

UCSF

UC San Francisco Previously Published Works

Title

Coanalysis of GWAS with eQTLs reveals disease-tissue associations.

Permalink

<https://escholarship.org/uc/item/9rj2k4pc>

Authors

Kang, Hyunseok Peter

Morgan, Alex A

Chen, Rong

et al.

Publication Date

2012

Peer reviewed

Coanalysis of GWAS with eQTLs reveals disease-tissue associations

Hyunseok Peter Kang, M.D.¹, Alex A. Morgan, Ph.D.¹, Rong Chen, Ph.D.¹,
Eric E. Schadt, Ph.D.², Atul J. Butte, M.D., Ph.D.¹

¹Division of Systems Medicine, Department of Pediatrics,
Stanford University School of Medicine, Stanford, CA

²Department of Genetics and Genome Sciences, Mount Sinai School of Medicine,
New York, NY

Abstract

Expression quantitative trait loci (eQTL), or genetic variants associated with changes in gene expression, have the potential to assist in interpreting results of genome-wide association studies (GWAS). eQTLs also have varying degrees of tissue specificity. By correlating the statistical significance of eQTLs mapped in various tissue types to their odds ratios reported in a large GWAS by the Wellcome Trust Case Control Consortium (WTCCC), we discovered that there is a significant association between diseases studied genetically and their relevant tissues. This suggests that eQTL data sets can be used to determine tissues that play a role in the pathogenesis of a disease, thereby highlighting these tissue types for further post-GWAS functional studies.

Introduction

Genome wide association studies (GWAS) of common complex or multifactorial diseases have proliferated enormously over the last few years. They have also been successful in identifying a large number of loci at extraordinary levels of significance. However, this success has presented a new challenge: translating these findings into a full understanding of how the loci affect complex disease traits. Most of the reported variants do not affect protein function in an obvious manner and indeed a large number lie in introns or intergenic regions, indicating that they may function through more subtle regulation of gene expression. This is compatible with the hypothesis that such genetic variants, commonly known as expression quantitative trait loci (eQTL), are an important factor in disease susceptibility (1).

eQTLs are discovered when samples are simultaneously studied using genotyping tools, yielding information on DNA variants, and expression measurement tools, yielding information on RNA levels. Each SNP can be represented by its alleles or genotypes, and statistically associated across samples with RNA levels. As a hypothetical example of one eQTL (Figure 1), as genotypes vary from CC, CT, TT at SNP rs7188573 in gene *MMP25*, expression levels of this gene decrease in monocytes. Note that the SNP may be within a certain distance from the gene, in which case the SNP is termed a *cis* eSNP, or might be distant, termed a *trans* eSNP. The degree of association is typically estimated statistically, using ANOVA or equivalent tests, controlled for multiple hypothesis testing. Nicolae, et al., recently demonstrated that SNPs associated with human traits are in general enriched for eQTLs (2). Other studies have shown that eQTLs calculated in the tissue of interest in a disease are enriched for disease-associated SNPs (3,4). These insights have been applied to use eQTLs to provide another layer of meaning to SNPs and prioritize GWAS results (5,6).

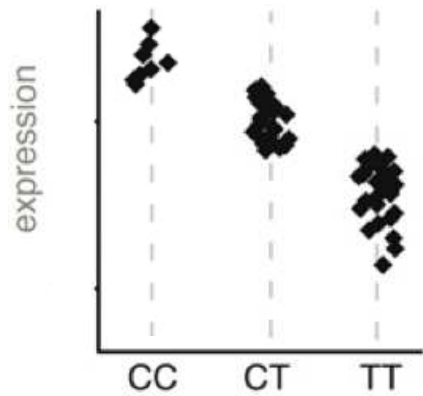


Figure 1. Hypothetical example of the different expression values of *MMP25* observed in monocytes with various genotypes at eSNP rs7188573.

