The *COL4A1* gene was 235 bp apart from *COL4A2* gene, and the *COL4A5* gene was 315 bp apart from *COL4A6* gene. The eQTLs located between the gene pairs are likely to be critical promoter regions. This was supported further by positive correlation of expression between the genes. On the other hand, functional roles of the eQTLs apart from the genes are hardly guessed. Some signals turned out to be an enhancer or a promoter of other genes by ChIA-PET. The rs16871986 associated with the genes of *COL4A5* and *COL4A6* was located within an enhancer of *cadherin related family member 3* (*CDHR3*) gene. The *CDHR3* mediates homophilic cell adhesion through its binding to calcium ions.[37,38] This leads to a possibility of direct or indirect induction of *COL4A5* and *COL4A6* by *CDHR3*. Similarly, the rs34845474 associated with the genes of *COL4A1* and *COL4A2* was located within the promoter of *palladin* (*PALLD*). The *PALLD* regulates actin cytoskeletal organization and cell adhesion formation.[39] Furthermore, it was identified as a key node in *focal adhesion kinase* (*FAK*) and *PI3K* of the PI3K/AKT pathway.[40] Collagens are synthesized for a variety of human tissues by fibroblast which can be activated by *FAK* and *PI3K*.[40] This suggests the *PALLD* might ultimately induce *COL4A1* and *COL4A2*.

The 13 eQTLs identified for both *COL4A1* and *COL4A2* were supported by positive correlation between the number of their minor alleles and *COL4A1*/*COL4A2* expression. Similarly, the 18 eQTLs identified for both *COL4A5* and *COL4A6* were also supported by positive correlation between the number of their minor alleles and their expression. In particular, all the minor alleles at the 13 eQTLs were observed only in 1 individual who had the largest expression of *COL4A2*, and all the minor alleles at the 18 eQTLs were observed only in 1 individual who had the largest expression of *COL4A5*.

This study also revealed strong epistases between the eQTLs for the genes encoding type IV collagen. For example, individuals with "CC" genotype at rs4276755 showed increased expression of *COL4A2* by nucleotide substitution of rs55869773 from "GG" to "GA/AA" whereas individuals with "CT/TT" genotype at rs4276755 showed decreased the expression ($P = 1.86 \times$

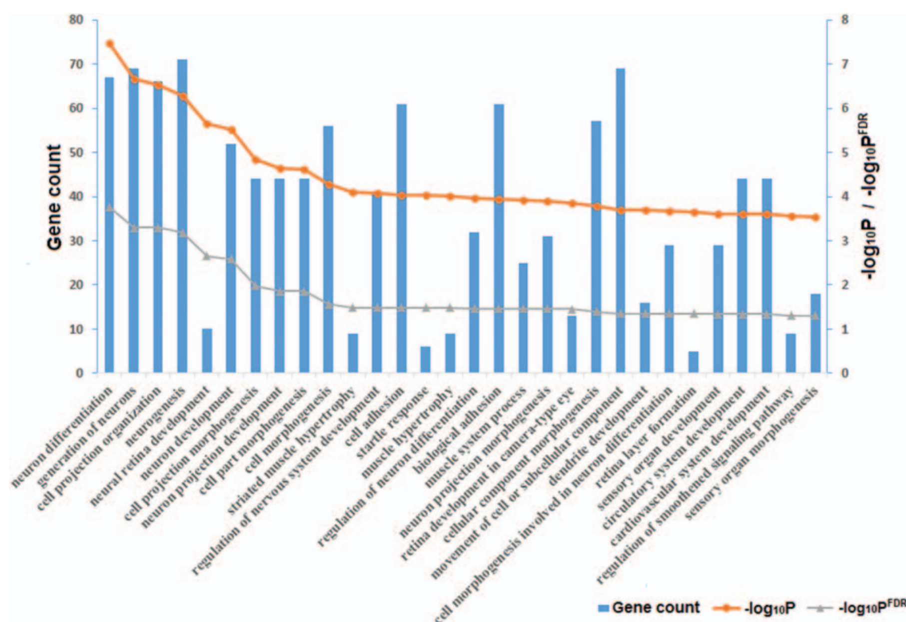


Figure 3. Functional enrichment ($P^{FDR} < 0.05$) of the genes containing eQTL for genes involved in the PI3K/AKT signaling pathway. eQTL = expression quantitative trait locus, PI3K/AKT = phosphatidylinositol-3-kinase/v-akt murine thymoma viral oncogene homolog.

10^{-237}; Supplementary Fig. S3, http://links.lww.com/MD/B493). We attempted to examine several epistasis among 3 eQTLs and found them significant (data not shown), which implied other epistasis among more than 2 eQTLs. This suggests a variety of interactions among eQTLs in regulating the gene expression.

We suggested that multiple genes involved in the PI3K/AKT pathway were simultaneously expressed with associations of multiple eQTLs. Further studies are required to understand relationship among the genes simultaneously expressed in the current study and their specific roles and mechanisms in the PI3K/AKT pathway.

