Mol Cell Toxicol (2013) 9:103-111
DOI 10.1007/s13273-013-0014-3

ORIGINAL PAPER

Semantic networks for genome-wide CNV associated with AST and ALT in Korean cohorts

Hyo-Young Kim[1,2], **Jun-Hyung Park**[2], **Heebal Kim**[1] **& Byeong-Chul Kang**[2]

Received: 12 September 2012 / Accepted: 26 November 2012

Abstract Copy number variation (CNV) is an emerging approach to study about human health and diseases. Liver-related biochemical tests (aspartate aminotransferase:AST, alanine aminotransferase: ALT) are useful for diagnosing a patient with liver injury. We analyzed a CNV-based GWAS of AST and ALT in 407 Korean. Affymetrix Human 6.0 Array was used to identify CNV, and CNV segmentation was performed using CNV analysis software. Univariate linear regression was used for the GWAS using R package. We identified 64 CNVs associated with AST or ALT, and screened 228 genes located within our CNVs. In this study, we focused on semantic networks about liver disease using knowledge integration software. This semantic networks about liver disease contained entities like gene, disease, pathway, chemical, drug, and contained relationships between two entities like gene-pathway, gene-disease, pathway-chemical, disease-pathway, chemical-drug. Application of semantic networks shown three clusters, including four diseases (hepatocellular carcinoma, liver neoplasm, liver cell adenoma, drug-induced liver injury), one pathway (hepatitis C pathway), and seven drugs (acetaminophen, chlormezanone, stavudine, enflurane, isoniazid, mebendazole, nitisinone).

Keywords Copy number variation, Liver, Modeling, Network

[1]Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Science, Seoul National University, Kwan-ak St. 599, Kwan-ak Gu, Seoul 151-742, Korea
[2]Insilicogen Inc, #909, Venture valley, 958, Gosaek-dong, Gwonseon-gu, Suwon-si, Gyeonggi-do 441-813, Korea
Correspondence and requests for materials should be addressed to B. C. Kang (✉ byeongchul@gmail.com)

The liver, dark reddish brown color, is the second largest glandular organ in the body and is located under the lib on the right side. This organ performs many functions, including remove and detoxify harmful substances from blood, storage of glycogen, and absorb nutrients (http://www.britishlivertrust.org.uk; http://www.liverfoundation.org). Liver function test is blood tests to evaluate about patient's liver state[1]. Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are an important biochemical markers for evaluating of inflammation degree about liver injury[2]. ALT is an enzyme mainly found in hepatocytes, and AST is another hepatocellular enzyme. Elevated levels of ALT and AST are a sign of hepatic cells damage[3].

In the past few years, enormous efforts have been made to investigate the role of single nucleotide polymorphism (SNP) and copy number variation (CNV) in health and disease[4,5]. Differences of copy number between individuals contribute to alter in expression of genes sensitive to a disease susceptibility or dosage effect[6,7]. Genome-wide association studies of SNP and CNV are actively and fast studying to discover the genetic basis of common complex diseases such as cancer, cardiovascular disorder, and autism[8]. Gerber *et al*. (2012) identified rs6983267 variant was associated with colorectal cancer at a significant genome-wide level[9].

Data mining is provide important insights into the data with complicated and huge quantity. Theses semantic network have given researcher aids to semantic information answer about complex questions through integration of the available data[10,11]. While many data mining studies and software development have been progressed various fields, few have focused on data mining about CNV-association studies[12].

Many studies are reported about AST or ALT as biochemical marker to identify liver diseases. Also CNV

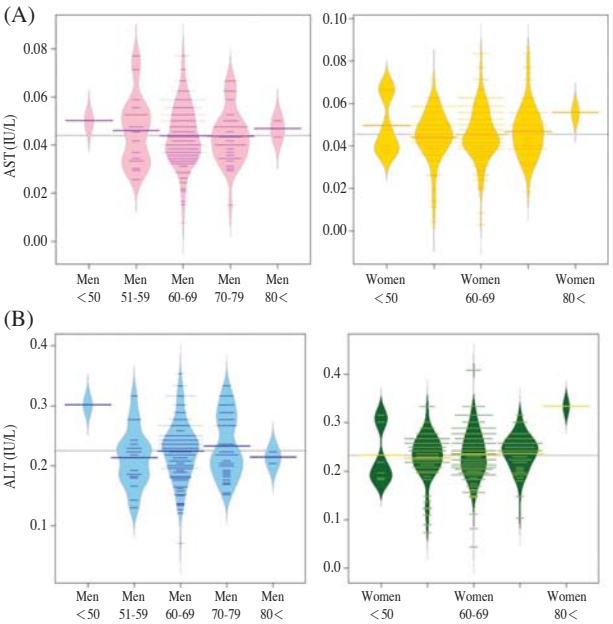

Figure 1. Beanplots for the distribution of AST (A) and ALT (B) in 407 Korean cohorts. The X-axis represents age in years.

analyses are emerging associated with human disease phenotypes. However, no studies identifying CNVs associated with biochemical marker for liver function test. Therefore, we focused on identifying a genome-wide copy numbers associated with AST or ALT in Korean population. Furthermore, we constructed semantic network by integrating datasets such as CNVs, gene, disease, chemical, drug and so on. We expect that this research provide valuable CNV-based genes and useful semantic network for CNVs associated with liver biochemical marker.