Since gene expression profiles vary in different tissues, it is only to be expected that some eQTLs are tissue specific, and it has been reported that 33-69% of eQTLs, depending on the analysis method and the tissue type, are not discovered in other tissues (3,7,8), suggesting that the differences in eQTLs across tissues may provide additional functional information. This prompted us to investigate the correlation between the statistical significance of eQTLs calculated in different tissues (blood monocytes, liver, and adipose tissue) and their odds ratios in diseases studied by the Wellcome Trust Case Control Consortium (WTCCC) (9). Our results reveal the potential for this type of correlation analysis to be used to determine the tissue of interest for a disease, ultimately providing information that narrows the research focus or opening up novel avenues of inquiry.

Methods

The overall experimental design is shown in Figure 2. We obtained publicly-available eQTL data generated from 3 different tissue types in healthy Caucasians: peripheral blood monocytes (8), liver, and adipose tissue (10). The first data set specifically involved monocytes isolated from the blood of 1490 healthy individuals recruited in the Gutenberg Heart Study. The eQTLs were mapped using analysis of variance (ANOVA), with a p-value cutoff of $5.78E-12$, corresponding to a family-wise error rate of 0.05. This study reported 37,403 associations, comprising 29,912 SNPs and 2,745 expression traits. The liver and adipose datasets involved samples obtained from 1008 morbidly obese individuals at the time of gastric bypass surgery. Association was determined using the Kruskal-Wallis test, and results were reported at a 10% false discovery rate (FDR) based on permutations of the SNP genotypes and gene expression levels, for a total of 24,513 eSNPs associated with 15,241 transcripts or 9931 distinct genes.

To compare against these, we used the results of the genome wide association studies reported by the Wellcome Trust Case Control Consortium for 7 different diseases: bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HTN), rheumatoid arthritis (RA), type 1 diabetes (T1D), and type 2 diabetes (T2D).

We will refer to the p-value of eQTL association for an eSNP as p_{eqtl} , and the p-value of disease association for a given SNP as p_{gwas} . For each of the three tissue types for which we had eQTL relationships, we studied the relationship of minimum eSNP p-value (p_{eqtl}) to odds ratio against any of the seven diseases. For each disease studied by the WTCCC, we selected eSNPs in each tissue that were in very strong linkage disequilibrium (LD, $R^2 = 1$) with any SNP having uncorrected $p_{\text{gwas}} < 0.01$ for that disease. Because testing multiple SNPs in LD with each other could result in spurious correlation, we selected one SNP with the minimum p_{eqtl} from each LD block and performed Kendall rank correlation analysis of $-\log(p_{\text{eqtl}})$ and the absolute value of $\log(\text{disease odds ratio})$ using the *cor.test* function in R (11). The resulting p-values of 21 statistical tests were provided as input to the *qvalue* package in R to calculate corresponding q-values (12).

