

# Strong cis-acting expression quantitative trait loci for the genes encoding SNHG5 and PEX6

Jihyeon Lee, Jihye Ryu, PhD, Chaeyoung Lee, PhD\*

## Abstract

Expression of quantitative trait loci (eQTLs) for the genes located in human chromosome 6 were examined. Data on RNA expression in lymphoblastoid cells of 373 unrelated Europeans were used to identify eQTLs.

Genome-wide analysis resulted in 24,447 nucleotide variants associated with gene expression ( $P < 2.16 \times 10^{-10}$ ). We found 36 variants with  $P < 10^{-100}$ , which were all associated with expression levels of the genes encoding small nucleolar RNA host gene 5 (SNHG5) and peroxisomal biogenesis factor 6 (PEX6). Enhancer eQTLs downstream of the SNHG5 gene might be candidate genetic factors for susceptibility to cancer. This is because nucleotide substitutions (eg, G→T at rs6922) of the enhancer eQTLs may cause low expression of SNHG5 gene, and low expression of snoRNA U50, a product generated from introns of the SNHG5 gene, can induce cancer. One presently identified eQTL for the PEX6 gene was rs10948059, which had been associated with prostate cancer from previous association studies. The results imply that variants associated with prostate cancer can be identified through expressional change in the PEX6 gene, but not in the overlapped glycine N-methyltransferase gene which had been considered as a candidate gene.

Further studies are required to understand their underlying mechanisms for the strong eQTLs for the SNHG5 and PEX6 genes.

**Abbreviations:** ChIA-PET = chromatin interaction analysis with paired-end-tag sequencing, eQTL = expression quantitative trait locus, GRM = genomic relationship matrix, GWAS = genome-wide association study, HLA = human leukocyte antigen, LD = linkage disequilibrium, MAF = minor allele frequency, RPKM = reads per kilobase per million mapped reads, SNP = single-nucleotide polymorphism.

**Keywords:** chromosome 6, eQTL, gene expression, PEX6, SNHG5, SNP

## 1. Introduction

As of 2003, more than 170 million base pairs had been completely sequenced for human chromosome 6.<sup>[1]</sup> Genome-wide association studies (GWAS) have unraveled a large number of single-nucleotide variants associated with human traits. According to the GWAS Catalog, 32,768 genetic associations were reported across hundreds of diseases and quantitative traits as of October 2016 ( $P < 10^{-5}$ ; <https://www.ebi.ac.uk/gwas>). The GWAS signals were more frequently found in chromosome 6 than in any other chromosomes, even in chromosome 1, which is 1.5 times longer than chromosome 6. The chromosome 6 has medical significance, with about 120 causal genes reported for complex diseases, such as cancers, heart disease, diabetes, Alzheimer disease, rheumatoid

arthritis, and multiple sclerosis.<sup>[2–4]</sup> In particular, major histocompatibility complex and related regions including the human leukocyte antigen (HLA) genes in chromosome 6 are critically related to infection, immunity, and inflammation.<sup>[2]</sup> Many nucleotide variants turned out to be important for susceptibility to complex diseases from GWAS. One example is rs9378815, an intergenic single-nucleotide polymorphism (SNP) on chromosome 6, which has been identified as a rheumatoid arthritis risk locus from a trans-ethnic GWAS meta-analysis using >100,000 subjects ( $P = 1.7 \times 10^{-10}$ ).<sup>[5]</sup> Another example is rs2046210, an SNP upstream of the gene encoding estrogen receptor alpha on chromosome 6, which was associated with susceptibility to breast cancer in a study involving over 31,000 subjects (eg,  $P = 1.5 \times 10^{-30}$  in Chinese women).<sup>[6]</sup> Nevertheless, their underlying mechanisms are unknown. Identifying nucleotide variants that influence RNA transcript expression levels is critical to understand underlying mechanisms of GWAS signals. The objective of the current study was to identify expression quantitative trait loci (eQTLs) for genes located in chromosome 6, which might have abundant potential regulatory sequences.