Discovery of CNV associated with AST or ALT

We focused on identify CNV-associated with AST or ALT as liver-related biochemical marker in 407 Korean cohorts. Figure 1 showed beanplots visualizing frequency distributions of individual observations for the AST and ALT on a one-dimensional scatterplot[13]. We did not show distribution differences between men and women for AST (A), but show distribution differences for ALT (B). Aged under 50 and older 80 were show highly gender differences. Therefore, we corrected age and gender to adjust the differences between samples in our statistical analysis. We assayed the genome-wide of copy-number variations on Affymetrix 6.0 arrays, and 3,046 CNVs were extracted using the copy-number analysis module option provided by Helix SVS software. Univariate linear regression analysis was identified 64 CNVs for AST and ALT at the $P < 0.05$ (Figure 2). The size of significant CNVs was 33 Mb and, which occupied about 10% of whole genome. The median size of CNVs was 115 Kb.

Genes enriched in CNV regions

We found a total of 228 genes fully located within our CNV regions (Table 1). To explore the biological meaningful of structural variations, we analyzed the gene set enrichment for the CNV regions. The top in biological process presented transcription from RNA polymerase II promoter, DNA-dependent transcription, and RNA biosynthetic process. Enriched functions in molecular functions were tropomyosin binding, including leukotriene B4 receptor activity, calcium ion binding, and carbonyl reductase activity. The most significantly enriched function in cellular components was intermediate filament, including intermediate filament cytoskeleton. Fifty percent of the pie was the proteasome activator complex.

Data integration for human liver disease

To answering complex questions, we integrated data sets for human liver disease, including entities such as gene, pathway, disease, chemical, drug, SNP, and CNV (Table 2). Figure 3 shows semantic network of integrated liver diseases. Gene entity was consisted of information, including Gene ID, NCBI Accession, position, curated integrating from UCSC Human Genome Browser and Comparative Toxicogenomics Database (CTD). Pathway, disease, chemical entities were mapped against CTD. The CTD is a public resource that advances understanding about chemical-gene-disease network on human health[14]. Drug was provided information for name, description, CAS number, indication, pharmacology, mechanism of action, toxicity, biotransformation, absorption, brands from DrugBank[15]. SNP was mapped against 1000 Genomes[16], and CNV was mapped against TCGA[17]. After that, to promote understanding about the mechanism of entities, the relations between two entities were created such as gene-pathway, gene-disease, gene-chemical, disease-chemical, disease-pathway, chemical-pathway, chemical-drug, gene-SNP, gene-CNV. Curated gene-pathway, gene-disease, gene-chemical, disease-chemical, disease-pathway, and chemical-pathway associations were mapped against from CTD. In particular, relation data with unknown interaction chemical-drug, gene-SNP, and gene-CNV was parsed using Python software. Curated associations are identified, and users helpful improve understanding about mechanisms about liver diseases.

180 genes shows associated with 4 liver diseases

Using the semantically integrated liver disease datasets, we analyzed the functional implications of structure

