GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Relationship of common expression quantitative trait loci genes to the immune system

J. Li[1,2,3]*, L. Wang[3]* , H. Li[3] *, R. Zhang[3], X. Li[3] and M. Guo[2]

[1]School of Life Science and Technology, Harbin Institute of Technology, Harbin, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[3]Department of Bioinformatics and Computer, Harbin Medical University, Harbin, China

*These authors contributed equally to this study
Corresponding authors: X. Li / M. Guo
E-mail: lixia6@yahoo.com / maozuguo@hit.edu.cn

Genet. Mol. Res. 12 (4): 6546-6553 (2013)
Received January 15, 2013
Accepted September 20, 2013
Published December 11, 2013
DOI http://dx.doi.org/10.4238/2013.December.11.6

ABSTRACT. Tissue-specific expression quantitative trait loci (eQTLs) are always linked to specific diseases. In this study, we focused on eQTLs common to multiple tissues and explored their functional mechanisms in disease for the first time. We found 11 common eQTL genes among multiple tissues. Five genes were validated through a genome-wide association study, 3 genes were validated using the Online Mendelian Inheritance in Man database, and the others were validated through text mining. Most of these disorders were related to the systemic immune system. In functional analyses using Gene Ontology and the Kyoto Encyclopedia of Genes and Genomes, these common eQTL genes were enriched in biological processes and pathways mostly related to the

immunity. Therefore, we believe that these common eQTL genes are related to immune system.

Key words: Multiple tissues; Expression quantitative trait locus; Immune system

INTRODUCTION

Recent developments in high-throughput genotyping technology, next-generation sequencing technologies, and genome-wide association studies (GWAS) have been widely applied to find loci and genes related to complex diseases such as coronary heart disease (McPherson et al., 2007), type 2 diabetes (Sladek et al., 2007), Crohn's disease (WTCCC, 2007), and host control of human immunodeficiency virus (HIV) type 1 (Fellay et al., 2007). However, explorations of the relationship between DNA sequence variations and phenotypes in GWAS ignore many middle molecular phenotypes, such as gene transcript levels and protein expression levels. These molecular phenotypes influenced by DNA sequence variation may also induce disease (Schadt et al., 2003; Nicolae et al., 2010). Expression quantitative trait locus (eQTL) mapping is a technique that uses genomic variation to explain expression traits. Using the level of gene expression as a middle molecular phenotype between genetic variation and clinical phenotypes in eQTLs, it provides a new way to mine risk genes and their functional mechanisms in studies of complex traits, especially complex diseases (Schadt et al., 2003; Nicolae et al., 2010). Therefore, integration of eQTL and GWAS genetic approaches not only reduce the dimensions of risk single-nucleotide polymorphism (SNP) loci in GWAS, but also effectively locate complex disease susceptibility genes and assist in functional verification of candidate genes.

Nica et al. (2011) and van Nas et al. (2010) have proven that tissue-specific eQTLs are always linked to specific diseases, so most eQTL studies use specific tissue samples - such as liver tissue samples for cirrhosis (Schadt et al., 2008) and brain tissue samples for Alzheimer's disease (Webster et al., 2009) - to investigate a disease. In this study, we focused on the common eQTLs among multiple tissues and explored the functional mechanisms of these common eQTLs in disease processes for the first time. We integrated 5 eQTL analyses from various tissues, found 11 significant common eQTL genes, and performed functional analysis, such as validation through GWAS, Online Mendelian Inheritance in Man (OMIM), text mining, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis.

MATERIAL AND METHODS

We used 5 eQTLs datasets from liver tissue samples (Schadt et al., 2008), normal brain tissue samples (Webster et al., 2009), cohort studies of Alzheimer's disease brain tissue samples (Myers et al., 2007), lymphoblastoid cell line samples (Montgomery et al., 2010), and monocyte samples (Zeller et al., 2010). Genome-wide genotyping and gene expression data are present in the sets. Following the statistical methods, significant thresholds, and multiple testing correction methods in the original literature, we selected significant eQTLs in these datasets for subsequent analyses. Detailed information for these datasets is shown in Table 1.

Table 1. Summary of the 5 eQTL datasets.