## 2. Methods

The study used RNA expression data for lymphoblastoid cells of 376 unrelated individuals from European populations of CEPH (CEU), Finns (FIN), British (GBR), and Toscani (TSI) to identify regulatory variants for expression of the 573 genes in chromosome 6, which were obtained from the study by Lappalainen et al.<sup>[7]</sup> The genes included protein-coding genes and long intergenic noncoding RNA genes. The expression level was calculated as the sum of reads per kilobase of transcript per million mapped reads values of all the transcripts of each gene for each individual. A read was counted in an exon if either start

Editor: Xiaolin Zhu.

Funding: 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 funding and conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Department of Bioinformatics and Life Science, Soongsil University, 369 Sangdo-ro, 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 © 2016 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 (2016) 95:52(e5793)

Received: 8 August 2016 / Received in final form: 23 November 2016 /

Accepted: 8 December 2016

<http://dx.doi.org/10.1097/MD.0000000000005793>

or end coordinate of the read was in the exon. To avoid overestimating gene expressions, the count was divided by the number of overlapping exons among split reads. Their corresponding genotypes were obtained from the 1000 Genomes project (phase I; <http://phase1browser.1000genomes.org/>) that originally conducted whole genome sequencing with an average depth of  $5\times$ , and whole exome sequencing with an average depth of  $80\times$  using 1092 individuals from 14 populations of Europeans, East Asians, Saharan Africans, and Americans. A total of 404,240 SNPs on chromosome 6 among 5,941,815 genome-wide SNPs were included in the current analysis after a filtration process (excluding SNPs with minor allele frequency [MAF]  $<0.05$ ). Three individuals with missing rate  $>0.05$  of the genotypes were excluded from the analysis. Ethical approval was not necessary because we dealt with publically available data.

Genetic associations of SNPs with RNA expression were conducted to determine the significance of regression coefficient for minor allele effect. We employed the following mixed model incorporating a genomic relationship matrix (GRM) that reflected polygenic covariance among individuals to avoid population stratification<sup>[8]</sup>:

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

where  $y$  is the vector of gene expression levels,  $\mu$  is the overall mean,  $\mathbf{1}$  is the vector of 1s,  $\beta$  is the fixed effect for minor allele of the SNP to be tested for association, and  $x$  is the vector of the numbers of minor alleles of the SNP, for example, 0(0, 0) for homozygote of the major allele, 1 (1, 0) for heterozygote, and 2 (1, 1) for homozygote of the minor allele in additive (dominant, recessive) model. Vector  $g$  is the vector of random polygenic effects with  $g \sim N(0, A\sigma_g^2)$ , where  $A$  is GRM estimated using SNPs, and  $\sigma_g^2$  is the polygenic variance component. Elements of  $A$  consist of the following pair-wise genomic relationship coefficients estimated using genotypes of SNPs in linkage equilibrium ( $r^2 > 0.8$ ):

$$a_{jk} = \frac{1}{N_S} \sum_{i=1}^{N_S} \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}$$

where  $a_{jk}$  is the genomic relationship coefficient between  $j^{\text{th}}$  and  $k^{\text{th}}$  individuals,  $N_S$  is the number of SNPs,  $x_{ij}$  ( $x_{ik}$ ) is the number (0, 1, or 2) of the minor allele at the  $i^{\text{th}}$  SNP for the  $j^{\text{th}}$  ( $k^{\text{th}}$ ) individual,  $p_i$  is the frequency of the minor allele at the  $i^{\text{th}}$  SNP.  $\varepsilon$  is the vector of random residuals with  $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$ , where  $\sigma_\varepsilon^2$  is the residual variance component and  $I$  is the identity matrix. All the association analyses were conducted using Genome-wide Complex Trait Analysis v1.24 freeware.<sup>[9]</sup>

Multiple testing was applied employing Bonferroni correction. A total of 231,629,520 association tests for 573 genes and 404,240 SNP markers resulted in the significance threshold value of  $2.16 \times 10^{-10}$ . The association tests were analyzed using PLINK ver.1.04 (Free Software Foundation, Inc. Boston, MA).

Linkage disequilibrium (LD) blocks were constructed based on the algorithm of Gabriel et al.<sup>[10]</sup> (Haploview version 4.2). The algorithm used a 95% confidence interval (CI) of pair-wise  $D'$  estimate between SNPs with MAF  $>0.05$ .