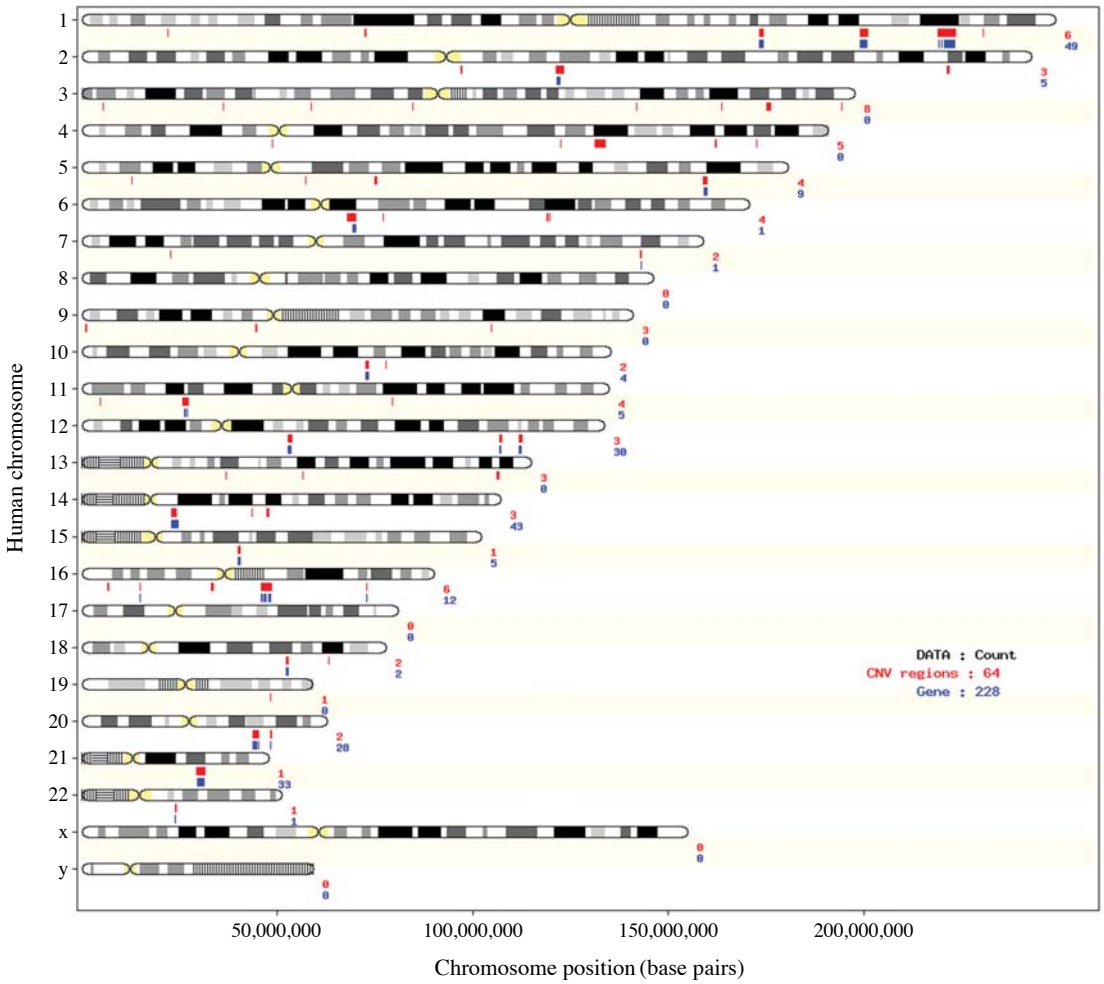

Figure 2. Visualization of the physical distribution of significant CNVs-associated with AST or ALT and genes fully located within the CNV regions on whole chromosome.

variants for AST and ALT. 180 genes were revealed associated with 4 liver diseases (hepatocellular carcinoma, liver neoplasm, liver cell adenoma, and drug-induced liver injury) in gene-disease interactions. Hepatocellular Carcinoma is a primary malignant liver neoplasm. It is the six most common cancer and the third leading cause of cancer death in the world[18]. Liver Neoplasms is an another name for hepatocellular carcinoma[19]. Liver Cell Adenoma is a hepatocellular benign epithelial tumor[20]. It occur most often in women who take higher dose of estrogen hormone pills. Since symptoms were generally not observed in patients, most was never detected. When discovered a large adenoma, it is surgically removed (http://www.liverfoundation.org). Drug-Induced Liver Injury (DILI) is known as hepatotoxicity. It caused by drugs agents and reactions, and more than 1000 drugs have been associated with significant hepatic injury[21]. DILI is classified into intrinsic and idiosyncratic types; intrinsic DILI is dose dependent whereas idiosyncratic DILI is not dose-related[22].

Hepatitis C pathway has been associated with AST or ALT

Analysis performed using the semantic integrated liver disease datasets identified one biochemical pathway relevantAST or ALT. Two genes (IRF9, OAS2) were revealed hepatitis C pathway (KEGG : 05160). Hepatitis C is a hepatitis C virus (HCV)-associated liver disease. HCV causes the liver to prevents its functions from working well, and is a main risk factor for hepatocellular carcinoma[23]. About 25% of people with HCV fully recover within six months, but about 75% of HCV-infected people develop chronic HCV, and chronic HCV can lead to cirrhosis, liver cancer, and liver

Table 1. The total number of 228 gene lists located within the CNV regions associated with AST or ALT.