Tissues	Liver	Brain (Myers)	Brain (Webster)	Lymphoblastoid cell lines (Mon)	Monocytes
Sample number	427	193	364	63	1490
Race	Caucasian white	European descent	United States	European descent	Gutenberg, Germany
SNP number	782476	366140	380157	1200000	29912
Transcript number	39280	14078	8650	15967	2745
Genome genotyping method	Affymetrix 500K panel/ Illumina 650Y panel	Affymetrix GeneChip Human Mapping 500K Array Set	Affymetrix GeneChip Human Mapping 500K Array Set	Illumina WG-6 version 2 array	Affymetrix Genome-Wide Human SNP Array 6.0
Call rate	>75%	>90%	>90%	NA	>98%
MAF	>4%	>1%	>1%	>5%	>1%
HWE	0.0001	0.05	0.05	0.05	0.0001
Gene expression chip	CDNA(Cy3/Cy5)	Illumina Human Refseq-8 Expression BeadChip	Illumina Human Refseq-8 Expression BeadChip	Illumina WG-6 version 2 array	The Illumina HT-12 BeadChip
Gene expression data preprocessing	R (rlm function)	BeadStudio software	BeadStudio software	FluxCapacitor	BeadStudio software
eQTL test method	Kruskal-Wallis H test	PLINK analysis toolset	PLINK analysis toolset	Spearman rank correlation analysis	TAMU ANOVA
Multiple test adjustment method	FDR	WARD	P value with 1000 permutation	P value with 1000 permutation	Family-wise error rate
Cis eQTL threshold value	5.00E-05	0.05	0.05	0.01	0.05
Trans eQTL threshold value	1.00E-08	0.05	0.05	NA	0.05
Significant eQTL number	6754	769	665	6599	37416
Significant cis eQTL number	4265	433	189	6599	34197
Significant trans eQTL number	2489	336	476	NA	3219
Gene number in significant eQTL	6108	258	432	1135	2747
SNP number in significant eQTL	3419	676	462	4230	29923

We unified a gene to be the official gene symbol. We directly extracted the data from the liver tissue dataset and 2 brain tissue datasets, as they used official gene symbols. In the data from the lymphoblastoid cell line samples, we used the David Gene ID Conversion Tool (Huang et al., 2009) to convert the Ensemble gene IDs to their corresponding official gene symbols; in the data of the monocyte samples, we extracted the official gene symbol data from the Sqlite3Explorer database provided in the original literature. We used reference SNP identifiers to determine SNPs.

We divided the significant eQTLs into cis and trans eQTLs. Molecular variation located on or near the genomic position of the regulated gene (usually within the selected gene or the range of less than 1 Mb from the 5' or 3' end) was identified as a cis eQTL; otherwise it was labeled a trans eQTL (Rockman and Kruglyak, 2006). In this study, we set 1 Mb as the threshold to separate cis and trans eQTLs. We only selected the cis eQTLs for subsequent analyses, as they have higher credibility and a more direct mechanism of action than that of trans eQTLs.

RESULTS

Intersection of the 5 eQTL gene sets

Because the 5 datasets used different chips and the diversities of the measured SNPs are large, we could not directly screen common SNPs among them. However, the most frequently detected genes are common, so we focused on the intersection of genes regulated by SNPs (said eQTL genes) among various tissues. Two datasets of brain tissue samples were included, so we first combined them and identified 658 eQTL genes. We then intersected them with 3 additional eQTL gene datasets to obtain a final 11 common eQTL genes among 4 tissues. These genes are listed in Table 2.

Table 2. Number of eSNPs in the 5 datasets and SNPs located on the genes.

Gene symbol	Number of eSNPs in the 5 datasets	Number of SNPs located on the gene	Number of joint SNPs
ATPIF1	10	71	1
CWF19L1	27	271	2
H1F0	15	63	0
LRRC14	11	83	2
HLA-A	177	2750	1
HLA-B	311	2306	4
HLA-DRB1	453	4933	1
RPS26	35	77	1
SIRT1	13	318	0
SLC4A7	14	954	0
ZNF266	47	406	8

Validation through GWAS, OMIM, and text mining

eQTL genes have been proven to be strongly related to genes found through GWAS, so we can validate more accurate disease-related genes by combining eQTL and GWAS (Nicolae et al., 2010). We took these 11 genes into the GWAS catalog (Hindorff et al., 2009, 2011) to search for SNPs using a default threshold P value of 10^{-5}. The results are shown in Table 3. We confirmed that 5 of these 11 genes are related to complex diseases through GWAS. In particular, the strongest SNP loci rs2523608 and rs2523590 on major histocompat-

ibility complex, class I, B (HLA-B) were associated with HIV-1 control (Pelak et al., 2010; Pereyra et al., 2010), and they are also linked to HLA-B gene expression in lymphoid cell lines (Montgomery et al., 2010). Some research has found that a variety of HLA-B gene subtypes is related to the progression of HIV infection (Gao et al., 2001; Carrington and O'Brien, 2003). We also confirmed that loci rs9268853 and rs9271100, which are associated with ulcerative colitis (Pereyra et al., 2010) and systemic lupus erythematosus (Han et al., 2009; Alcina et al., 2012), affect the expression of HLA class II, DR beta 1 (HLA-DRB1) in lymphoid cell lines (Montgomery et al., 2010). The results showed that the eQTL genes shared among tissues are consistent with disease risk genes and loci in GWAS.