Promoters/enhancers including the significant SNPs were examined using genome-wide chromatin interaction analysis with paired-end-tag sequencing (ChIA-PET) data, which resulted from the study by Li et al.<sup>[11]</sup> The paired-end-tag (PET) sequences were mapped with the human reference genome of hg19. High-confidence intrachromosomal PET clusters with  $P_{\text{FDR}} < 0.05$  were

included with genomic span of 8Kb to 1Mb and PET count  $>3$  for each PET cluster from the saturated libraries. Interaction anchors flanking transcription start site  $\pm 5$ Kb were defined as promoters.

Enrichment analysis was conducted to analyze associations of significant SNPs with diseases and biological processes using genomic regions enrichment of annotations tool.<sup>[12]</sup> Diseases and biological processes based on the Gene Ontology (<http://www.geneontology.org/>) were used as input annotation terms. Each gene involved in input annotation terms was assigned to a regulatory domain that extends up to 1Mb in both directions. Significance of enrichment was determined by a hypergeometric test over genes. The significance threshold was adjusted by false discovery rate (FDR) to control false-positives.

Other enrichment analyses were also performed to determine enrichment of cis versus trans regulation, promoter/enhancer versus nonpromoter/nonenhancer regions, and downstream versus upstream regions. Their significances were all determined with  $P < 0.01$  by the Fisher exact test.

### 3. Results

The eQTL analysis revealed that 24,447 SNPs were associated with expression of the 88 genes located on chromosome 6 ( $P < 2.16 \times 10^{-10}$ ). Most were identified using the additive model (24,213 SNPs). They were enriched ( $P_{\text{FDR}} < 0.05$ , where  $P_{\text{FDR}}$  is  $P$  value adjusted by FDR) with genes involved in a variety of diseases (Supplementary Table 1, <http://links.lww.com/MD/B483>) and biological processes (Supplementary Table 2, <http://links.lww.com/MD/B483>). The SNPs were enriched ( $P < 0.01$ ) in promoter and enhancer regions resulting from a ChIA-PET study.<sup>[11]</sup> There were 18,872 cis-eQTLs within 1Mb from target genes and 5575 trans-eQTLs  $>1$  Mb away from target genes, and they were largely cis-regulatory eQTLs (77% = 18,872/24,447) by the enrichment analysis ( $P < 0.01$ ). The cis-eQTLs appeared considerably clustered with 68% (12,886/18,872) located within 100Kb. Also, 4960 SNPs were associated with expression of multiple genes, which resulted in 18,378 unique SNPs. In particular, 36 SNPs were observed with extremely low  $P$  value ( $P < 10^{-100}$ ; Table 1). All 36 SNPs influenced RNA expression of the genes encoding small nucleolar RNA host gene 5 (SNHG5) and peroxisomal biogenesis factor 6 (PEX6). The SNPs were all located in or near the corresponding gene. On chromosome 6, the SNHG5 gene is located from 85,676,990 to 85,678,736 bp (GRCh38/hg38), and the PEX6 gene is located from 42,964,335 to 42,979,150 bp (GRCh38/hg38). Considerable numbers of SNPs flanking the genes were associated with their expression ( $P < 2.16 \times 10^{-10}$ ; Fig. 1). Most of the strongly significant SNPs with  $P < 10^{-100}$  were found downstream of the SNHG5 gene and within the PEX6 gene. The SNPs were observed in strong linkage ( $r^2 > 0.8$ ), forming LD blocks (Fig. 2).

### 4. Discussion

The current study identified 24,447 SNPs associated with RNA expression of genes located in chromosome 6, including cis-eQTLs (77%) that may affect transcription initiation complex on the same physical chromosome. The cis-eQTLs tended to be considerably clustered with 68% located within 100Kb from their corresponding target gene, which concurred with the previous study where 80% of cis-eQTLs was identified within 100Kb using 9 types of tissues.<sup>[13]</sup>

The current study showed the pleiotropy that 4960 SNPs were associated with expression of multiple genes. This implied direct