Trait	Copy number regions	No. Genes	Genes
AST	Chr1:199155367-200961788	25	RNPEP,PTPN7,IGFN1,ELF3,PPP1R12B,UBE2T,SHISA4,SYT2, CACNA1S, TNNT2,TIMM17A,KIF21B,LAD1,NAV1,IPO9,GPR37L1,ARL8A,LGR6, LMOD1,PKP1,CSRP1,PHLDA3,TNNI1,TMEM9,RPS10P7
	Chr1:219016303-223441941	21	HHIPL2,SUSD4,DUSP10,TP53BP2,CNIH4,TLR5,DNAH14,C1orf65, TAF1A,MIA3,HLX,AIDA,NVL,WDR26,CNIH3,FAM177B,DISP1,FBXO28, DEGS1,CAPN2, CAPN8
	Chr10:72716105-73401893	4	CDH23,SLC29A3,C10orf54,PSAP
	Chr11:25870713-27225055	5	MUC15,SLC5A12,FIBIN,BBOX1,ANO3
	Chr15:39996964-40459070	6	PLA2G4E,VPS39,TMEM87A,PLA2G4D,PLA2G4F,GANC
	Chr16:14897352-14967222	1	NPIP
	Chr20:48228324-48562412	1	CEBPB
	Chr21:29429536-31660765	33	CLDN17,KRTAP26-1,KRTAP27-1,KRTAP23-1,KRTAP13-2,KRTAP13-4, KRTAP15-1,KRTAP19-2,KRTAP19-3,KRTAP19-4,KRTAP19-5, KRTAP19-6,KRTAP19-7,KRTAP6-3,KRTAP6-2,KRTAP22-1,KRTAP6-1, KRTAP20-1,KRTAP20-2,KRTAP21-2,KRTAP21-1,KRTAP8-1,KRTAP11-1, GRIK1,BACH1,KRTAP13-1,KRTAP13-3,KRTAP19-8,KRTAP20-3, KRTAP19-1,KRTAP24-1,CLDN8,KRTAP25-1
	Chr5:158988746-159912305	9	CCNJL,TTC1,SLU7,ADRA1B,FABP6,C5orf54,C1QTNF2,PTTG1,PWWP2A
	Chr6:67859000-70188705	1	BAI3
AST/ ALT	Chr1:173288426-174429270	3	TNN,TNR, SCARNA3
	Chr2:121450996-123193472	5	TSN,CLASP1,MKI67IP,TFCP2L1,RNU4ATAC
	Chr14:23001245-24313152	43	REC8,GZMH,NRL,CHMP4A,PSME2,DHRS2,RIPK3,AP1G2,SDR39U1, PSME1,LTB4R2,RABGGTA,IRF9,FITM1,C14orf21,NFATC4,GZMB, THTPA,LTB4R,NGDN,PCK2,TINF2,TM9SF1,FAM158A,DHRS4,TGM1, DHRS1,ADCY4,IPO4,GMPR2,TSSK4,JPH4,CMA1,RNF31,DHRS4L2, LRRC16B,CPNE6,CIDEB,CBLN3,NEDD8,CTSG,C14orf167,C14orf165
ALT	Chr7:142929078-143198980	1	FAM115C
	Chr12:107007904-107314707	2	WSCD2,CMKLR1
	Chr12:111893559-112564198	12	OAS2,SDS,SDSL,PLBD2,RASAL1,TPCN1,DTX1,DDX54,C12orf52, IQCD,SLC24A6,LHX5
	Chr12:52741396-53668999	16	ZNF385A,LACRT,ITGA5,SMUG1,HNRNPA1,DCD,PPP1R1A,MUCL1, COPZ1,CBX5,GPR84,NCKAP1L, PDE1B,NFE2,GTSF1,GLYCAM1
	Chr16:46036433-47063550	5	SIAH1,ABCC11,LONP2,PHKB,ABCC12
	Chr16:47069610-48586106	4	ZNF423,N4BP1,CBLN1,C16orf78
	Chr16:72954547-73014090	1	CLEC18B
	Chr18:52339709-52856818	2	WDR7,TXNL1
	Chr20:43763418-45210381	27	ACOT8,TP53RK,CD40,SNX21,PLTP,UBE2C,WFDC13,WFDC3,TNNC2, ZSWIM3,ZSWIM1,NEURL2,NCOA5,SLC35C2,ZNF334,SLC12A5, SLC2A10,PCIF1,CDH22,C20orf123,ELMO2,SLC13A3,SPINT4,DNTTIP1, MMP9,ZNF335,CTSA
	Chr22:24019061-24248709	1	LRP5L

injury[13]. Most acute or chronic HCV-infected people have no symptoms, but can occur symptoms such as tiredness, dark urine, itchy skin, poor appetite, abdominal pain, muscle soreness, and jaundice[24]. Treating for acute HCV was recommended rest, drinking large amount of fluids, eating healthy food, and avoiding alcohol. Patients with chronic HCV was treated with taking two oral medicines boceprevir and telaprevir, protease inhibitors that binds to the NS3 active site[25].

7 drugs have been associated with AST or ALT

Drug is very clinically important cause of liver injury. Many drugs have been reported to cause liver injury[26]. Gene-Drug interaction was established on the semantic integrated human liver disease datasets. Gene A is associated with drug D because gene A has a curated interaction with disease B, and disease B has a curated association with chemical C, and chemical C has a curated association with drug D. By smart query wizards (Figure 4(A)), seven drugs (acetaminophen, chlormezanone, stavudine, enflurane, isoniazid, mebendazole, and nitisinone) were associated with AST and ALT. Acetaminophen (APAP) is metabolized primarily in the liver, and APAP-overdose is the predominent cause of hepatic injury[27]. Stavudine is an antiviral me-

Table 2. Integrated public data into BioXM in the current version of the human liver disease knowledge base.