Acknowledgments

The authors thank the Basic Science Research Program through the National Research Foundation of Korea, Ministry of Education, Science, and Technology (Grant No. NRF-2012M3A9D1054705) for the support.

References

[1] Engelman JA, Luo J, Cantley LC. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. Nat Rev Genet 2006;7:606–19.
[2] Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. Cell 2007;129:1261–74.
[3] Zoncu R, Efeyan A, Sabatini DM. mTOR: from growth signal integration to cancer, diabetes and ageing. Nat Rev Mol Cell Biol 2011;12:21–35.
[4] Hennessy BT, Smith DL, Ram PT, et al. Exploiting the PI3K/AKT pathway for cancer drug discovery. Nat Rev Drug Discov 2005;4:988–1004.
[5] Brugge J, Hung MC, Mills GB. A new mutational AKTivation in the PI3K pathway. Cancer Cell 2007;12:104–7.
[6] Porta C, Paglino C, Mosca A. Targeting PI3K/Akt/mTOR signaling in cancer. Front Oncol 2014;4:64.
[7] Dienstmann R, Rodon J, Serra V, et al. Picking the point of inhibition: a comparative review of PI3K/AKT/mTOR pathway inhibitors. Mol Cancer Ther 2014;13:1021–31.
[8] Massacesi C, Di Tomaso E, Urban P, et al. PI3K inhibitors as new cancer therapeutics: implications for clinical trial design. Onco Targets Ther 2016;9:203–10.
[9] Brown KK, Toker A. The phosphoinositide 3-kinase pathway and therapy resistance in cancer. F1000Prime Rep 2015;7:13.
[10] Samuels Y, Wang Z, Bardelli A, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science 2004;304:554.
[11] Liaw D, Marsh DJ, Li J, et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 1997;16:64–7.
[12] Blumenthal GM, Dennis PA. PTEN hamartoma tumor syndromes. Eur J Hum Genet 2008;16:1289–300.
[13] Hildebrandt MA, Yang H, Hung MC, et al. Genetic variations in the PI3K/PTEN/AKT/mTOR pathway are associated with clinical outcomes in esophageal cancer patients treated with chemotherapy. J Clin Oncol 2009;27:857–71.
[14] Hutz JE, Manning WA, Province MA, et al. Genomewide analysis of inherited variation associated with phosphorylation of PI3K/AKT/mTOR signaling proteins. PLoS One 2011;6:e24873.
[15] Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.
[16] 't Hoen PA, Friedländer MR, Almlöf J, et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol 2013;31:1015–22.
[17] Marco-Sola S, Sammeth M, Guigo R, et al. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 2012;9:1185–8.
[18] Stegle O, Parts L, Durbin R, et al. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol 2010;6:e1000770.
[19] Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
[20] Browning BL, Browning SR. A unified approach to genotype imputation and haplotype and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009;84:210–23.
[21] Li Y, Sidore C, Kang HM, et al. Low-coverage sequencing: implications for design of complex trait association studies. Genome Res 2011;21:940–51.
[22] Ryu D, Ryu J, Lee C. Genome-wide association study reveals sex-specific selection signals against autosomal nucleotide variants. J Hum Genet 2016;61:423–6.
[23] Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.
[24] Pe'er I, Yelensky R, Altshuler D, et al. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008;32:381–5.
[25] Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
[26] Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5.
[27] Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
[28] Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790–7.
[29] Li G, Ruan X, Auerbach RK, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012;148:84–98.
[30] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
[31] Chen J, Bardes EE, Aronow BJ, et al. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 2009;37:W305–311.
[32] Velling T, Kusche-Gullberg M, Sejersen T, et al. cDNA cloning and chromosomal localization of human alpha(11) integrin. A collagen-binding, I domain-containing, beta(1)-associated integrin alpha-chain present in muscle tissues. J Biol Chem 1999;274:25735–42.
[33] Timpl R. Structure and biological activity of basement membrane proteins. Eur J Biochem 1989;180:487–502.
[34] Hudson BG, Reeders ST, Tryggavason K. Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J Biol Chem 1993;268:26033–6.
[35] Pöschl E, Pollner R, Kühn K. The genes for the alpha 1(IV) and alpha 2 (IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. EMBO J 1988;7:2687–95.
[36] Sugimoto M, Oohashi T, Ninomiya Y. The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proc Natl Acad Sci U S A 1994;91:11679–83.
[37] Hulpiau P, van Roy F. Molecular evolution of the cadherin superfamily. Int J Biochem Cell Biol 2009;41:349–69.
[38] Bønnelykke K, Sleiman P, Nielsen K, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet 2014;46:51–5.
[39] Goicoechea SM, Arneman D, Otey CA. The role of palladin in actin organization and cell motility. Eur J Cell Biol 2008;87:517–25.
[40] McLane JS, Ligon LA. Palladin mediates stiffness-induced fibroblast activation in the tumor microenvironment. Biophys J 2015;109:249–64.