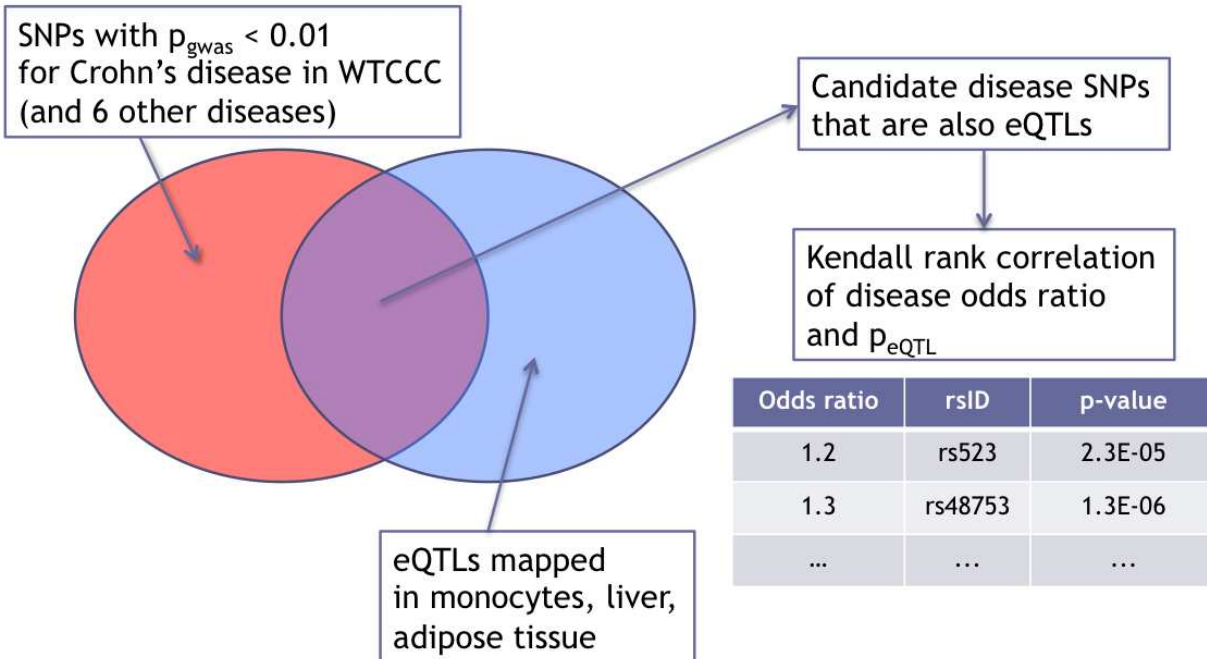


Figure 2. Experimental design.

Results

Our goal was to determine whether one can computationally assign a functional tissue to a disease, given only the genetic architecture found through GWAS, using tissue-specific eQTL data sets. As shown in Figure 2, we intersected two sets of variants, one from each disease studied by the WTCCC and the other from eSNPs reported in each tissue type.

The total numbers of eSNPs reported to be significant in each tissue, and the number that were associated with each disease at $p < 0.01$, are shown in Table 1. For the eSNPs listed in this table, we investigated the correlation between p_{eqtl} and disease odds ratio. For example, Figure 3 shows the rank of $\log(\text{odds ratio})$ plotted against $\log(p_{\text{eqtl}})$ for monocyte eSNPs and type 1 diabetes. The log of Kendall rank correlation p-value for each disease is shown in Figure 4.

Table 1. Number of eSNPs for each tissue type that are associated with each disease at $p < 0.01$.

Tissue type	Monocytes	Liver	Adipose
Total # of eSNPs	30,898	7380	15970
Disease	Number of eSNPs with $p < 0.01$ for disease		
Bipolar disorder	213	70	113
Coronary artery disease	183	57	108
Crohn's disease	211	66	128
Hypertension	182	49	117
Rheumatoid arthritis	328	84	139
Type 1 diabetes	481	113	206
Type 2 diabetes	178	72	119