Entity	Source DB	Records	Relation	Records
Gene	http://www.geneontology.org	46,354	Gene-Disease Gene-Pathway Gene-Chemical	18,391,755 62,057 308,405
Somatic mutation	http://www.sanger.ac.uk/genetics/CGP/cosmic	242,217	SomaticMutation-Gene	32,695
GO	http://www.geneontology.org	36,855	GO-Gene	185,929
Pathway	http://ctdbase.org	362	Pathway-Disease Pathway-Chemical	43,139 196,073
Disease	http://ctdbase.org	9,647	Disease-Chemical	401,145
ClinicalTrials	http://www.clinicaltrials.gov	1,273	ClinicalTrials-Disease	1,210
Drug	http://www.drugbank.ca	6,712	Drug-ClinicalTrials	1,419
Chemical	http://ctdbase.org	153,021	Chemical-Drug	1,702
CNV	http://cancergenome.nih.gov	21,591	CNV-Gene	31,740
SNP	http://www.1000genomes.org	15,484	SNP-Gene	37,060
Total		518,032	Total	19,498,452

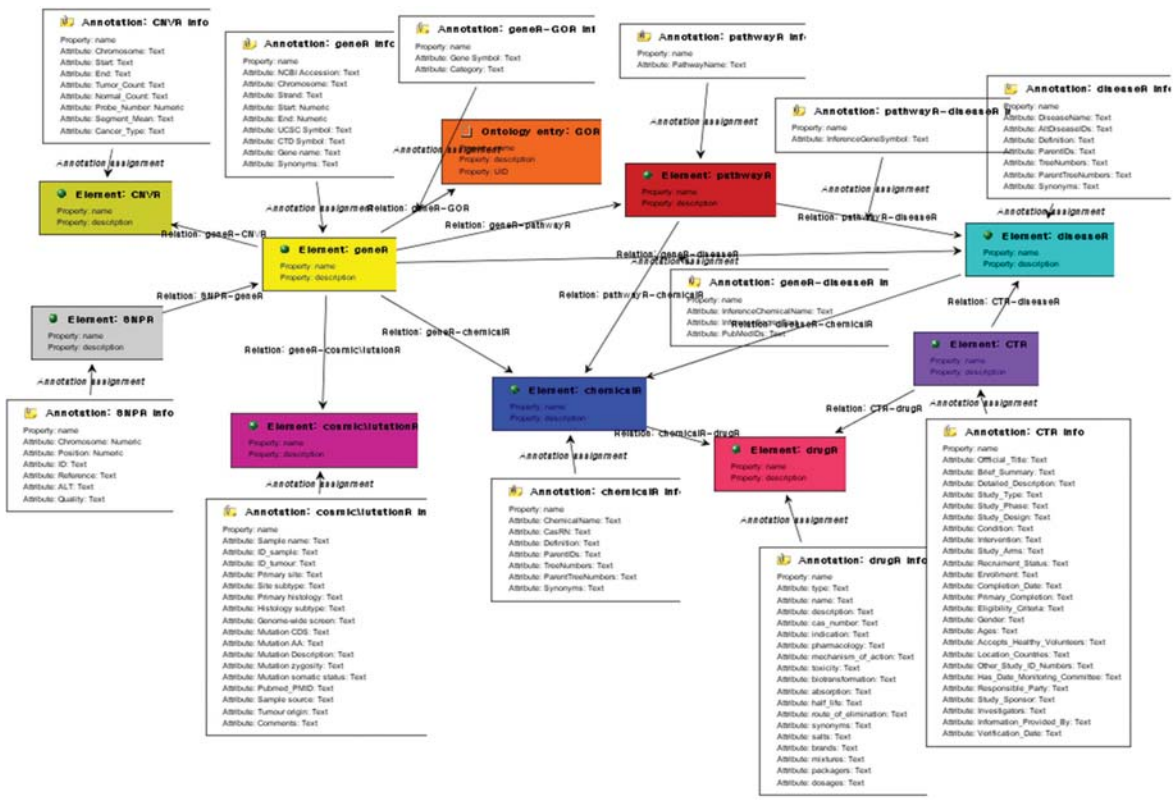

Figure 3. Scheme of semantic data integration model for human liver disease. The color box represents entities with description (white box), such as gene, chemical, pathway, disease, drug, GO, CNV, and SNP. The black arrows indicated the relations between two entities such as gene-pathway, gene-disease, and drug-chemical.

dication that active against human immunodeficiency virus. It can cause severe or often life-threatening effects on liver, and can increase risk of liver damage while taking it (http://www.drugs.com/mtm/stavudine.html). Isoniazid is an antibiotic, which prevents tuberculous bacteria (http://www.drugs.com/mtm/isoniazid.html). Mebendazole is an anti-worm medication and used for prevents infections of such as pinworm[28], whipworm[29], roundworm[30], and hookworm[31]. Nitisinone is used to treat hereditary tyrosinemia type 1. It