**Table 1****Genetic associations of nucleotide variants with mRNA expression of genes located in chromosome 6\*.**

Gene <sup>†</sup>	SNP	Position	Allele <sup>‡</sup>	MAF <sup>§</sup>	<i>P</i> <sub>HWE</sub> <sup>  </sup>	Beta <sup>¶</sup>	<i>P</i>	E/ <i>P</i> <sup>#</sup>
SNHG5	rs1059307	85678170	T/G	0.4353	0.250	-69.37	4.62E-130	P
SNHG5	rs3087978	85678505	T/C	0.4393	0.245	-67.54	1.76E-113	P
SNHG5	rs9450311	85666868	A/G	0.4315	0.245	-67.54	1.76E-113	
SNHG5	rs12193777	85536991	A/C	0.4321	0.206	-67.59	9.89E-113	E
SNHG5	rs12203963	85513598	C/T	0.4323	0.206	-67.59	9.89E-113	E
SNHG5	rs12206573	85548896	A/G	0.4319	0.206	-67.59	9.89E-113	E
SNHG5	rs2758848	85637887	C/T	0.4321	0.206	-67.59	9.89E-113	
SNHG5	rs2842605	85625432	A/T	0.4317	0.206	-67.59	9.89E-113	
SNHG5	rs2842611	85592316	A/G	0.4507	0.206	-67.59	9.89E-113	
SNHG5	rs2842615	85631130	G/C	0.4315	0.206	-67.59	9.89E-113	
SNHG5	rs4640862	85530429	A/G	0.4321	0.206	-67.59	9.89E-113	
SNHG5	rs4707211	85522992	A/G	0.4319	0.206	-67.59	9.89E-113	
SNHG5	rs6454472	85514335	G/T	0.4565	0.206	-67.59	9.89E-113	E
SNHG5	rs7766485	85553837	C/T	0.4317	0.206	-67.59	9.89E-113	
SNHG5	rs9444351	85538081	C/T	0.4319	0.206	-67.59	9.89E-113	E
SNHG5	rs9444352	85547817	A/C	0.4167	0.206	-67.59	9.89E-113	
SNHG5	rs9444354	85582217	A/G	0.4327	0.206	-67.59	9.89E-113	
SNHG5	rs4707212	85524666	T/C	0.4319	0.245	-67.38	1.11E-111	
SNHG5	rs1173418	85693198	A/G	0.4389	0.172	-67.52	1.08E-111	
SNHG5	rs9450292	85526898	A/G	0.4321	0.172	-67.55	2.67E-111	
SNHG5	rs9342045	85715430	T/C	0.4289	0.344	-66.69	9.88E-111	
SNHG5	rs2842600	85615914	C/T	0.4535	0.248	-67.03	9.79E-111	
SNHG5	rs2758849	85634578	G/A	0.4369	0.168	-67.41	3.87E-110	
PEX6	rs6907751	42977112	C/T	0.3814	0.470	2.994	9.44E-109	
PEX6	rs10948059	42960723	C/T	0.3922	0.536	2.979	6.98E-108	E
SNHG5	rs2842610	85588345	A/G	0.4331	0.206	-67.00	1.54E-108	
SNHG5	rs4348284	85640676	C/T	0.4507	0.172	-67.10	2.04E-108	
SNHG5	rs9344539	85643903	A/G	0.4321	0.203	-67.19	1.99E-107	
SNHG5	rs11752795	85577411	A/C	0.4529	0.114	-67.41	1.81E-106	
SNHG5	rs9344540	85654716	A/G	0.4331	0.287	-66.60	2.75E-105	
SNHG5	rs6927475	85497055	A/G	0.4313	0.206	-66.59	1.12E-105	
SNHG5	rs9450285	85498378	G/A	0.4309	0.206	-66.59	1.12E-105	
SNHG5	rs9362232	85643904	A/C	0.4321	0.394	-65.92	5.90E-104	
PEX6	rs9462857	42975360	T/C	0.3407	0.534	2.928	2.72E-101	
PEX6	rs9462859	42979205	G/A	0.3339	0.534	2.928	2.72E-101	
PEX6	rs9986447	42975041	A/G	0.2897	0.467	2.936	4.53E-101	

MAF=minor allele frequency, SNP=single-nucleotide polymorphism.

\* Only SNPs strongly associated with gene expression are presented ( $P < 10^{-100}$ ).

† Its expression was associated with SNP.