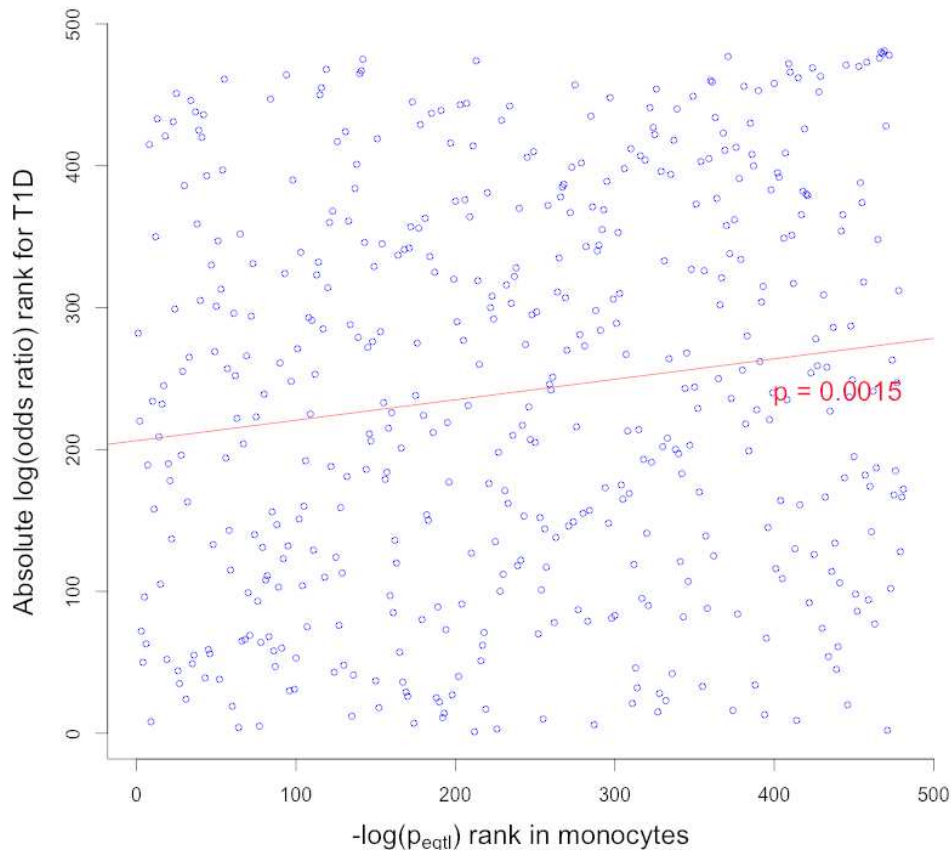


Figure 3. Plot of rank of absolute log(odds ratio) vs. $-\log(p_{\text{eqtl}})$ for monocyte eSNPs and type 1 diabetes.

For SNPs reported to be associated with gene expression in monocytes, we found that the correlation was significant at $p < 0.05$ in three diseases, Crohn's disease, rheumatoid arthritis, and type 1 diabetes. These are all autoimmune diseases in which macrophages play various roles. For example, there are recent reports that Crohn's disease may be a primary immunodeficiency of macrophages (13), and these cells are also involved in the pathogenesis of type 1 diabetes (14,15). Monocytes are also of central importance in rheumatoid arthritis (16). Closer inspection revealed several SNPs in the vicinity of -- but not within -- *CARD9* that were strongly associated with the expression of this gene. *CARD9* has previously been implicated in Crohn's disease, although the association was not replicated in the WTCCC study (17).

In liver tissue, we found that the correlation was significant for type 2 diabetes, and borderline for coronary artery disease. The liver is of course intimately involved in the metabolism of glucose and lipids, and liver pathology, such as nonalcoholic fatty liver disease seen in metabolic syndrome, is linked to coronary artery disease (18) and type 2 diabetes (19). We did not find any significant correlations in adipose tissue. The Kendall rank correlation statistics and q-values are summarized in Table 2. For the significant correlations, the q-values ranged from 0.025 to 0.205.