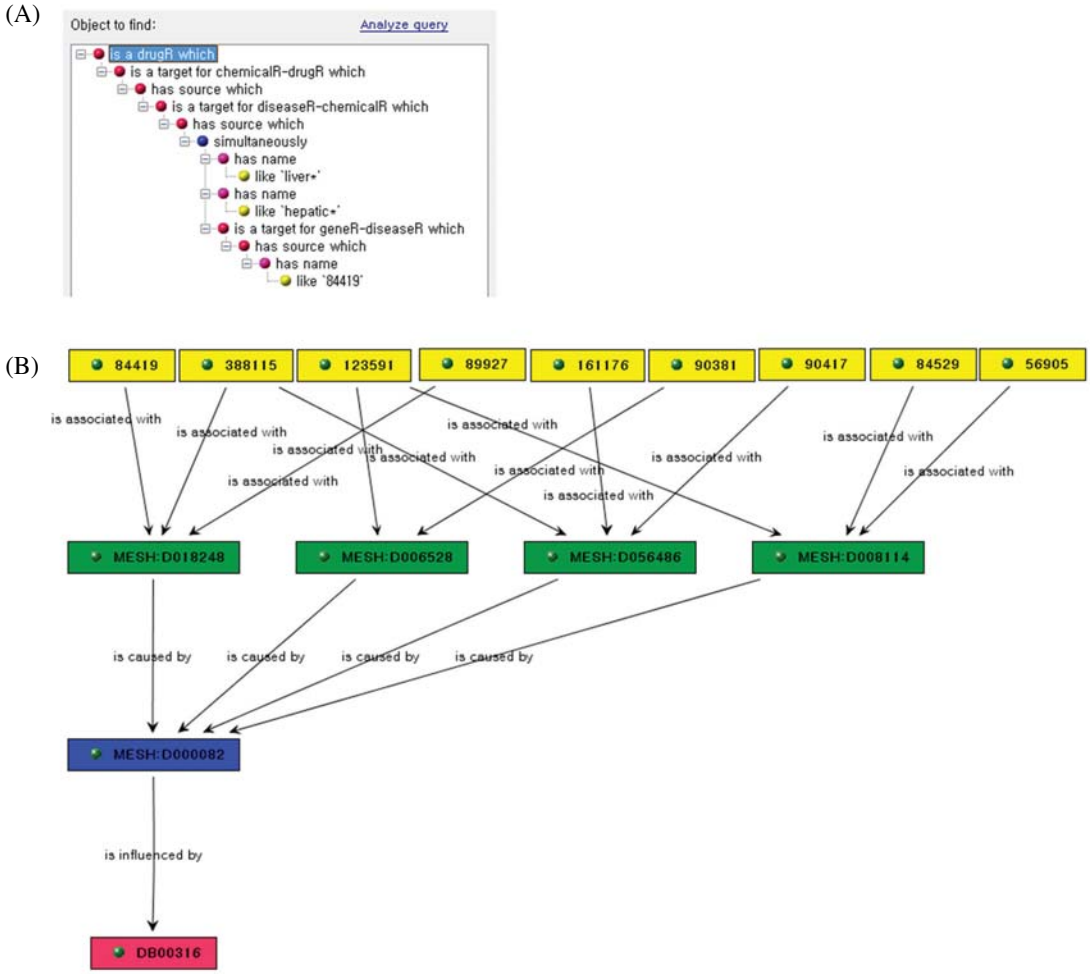

Figure 4. Query wizards : Find all drugs which related to liver disease by CNV-based genes (A); Semantic Network : Drug selected using CNV-based genes associated with liver disease (B). Drug DB00316 (red) is related to chemical MESH:D000082 (blue), and MESH:D000082 has interaction with diseases MESH:D018248, MESH:D006528, MESH:D056486, MESH:D008114 (green), and 4 diseases has a curated association with genes (yellow) such as 84419, 38115, 123591, and 89927. green is disease, and blue is chemical.

keeps causing harm to liver tissue, and its symptom is liver failure (http://www.drugs.com/mtm/nitisinone.html).

Discussion

As reported an important role of copy number in complex diseases, CNVs have become a more attractive field[4,6]. Differences of copy number between individuals contribute to alter in expression of genes sensitive to a disease susceptibility or dosage effect[6,7]. Liver function test is blood tests to evaluate about patient's liver state[1]. Alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are an important biochemical markers for evaluating of inflammation degree about liver injury[2]. Therefore, we focused on identifying copy numbers associated with AST or ALT in Korean population.

We identified 64 CNVs associated with liver biochemical markers AST or ALT using univariate linear regression. The size is approximately 10% of the genome in Korean cohorts. This result a little higher than found CNVs in same Korean populations[32]. Yim *et al.* (2010) identified CNVs on Affymetrix Genome-wide Human 6.0 Arrays in Korean population, and found that CNV covered about 8% of the whole genome. It seems that copy number differences could vary due to the incompletely knowledge about detection criterion for CNV calling and reference CNV[32,33].