Table 3. Gene-related diseases and risk loci in GWAS.

Gene symbol	GWAS disease/trait	Strongest risk-SNP
HLA-A	Drug-induced liver injury (amoxicillin-clavulanate)	rs2523822
	Adverse response to carbamapezine	rs1061235
	Nasopharyngeal carcinoma	rs2860580/rs2517713
	Vitiligo	rs6904029
	HIV-1 control	rs7758512
HLA-B	Nasopharyngeal carcinoma	rs2894207
	Ankylosing spondylitis	rs4349859
	Vitiligo	rs11966200
	HIV-1control	rs2523608/rs2395029/rs2523590/rs9264942
	Multiple sclerosis	rs2523393
	Drug-induced liver injury (flucloxacillin)	rs2395029
	Height	rs13437082
	Type 1 diabetes	rs2524054
HLA-DRB1	Rheumatoid arthritis	rs7765379/rs13192471/rs6910071/rs9272219/rs6457620
	Ulcerative colitis	rs9268853/rs2395185
	Systemic lupus erythematosus	rs9271100
	Nephropathy	rs9275596
	Type 1 diabetes	rs2647044
	Systemic sclerosis	rs3129763
	Inflammatory bowel disease	rs9271366/rs2006996
	Response to interferon beta therapy	rs9272105
	Drug-induced liver injury (amoxicillin-clavulanate)	rs9274407
	Chronic lymphocytic leukemia	rs674313
	Immunoglobulin A	rs9271366/rs2187668
	Lumiracoxib-related liver injury	rs3129900
	Multiple sclerosis	rs2040406/rs9271366/rs3135388/rs3129934
RPS26	Psoriasis	rs12580100
	Type 1 diabetes	rs11171739
SLC4A7	Breast cancer	rs4973768

To explore further whether the mutations or deletions of these genes lead to genetic diseases, we queried these genes in OMIM (Hamosh et al., 2000). The results are shown in Table 4. We used ankylosing spondylitis as an example. Several studies (Monnet et al., 2004; Gu et al., 2009) have shown that HLA-B27 is a risk factor for ankylosing spondylitis. Because the HLA system is genetically determined, the genetic pathogenesis of ankylosing spondylitis is an important factor. In addition to genetic factors, immune disorders and systemic inflammation of the body are related to spondyloarthropathy. Furthermore, some research has confirmed that HLA-B27 is associated with other immune system and inflammatory diseases, such as psoriasis (Sheehan, 2004), acute anterior uveitis (Chang et al., 2005), and Crohn's disease (Purrmann et al., 1988; Khan, 1989).

We adopted text mining for another 7 genes and confirmed that ATPase inhibitory factor 1 is related to tumor formation (Sanchez-Cenizo et al., 2010); H1 histone family member 0

is related to systemic lupus erythematosus (Armananzas et al., 2009); Sirtuin 1 forms a pathway with microRNA-134 to regulate memory and plasticity (Gao et al., 2010) and is related to insulin resistance (Liang et al., 2009) and cancer, as demonstrated previously in a mouse model (Herranz et al., 2010); and a segment of leucine rich repeat containing 14 is related to bladder cancer (Rothman et al., 2010), schizophrenia, bipolar disorder (Wang et al., 2010), and attention deficit hyperactivity disorder (Lasky-Su et al., 2008). The results of analyses showed that the 11 eQTL genes common to multiple tissues were related with a variety of diseases or biological traits - mostly disorders of the systemic immune system such as cancers and autoimmune diseases.

Table 4. Results queried in OMIM of the eQTL genes.

Gene/locus	Phenotype	Gene/locus MIM number	Location
HLA-A	Severe cutaneous adverse reaction	142800	6p21.3
HLA-B	Abacavir hypersensitivity	142830	6p21.33
	Drug-induced liver injury due to flucloxacillin		
	Spondyloarthropathy		
	Stevens-Johnson syndrome, carbamazepine-induced		
	Synovitis, chronic		
RPS26	Diamond-Blackfan anemia 10	603701	12q13.2

GO and KEGG functional enrichment analysis

To explore further the biological processes in which these 11 genes are involved, we performed GO and KEGG functional enrichment analysis using the DAVID Functional Annotation Tool (Huang et al., 2009). The significant results are detailed in Table 5 with a P value of <0.05 as the threshold.