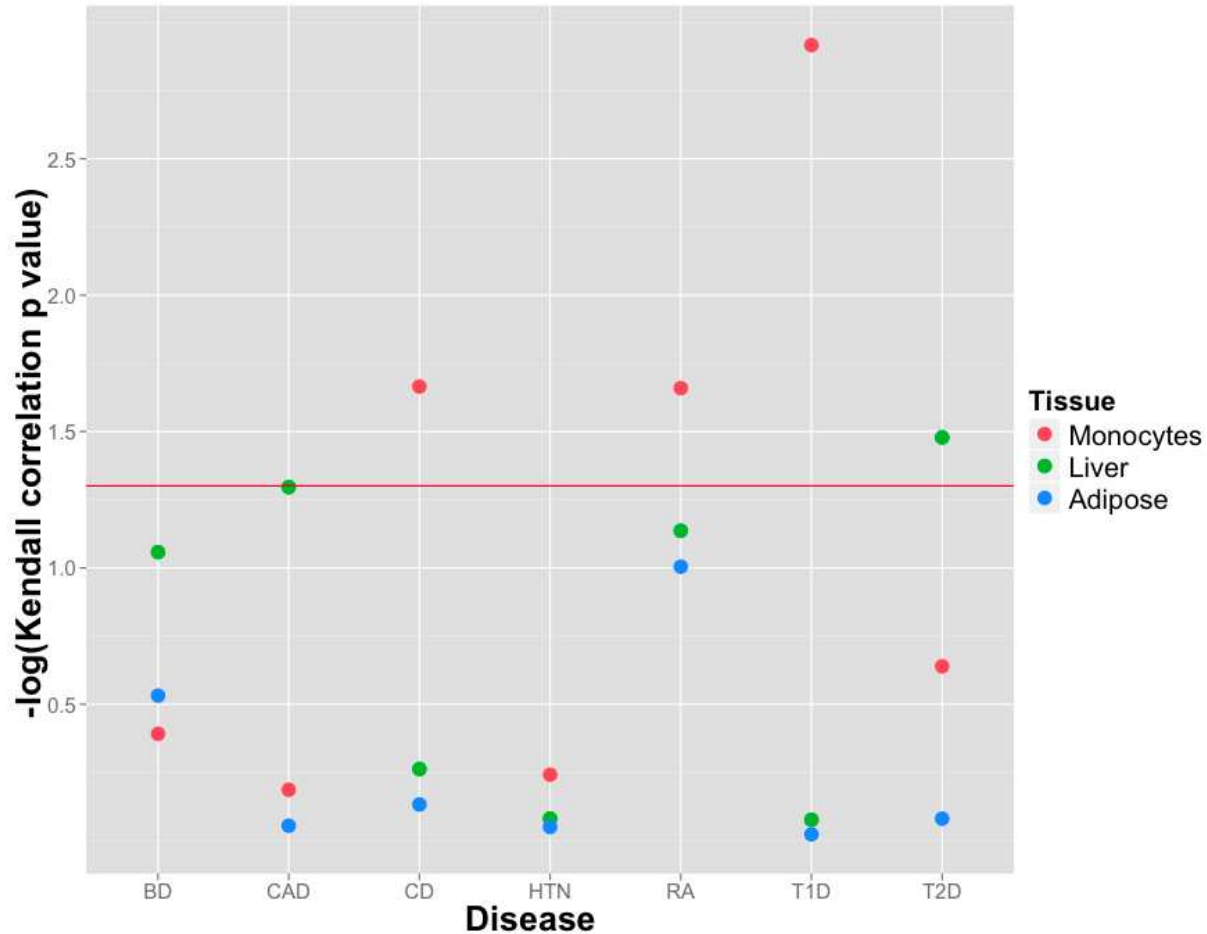


Figure 4. Correlation statistics for eSNP p-value and disease odds ratio. The red line represents a cutoff of $p = 0.05$.

Table 2. Summary of Kendall rank correlation statistics and q-values. Bold text indicates significant correlations.

Tissue type	Monocytes			Liver			Adipose		
	τ	P	q	τ	P	q	τ	P	q
Disease									
Bipolar disorder	-0.038	0.406	0.747	0.140	0.088	0.250	-0.067	0.294	0.594
Coronary artery disease	-0.023	0.651	0.902	0.178	0.0505	0.205	-0.010	0.882	0.902
Crohn's disease	0.106	0.022	0.148	0.051	0.546	0.892	0.020	0.737	0.902
Hypertension	0.028	0.573	0.892	-0.022	0.830	0.902	-0.009	0.891	0.902
Rheumatoid arthritis	0.085	0.022	0.148	0.133	0.073	0.246	0.094	0.099	0.250
Type 1 diabetes	0.099	0.001	0.025	-0.013	0.839	0.902	0.003	0.948	0.914
Type 2 diabetes	0.061	0.230	0.516	0.171	0.033	0.168	0.013	0.831	0.902

Discussion

We systematically investigated the correlation between the statistical significance of eQTLs calculated in different tissues and their odds ratios for different diseases studied by the WTCCC. Although several groups have used these ‘tissue eQTLs’ as an indicator of functional significance in interpreting GWAS results, to our knowledge, this is the first time the correlation of tissues to diseases has been systematically evaluated. If we consider eQTLs to be a marker of the functional significance of SNPs, our analysis reveals that in tissues that play a role in the disease this significance correlates with the disease odds ratio. These results suggest that the reverse may be true as well: tissues

in which there is a correlation between functional significance and odds ratio are more likely to play a role in that disease. While the statistical associations are not very strong, the false discovery rates of 0.025 to 0.205 indicate that the majority of them are not spurious. We plan to validate these findings by performing the analysis on additional eQTL and GWAS datasets. Ultimately, eSNP-gene relationships in the tissue of interest may reveal novel candidate disease SNPs, providing additional clues to the pathogenesis of the disease. For example, rs7698608, a SNP with $p = 0.0005$ for type 2 diabetes, is associated with the expression of *CISD2* in the liver. *CISD2* is a causative gene for Wolfram syndrome type 2, and a glucose intolerance phenotype has been observed in *CISD2*-/*C57BL/6* mice (20).