228 genes fully located within our CNV regions were identified by mapping to the RefGene (hg18)

from UCSC Genome Browser. To explore the biological meaningful of structural variations, we performed the gene set enrichment for the CNV regions. The most enriched in biological process included transcription from RNA polymerase II promoter, DNA-dependent transcription, and RNA biosynthetic process. Transcription from RNA polymerase II promoter is regulation of transcription from an RNA polymerase II promoter by a ligand-bound hormone receptor. DNA-dependent transcription is the cellular RNA synthesis on a DNA template. RNA biosynthetic process is the chemical pathways and reactions resulting involving RNA (http://amigo.geneontology.org). Enriched functions in molecular functions were tropomyosin binding, including leukotriene B4 receptor activity, calcium ion binding, and carbonyl reductase activity. Calcium ion binding is selectively interacting with CA^{2+}. The important role of calcium ion in liver metabolism has been reported in recent researches[34,35]. Calcium ion binding protein is identified in the cytosol of rat liver[36], and it has a decreases of various enzyme by CA^{2+} in rat liver cell[37]. The most significantly enriched function in cellular components was intermediate filament, including intermediate filament cytoskeleton. Fifty percent of the pie was the proteasome activator complex (PAC). Because other term was statistically very significant, this term has many percentage. PAC enhances the hydrolysis of nonubiquitinated peptide by binding to the proteasome (http://www.ebi.ac.uk).

To determine whether the genes within our CNV regions showed the consistency of those early reported, we constructed semantic networks for human liver disease data sets using BioXM software. This semantic networks provides comprehensive and easy to use resource for human liver disease. Also it enables the retrieval of relationship networks including gene-disease, gene-pathway, gene-chemical, chemical-drug, and pathway-chemical. Our configuration-based approach to semantic integration network is closing the gap between public and experimental data. Recently, two studies of semantic networks have been published, which finding molecular signature of chemical[11] and managing toxicogenomic laboratory experiment[10]. This work supports to build such as gene-disease-chemical-drug relationship.

To the functional studies of the CNV-based genes, we performed gene functional classification using our semantic network. The significant results showed the three clusters, including four diseases (hepatocellular carcinoma, liver neoplasm, liver cell adenoma, drug-induced liver injury), one pathway (hepatitis C pathway), and seven drugs (acetaminophen, chlormezanone, stavudine, enflurane, isoniazid, mebendazole, nitisinone). Liver Cell Adenoma is a benign neoplasm occurred from liver cell (hepatocytes)[38], occur most often in young women[39]. It is important to recognize since it can be advanced a hepatocellular carcinoma (http://www.medicalgeek.com/). Liver Neoplasm is same name for liver (hepatic) cancer, and is an abnormal liver tissue (http://www.rightdiagnosis.com). Hepatocellular carcinoma (HCC) is the most common liver cancer. It occurs most often in men than women[40], and is usually seen in people 50 years of age or older (http://www.nlm.nih.gov). This cancer in Africa and Asia is more common than North or South America and Europe[41]. Drug is very clinically important cause of liver injury. Many drugs have been reported to cause liver (hepatic) injury[26]. Stavudine is an antiviral medication, which active against human immunodeficiency virus infection[42], and isoniazid is an antibiotic, which resists tuberculous bacteria (TB)[42], and mebendazole is an anti-worm medication, which used for prevents infections of such as pinworm, round worm, and hookworm[42]. Nitisinone is used to treat hereditary tyrosinemia type 1[43]. These drugs keep causing harm to liver tissue, and treated to cause liver injury.

We identified CNVs significantly genome-wide associated with the AST and ALT in Korean cohorts. CNV-based genes analyzed by functional studies using semantic networks for human liver disease datasets. As seen in Figure 4(B), we show the example to demonstrate how to robustly connect between public data and CNV analysis data. In this study, our results require validation for candidate CNV-based genes associated with liver using quantitative PCR. We expect that this research provide valuable CNV-based genes and useful semantic integration for liver disease sets.

Materials & Methods

Research subjects description

This study used 407 Korean cohorts (N=154 men and N=253 women) from the Korea Associated Resource 2 project. The Korean National Institutes of Health approved this study in 2009. All-study subjects signed an NIH-approved informed-consent forms. Genomic DNAs were isolated from peripheral blood of healthy participants, aged between 35 and 80 (mean 62 years). And a total of 407 DNA samples were genotyped using 500 ng of each genomic DNA on the Affymetrix Genome-wide Human 6.0 array. To the association study of liver function, we used the predicted values of AST and ALT from cohorts. To approximate a normal distribution, the value of AST and ALT were transformed to $1/(y)$ and 1/square root (y).

Defining of CNVs

We analyzed CNVs on Affymetrix Genome-Wide Human 6.0 array using the copy-number analysis module (CNAM) in the Helix SVS software (Golden Helix Inc, ver. 7.0)[44]. The CNAM module was read the 407 CEL files with intensity values, performed normalization on probe intensities by dividing the mean probe intensity of all chips as a reference, and created log2 ratio of the copy numbers. We chose the options of a multivariate algorithm used the segmentation with a moving window of 5,000 sizes, a maximum 40 segments per window, a minimum 1 marker per segment, a significance level of 0.01 for pair-wise permutations 1,000. The multivariate segmenting are algorithm for simultaneous CNVs discovery.