Table 5. Results of gene ontology (GO) and KEGG functional enrichment analysis of these 11 genes.

GO ID or KEGG ID	Term name	Genes	P value
GO:0019882	Antigen processing and presentation	HLA-A, HLA-B, HLA-DRB1	0.001
GO:0002474	Antigen processing and presentation of peptide antigen via MHC class I	HLA-A, HLA-B	0.010
GO:0048002	Antigen processing and presentation of peptide antigen	HLA-A, HLA-B	0.016
GO:0051346	Negative regulation of hydrolase activity	ATPIF1, SIRT1	0.028
GO:0045934	Negative regulation of nucleobase, nucleoside, nucleotide, and nucleic acid	ATPIF1, SIRT1, RPS26	0.034
GO:0051172	Negative regulation of nitrogen compound metabolic process	ATPIF1, SIRT1, RPS26	0.035
hsa05330	Allograft rejection	HLA-A, HLA-B, HLA-DRB1	1.46E-4
hsa05332	Graft-versus-host disease	HLA-A, HLA-B, HLA-DRB1	1.71E-4
hsa04940	Type I diabetes mellitus	HLA-A, HLA-B, HLA-DRB1	1.99E-4
hsa05320	Autoimmune thyroid disease	HLA-A, HLA-B, HLA-DRB1	2.94E-4
hsa05416	Viral myocarditis	HLA-A, HLA-B, HLA-DRB1	5.72E-4
hsa04612	Antigen processing and presentation	HLA-A, HLA-B, HLA-DRB1	7.81E-4
hsa04514	Cell adhesion molecules	HLA-A, HLA-B, HLA-DRB1	1.09E-3

In the GO analysis, these genes were enriched in the categories antigen processing and presentation; negative regulation of hydrolase activity; negative regulation of nucleobase, nucleoside, nucleotide, and nucleic acid; and negative regulation of nitrogen compound metabolic process. In the KEGG pathway analysis, these genes, especially HLA-A, HLA-B, and HLA-DRB1 were enriched in the categories allograft rejection, graft-versus-host disease, type I diabetes mellitus, autoimmune thyroid disease, viral myocarditis, and antigen processing and presentation. Clearly,

these biological processes and pathways are primarily related to the immune system.

DISCUSSION

In this study, we analyzed 11 common eQTL genes among 5 tissues. Among them, 5 genes were validated to be associated with complex diseases through GWAS, 3 genes were validated to be associated with diseases or biological traits through OMIM analysis, and others were validated with text mining. Most of the associations were disorder related to the systemic immune system - for example, cancers and autoimmune diseases. In the GO and KEGG functional enrichment analyses, these genes - especially HLA-A, HLA-B, and HLA-DRB1 - were enriched in biological processes and pathways such as antigen processing and presentation, allograft rejection, and autoimmune thyroid disease that are mostly related to the immune system.

Therefore, we believe that these eQTL genes common to multiple tissues are most likely to be related to immune system. However, few genome-wide eQTL analysis datasets are available, and we used only 1 or 2 datasets for each tissue. When more genome-wide eQTL datasets are available, we will perform deeper analysis for common eQTL genes.

Conflicts of interest

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

Research partially supported by the National Natural Science Foundation of China (grants #61300116, #61271346, #91335112, #60932008, and #61172098), the Specialized Research Fund for the Doctoral Program of Higher Education of China (grant #20112302110040), the Natural Science Foundation of Heilongjiang Province (grant #QC2013C063), the Fund of Heilongjiang Education Department (grants #12531298 and #12531299), the Fund of Heilongjiang Health Department (grants #2011-204 and #2011-251).

REFERENCES

Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, et al. (2012). Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One* 7: e29819.

Armananzas R, Calvo B, Inza I, Lopez-Hoyos M, et al. (2009). Microarray analysis of autoimmune diseases by machine learning procedures. *IEEE Trans. Inf. Technol. Biomed.* 13: 341-350.

Carrington M and O'Brien SJ (2003). The influence of HLA genotype on AIDS. *Annu. Rev. Med.* 54: 535-551.

Chang JH, McCluskey PJ and Wakefield D (2005). Acute anterior uveitis and HLA-B27. *Surv. Ophthalmol.* 50: 364-388.