In conclusion, analyzing the correlation between tissue eQTL significance and disease odds ratio may provide another layer of tissue-specific information that can be used to decipher GWAS results. In order to facilitate this, more studies of tissue specific eQTLs must be performed and the data shared with the scientific community. Many researchers have been generous with their data, however the trend so far is to share mainly the p-values for the SNPs, and most of the datasets have not been collected in a centralized location. Providing the information equivalent to odds ratio in GWAS (the direction and magnitude of effect on gene expression) and submitting to a centralized database such as the GTEx (Genotype-Tissue Expression) eQTL Browser (<http://www.ncbi.nlm.nih.gov/gtex/test/GTEX2/gtex.cgi>) will help other investigators fully leverage the power of this data.

Acknowledgements

We thank Alex Skrenchuk and Mike Seda from Stanford University for computer support. This work was supported by Lucile Packard Foundation for Children's Health, the Hewlett Packard Foundation, and the National Library of Medicine (R01 LM009719 and T15 LM007033). This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

References

1. Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum. Genet.* 2003 Jul.;113(2):149–153.
2. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 2010;6(4):e1000888.
3. Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet.* 2010 Dec. 10;87(6):779–789.
4. Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF, et al. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Molecular Psychiatry.* 2011 Feb. 22;
5. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 2010 May;6(5):e1000932.
6. Fransen K, Visschedijk MC, van Sommeren S, Fu JY, Franke L, Festen EAM, et al. Analysis of SNPs with an effect on gene expression identifies *UBE2L3* and *BCL3* as potential new risk genes for Crohn's disease. *Human Molecular Genetics.* 2010 Sep. 1;19(17):3482–3488.
7. Michaelson J, Alberts R, Schughart K, Beyer A. Data-driven assessment of eQTL mapping methods. *BMC Genomics.* 2010 Sep. 17;11(1):502.
8. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS ONE.* 2010;5(5):e10693.

9. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007 Jun. 7;447(7145):661–678.
10. Greenawalt DM, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res*. 2011 May 20;
11. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
12. Storey JD. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003 Jul. 28;100(16):9440–9445.
13. Casanova J-L, Abel L. Revisiting Crohn's disease as a primary immunodeficiency of macrophages. *J. Exp. Med*. 2009 Aug. 31;206(9):1839–1843.
14. Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature*. 2010 Sep. 23;467(7314):460–464.
15. Eizirik DL, Colli ML, Ortis F. The role of inflammation in insulinitis and β -cell loss in type 1 diabetes. *Nat Rev Endocrinol*. 2009 Apr.;5(4):219–226.
16. Kinne RW, Stuhlmüller B, Burmester G-R. Cells of the synovium in rheumatoid arthritis. *Macrophages. Arthritis Res. Ther*. 2007;9(6):224.
17. Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet*. 2008 May;82(5):1202–1210.
18. Treeprasertsuk S, Lopez-Jimenez F, Lindor KD. Nonalcoholic Fatty Liver Disease and the Coronary Artery Disease. *Dig Dis Sci*. 2010 May 13;56(1):35–45.
19. Sung KC, Kim SH. Interrelationship between Fatty Liver and Insulin Resistance in the Development of Type 2 Diabetes. *Journal of Clinical Endocrinology & Metabolism*. 2011 Apr. 7;96(4):1093–1097.
20. Chen YF, Kao CH, Chen YT, Wang CH, Wu CY, Tsai CY, et al. Cisd2 deficiency drives premature aging and causes mitochondria-mediated defects in mice. *Genes & Development*. 2009 May 15;23(10):1183–1194.