‡ Major allele/minor allele.

§ Minor allele frequency.

|| *P* value for testing Hardy-Weinberg equilibrium. All the SNPs with  $P < 10^{-100}$  were in Hardy-Weinberg equilibrium ( $P_{HWE} > 0.05$ ).

¶ Regression coefficient estimated using the additive model. Positive estimate means that minor allele increased expression of the gene, and negative estimate means that minor allele decreased expression.

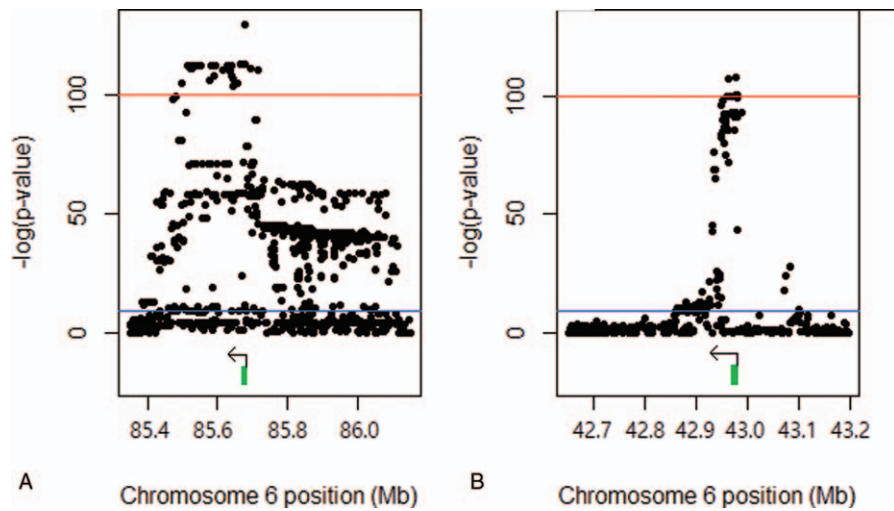
# Enhancer/promoter. The SNPs in second column are within enhancer or promoter of gene in first column.

or indirect functional relationships among genes simultaneously regulated by eQTLs. For example, rs2020202 was associated with expression of 9 histone genes (HIST1B, HIST1E, HIST2AH, HIST2AJ, HIST2BM, HIST3B, HIST3I, HIST3J, and HIST4B), 26 SNPs were associated with expression of 3 HLA genes (HLA-DQB2, HLA-DRB6, and HLA-DRB1), and 9 SNPs were associated with expression of 3 serpin genes (SERPINB1, SERPINB6, and SERPINB9).

In particular, 36 SNPs showed strong associations with  $P < 10^{-100}$ , and they were all associated with expression of the gene encoding SNHG5 or PEX6. All the SNPs were located flanking the corresponding gene within 200 Kb. An enrichment analysis showed that the SNPs strongly associated with expression of the SNHG5 were enriched downstream of the gene ( $P < 0.01$ ). However, SNPs with  $P < 2.16 \times 10^{-10}$  were enriched upstream of the gene ( $P < 0.01$ ). The difference in the degree of significance implied that downstream and upstream signals may act as regulatory sequences with different importance. The association

signal of rs6922 ( $P = 1.26 \times 10^{-54}$ ) was located far from the target gene (SNHG5; 181,402 bp apart), and thus it was suspected as an enhancer variant. This speculation led us to compare it with the sequences resulting from a ChIA-PET study.<sup>[11]</sup> As a result, we found that 1 anchor of an interacting PET cluster included rs6922, and the other anchor was the promoter of the SNHG5 gene. The current study also showed that nucleotide substitution from G to T at rs6922 decreased the expression of SNHG5 (beta = -53.7). The SNPs with  $P < 2.16 \times 10^{-10}$  were enriched in the promoter and enhancer regions of the SNHG5 gene ( $P < 0.01$ ; 7 SNPs in the promoter and 23 SNPs in the enhancer regions).