Fellay J, Shianna KV, Ge D, Colombo S, et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science* 317: 944-947.

Gao J, Wang WY, Mao YW, Graff J, et al. (2010). A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature* 466: 1105-1109.

Gao X, Nelson GW, Karacki P, Martin MP, et al. (2001). Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N. Engl. J. Med.* 344: 1668-1675.

Gu J, Huang J, Li C, Zhao L, et al. (2009). Association of chromosome 2q36.1-36.3 and autosomal dominant transmission in ankylosing spondylitis: results of genetic studies across generations of Han Chinese families. *J. Med. Genet.* 46: 657-662.

Hamosh A, Scott AF, Amberger J, Valle D, et al. (2000). Online mendelian inheritance in man (OMIM). *Hum. Mutat.* 15: 57-61.

Han JW, Zheng HF, Cui Y, Sun LD, et al. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* 41: 1234-1237.

Herranz D, Munoz-Martin M, Canamero M, Mulero F, et al. (2010). Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. *Nat. Commun.* 1: 3.

Hindorff L, MacArthur J, Wise A and Junkins H (2011). A Catalog of Published Genome-Wide Association Studies. Available at [www.genome.gov/gwastudies]. Accessed June 19, 2012.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 106: 9362-9367.

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

Khan MA (1989). HLA-B27 and B12(B44) in Crohn's disease with ankylosing spondylitis. *J. Rheumatol.* 16: 851-852.

Lasky-Su J, Neale BM, Franke B, Anney RJ, et al. (2008). Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am. J. Med. Genet. B: Neuropsychiatr. Genet.* 147B: 1345-1354.

Liang F, Kume S and Koya D (2009). SIRT1 and insulin resistance. *Nat. Rev. Endocrinol.* 5: 367-373.

McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* 316: 1488-1491.

Monnet D, Breban M, Hudry C, Dougados M, et al. (2004). Ophthalmic findings and frequency of extraocular manifestations in patients with HLA-B27 uveitis: a study of 175 cases. *Ophthalmology* 111: 802-809.

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, et al. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464: 773-777.

Myers AJ, Gibbs JR, Webster JA, Rohrer K, et al. (2007). A survey of genetic human cortical gene expression. *Nat. Genet.* 39: 1494-1499.

Nica AC, Parts L, Glass D, Nisbet J, et al. (2011). The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* 7: e1002003.

Nicolae DL, Gamazon E, Zhang W, Duan S, et al. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS. Genet.* 6: e1000888.

Pelak K, Goldstein DB, Walley NM, Fellay J, et al. (2010). Host determinants of HIV-1 control in African Americans. *J. Infect. Dis.* 201: 1141-1149.

Pereyra F, Jia X, McLaren PJ, Telenti A, et al. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330: 1551-1557.

Purrmann J, Zeidler H, Bertrams J, Juli E, et al. (1988). HLA antigens in ankylosing spondylitis associated with Crohn's disease. Increased frequency of the HLA phenotype B27, B44. *J. Rheumatol.* 15: 1658-1661.

Rockman MV and Kruglyak L (2006). Genetics of global gene expression. *Nat. Rev. Genet.* 7: 862-872.

Rothman N, Garcia-Closas M, Chatterjee N, Malats N, et al. (2010). A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.* 42: 978-984.

Sanchez-Cenizo L, Formentini L, Aldea M, Ortega AD, et al. (2010). Up-regulation of the ATPase inhibitory factor 1 (IF1) of the mitochondrial H+-ATP synthase in human tumors mediates the metabolic shift of cancer cells to a Warburg phenotype. *J. Biol. Chem.* 285: 25308-25313.

Schadt EE, Monks SA, Drake TA, Lusis AJ, et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297-302.

Schadt EE, Molony C, Chudin E, Hao K, et al. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS. Biol.* 6: e107.

Sheehan NJ (2004). The ramifications of HLA-B27. *J. R. Soc. Med.* 97: 10-14.

Sladek R, Rocheleau G, Rung J, Dina C, et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445: 881-885.

van Nas A, Ingram-Drake L, Sinsheimer JS, Wang SS, et al. (2010). Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* 185: 1059-1068.

Wang KS, Liu XF and Aragam N (2010). A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr. Res.* 124: 192-199.

Webster JA, Gibbs JR, Clarke J, Ray M, et al. (2009). Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* 84: 445-458.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-678.

Zeller T, Wild P, Szymczak S, Rotival M, et al. (2010). Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 5: e10693.