Nucleotide substitution of such variants might influence the expression of SNHG5 gene and is candidate genetic factor for susceptibility to cancer. This is supported by previous studies in which a variety of cancers were induced by low expression of snoRNAU50, a product generated from introns 4 and 5 of the SNHG5 gene. Expression of U50 is down-regulated in prostate



**Figure 1.** Genetic associations of nucleotide variants with mRNA expression of SNHG5 (A) and PEX6 (B). The blue line indicates the significance threshold with  $P < 2.16 \times 10^{-10}$ , and the red line indicates the significance threshold with  $P < 10^{-100}$ . The green bar and the black arrow indicate region and direction of each gene, respectively.

and breast cancer samples.<sup>[14,15]</sup> These studies also showed that the expression of U50 can inhibit colony formation of prostate and breast cancer cells. SNHG5 gene is located at the chromosomal translocation breakpoint involved in B-cell lymphoma (provided by RefSeq, July 2008). It was recently reported that the serum levels of SNHG5 were up-regulated in patients with malignant melanoma, and that SNHG5 may function in melanomagenesis or melanoma metastasis.<sup>[16]</sup> Further studies are needed to understand the underlying mechanisms of nucleotide substitutions on cancers and their subtypes.

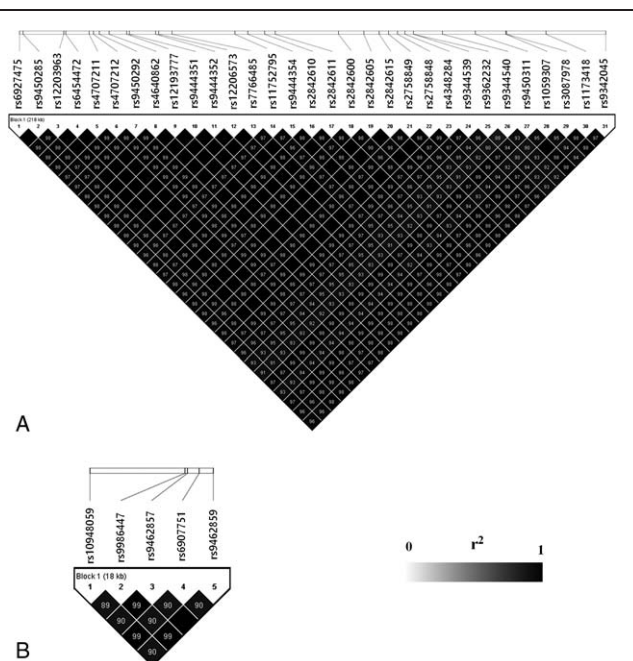
The current study revealed many intragenic eQTLs for the PEX6 gene. The gene encodes a member of the ATPases

associated with diverse cellular activities. In particular, the PEX6 plays a key role in biogenesis of peroxisome, which is a small eukaryotic organelle specialized to perform oxidative reactions. Its mutations cause peroxisome biogenesis disorders of complementation groups 4 and 6.<sup>[17]</sup> This suggests that the eQTLs detected from the current study might influence on susceptibility to the peroxisome biogenesis disorders by abnormal expression of PEX6.

Two of the eQTLs for PEX6 had been identified for association with prostate cancer from 2 previous studies which revealed rs10948059 and rs9462856 linked to it in European populations, respectively.<sup>[18,19]</sup> These findings were interpreted that the genetic associations with prostate cancer were produced by regulating the expression of the glycine N-methyltransferase (GNMT) gene that overlaps with the PEX6 gene. The authors suspected the GNMT gene because the rs10948059 and rs9462856 were their intronic variants. Since the nucleotide variants were revealed as a strong eQTL for PEX6 gene from the current study ( $P = 6.98 \times 10^{-108}$ ), the variants associated with prostate cancer might be identified through expressional change in PEX6. This was supported by the results of the Human Protein Atlas in which PEX6 is down-regulated in prostate cancer cells (<http://www.proteinatlas.org/ENSG00000124587-PEX6/cancer>).

We found that the strong eQTLs ( $P < 10^{-100}$ ) using additive model (“AA” vs “AB” vs “BB,” where A is major allele and B is minor allele<sup>[20]</sup>) were all significant using dominance (“AA + AB” vs “BB”) and recessive (“AA” vs “AB + BB”) models ( $P < 2.16 \times 10^{-10}$ ). However, they were not as strong as observed using the additive model, and all showed  $P > 10^{-100}$ . The best fit into the additive model implied that the expressional regulation of eQTLs was correlated with the number of minor alleles.

The current study also identified eQTLs with  $10^{-100} < P < 2.16 \times 10^{-10}$ . In particular, some eQTLs for PEX6 (rs9471950, rs9471951, rs7754294, rs13203402, rs7740252, rs7758576, and rs7759112;  $5.63 \times 10^{-12} < P < 1.91 \times 10^{-10}$ ) turned out to be enhancer variants of the serum response factor (SRF) gene, which resulted from chromatin interaction analysis with paired-end-tag using RNA polymerase II.<sup>[11]</sup> The regulatory variants could be plausible because SRF is a well-known transcript factor for the PEX6 gene.



**Figure 2.** Linkage disequilibrium blocks for strongly significant SNPs ( $P < 10^{-100}$ ) in and around the SNHG5 (A) and PEX6 (B) genes.

In this study, we employed a mixed model incorporating pair-wise polygenic relationship among individuals to control population stratification, which may reduce false-positive eQTLs.<sup>[8]</sup> The analysis revealed strong cis-acting eQTLs for SNHG5 and PEX6, and offered insights about their cellular mechanism of transcriptome variation. However, this study was limited to gene expression in lymphoblastoid cells. Other strong eQTLs would be identified with different cell types. In addition, this study had another limitation in understanding specific functions of eQTLs in gene expression (eg, splicing regulation and transcript stability).

In conclusion, the current study revealed novel eQTLs for SNHG5 and PEX6 genes in chromosome 6. Nucleotide substitutions of the eQTLs might be candidate factors for a variety of cancers by regulating expression of the 2 genes. Further studies are needed to understand their underlying mechanisms.

## References

- [1] Mungall AJ, Palmer SA, Sims SK, et al. The DNA sequence and analysis of human chromosome 6. *Nature* 2003;425:805–11.
- [2] Borchers CH, Kast J, Foster LJ, et al. The Human Proteome Organization Chromosome 6 Consortium: integrating chromosome-centric and biology/disease driven strategies. *J Proteomics* 2014;100:60–7.
- [3] Flicek P, Amode MR, Barrell D, et al. Ensembl 2011. *Nucl Acids Res* 2011;39(Database issue):D800–6.
- [4] Communications LHNCfB: conditions related to genes on chromosome 6. Lister Hill National Center for Biomedical Communications; 2012.
- [5] Okada Y, Wu Di, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
- [6] Cai Q, Wen W, Qu S, et al. Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of Chinese, Japanese, and European ancestry. *Cancer Res* 2011;71:1344–55.
- [7] Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11.
- [8] Shin J, Lee C. Statistical power for identifying nucleotide markers associated with quantitative traits in genome-wide association analysis using a mixed model. *Genomics* 2013;105:1–4.
- [9] 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.
- [10] Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
- [11] 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.
- [12] McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
- [13] The GTEx Consortium The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60.
- [14] Dong XY, Rodriguez C, Guo P, et al. SnoRNA U50 is a candidate tumor suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* 2008;17:1031–42.
- [15] Dong XY, Guo P, Boyd J, et al. Implication of snoRNA U50 in human breast cancer. *J Genet Genomics* 2009;36:447–54.
- [16] Ichigozaki Y, Fukushima S, Jinnin M, et al. Serum long non-coding RNA, snoRNA host gene 5 level as a new tumor marker of malignant melanoma. *Exp Dermatol* 2016;25:67–9.
- [17] Matsumoto N, Tamura S, Moser A, et al. The peroxin Pex6p gene is impaired in peroxisomal biogenesis disorders of complementation group 6. *J Hum Genet* 2001;46:273–7.
- [18] Chen M, Huang YL, Huang YC, et al. Genetic polymorphisms of the glycine-N-methyl transferase and prostate cancer risk in the health professionals follow-up study. *PLoS One* 2014;9:e94683.
- [19] Ianni M, Porcellini E, Carbone I, et al. Genetic factor regulating inflammation and DNA methylation associated with prostate cancer. *Prostate Cancer Prostatic Dis* 2013;16:56–61.
- [20] Zhao F, Song M, Wang Y, et al. Genetic model. *J Cell Mol Med* 2016;20:765.