Association study between CNV and AST or ALT

To show the impact of copy number variation on AST or ALT, we used univariate linear regression. This genetic model was adjusted for gender and age and analyzed the log2 ratios of each CNV associated with AST or ALT as continuous response variables. The thresholds of statistical significance was determined a probability (P) value < 0.05.

For copy number as continuous variables.

$$Y=\beta_0+\beta_1 CNV+\beta_2 Gender+\beta_3 Age+\varepsilon$$

where β is a p vector of regression coefficients. All statistical analysis and parsing were performed using R and Python software.

CNV-based gene set enrichment analysis

We downloaded the hg18 RefGene lists from UCSC Genome Browser[45]. The genes were mapped to fully located within our CNV regions associated with AST or ALT. For the genes, gene set enrichment analysis was performed using Gene Ontology set from the Database for Annotation, Visualization and Integrated Discovery (DAVID) (ver. 6.7)[46]. The GO sets include biological process (BP), molecular process (MF), and cellular component (CC).

Semantic integration for human liver disease datasets

We constructed semantic networks for human liver disease data sets using BioXM software (ver. 2.2), which customizable knowledge base management for scientific data. The semantic network consisted of entities such as gene, disease, pathway, chemical, drug, SNP, CNV, and relations between two entities such as gene-GO, gene-pathway, gene-disease, gene-chemical,-gene-SNP, gene-CNV, pathway-chemical, chemical-drug. Semantic network provides the connection information between participating object instantly. We generalized this complex semantic network of detecting entity and connection for the answer to complex questions. Therefore this semantic data integration for liver disease enables us to create new knowledge networks with flexible workflow modeling. All analysis were performed using Python software and R package.

Acknowledgements The consortium for large-scale genotyping data was supported by the Korea Association Resource 2 project that was funded by the Korean National Institute of Health, Republic of Korea. This work was supported by the IT R&D program of MKE/KEIT. [KI00 18-10039594, Development of Molecular Diagnostic System for Personalized Cancer Treatment].

References

1. Thapa, B. & Walia, A. Liver function tests and their interpretation. *Indian J Pediatr* **74**:663-671 (2007).
2. Ruhl, C. E. & Everhart, J. E. Upper limits of normal for alanine aminotransferase activity in the United States population. *Hepatology* **55**:447-454 (2012).
3. Hsueh, C. J., Wang, J. H., Dai, L. & Liu, C. C. Development of an Electrochemical-Based Aspartate Aminotransferase Nanoparticle Ir-C Biosensor for Screening of Liver Diseases. *Biosensors* **2**:234-244 (2012).
4. Lee, B. Y., Cho, S., Shin, D. H. & Kim, H. Genome-wide association study of copy number variations associated with pulmonary function measures in Korea Associated Resource (KARE) cohorts. *Genomics* **97**: 101-105 (2011).
5. McCarroll, S. A. & Altshuler, D. M. Copy-number variation and association studies of human disease. *Nature Genetics* **39**:S37-S42 (2007).
6. Kim, H. Y., Yu, J. & Kim, H. Analysis of copy number variation in 8,842 Korean individuals reveals 39 genes associated with hepatic biomarkers AST and ALT. *BMB reports* **43**:547-553 (2010).
7. Redon, R. *et al*. Global variation in copy number in the human genome. *Nature* **444**:444-454 (2006).
8. Zhang, Y. *et al*. Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. *BMC Medical Genomics* **3**:1 (2010).
9. Houlston, R. S. *et al*. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature Genetics* **40**:1426-1435 (2008).
10. Kim, H. Y. *et al*. Semantic data integration for toxicogenomic laboratory experiment management systems. *Toxicology and Environmental Health Sciences* **3**:135-143 (2011).
11. Shin, G. H. *et al*. mRNA-centric semantic modeling

for finding molecular signature of trace chemical in human blood. *Mol Cell Toxicol* **8**:35-41 (2012).
12. Forbes, S. A. *et al*. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research* **39**:D945-D950 (2011).
13. Kampstra, P. Beanplot: A boxplot alternative for visual comparison of distributions (2008).
14. Davis, A. P. *et al*. Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical gene disease networks. *Nucleic Acids Research* **37**: D786-D792 (2009).
15. Wishart, D. S. *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research* **34**:D668-D672 (2006).
16. Overbeek, R. *et al*. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* **33**:5691-5702 (2005).
17. Higgins, M. E., Claremont, M., Major, J. E., Sander, C. & Lash, A. E. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Research* **35**: D721-D726 (2007).
18. Taniguchi, K. *et al*. I. Mutational spectrum of beta-catenin, AXIN1, and AXIN2 in hepatocellular carcinomas and hepatoblastomas. *Oncogene* **21**:4863 (2002).
19. Liu, K., Hogan, W. R. & Crowley, R. S. Natural Language Processing methods and systems for biomedical ontology learning. *Journal of Biomedical Informatics* **44**:163-179 (2011).
20. Zucman-Rossi, J. Benign Liver Tumors. *Molecular Pathology of Liver Diseases* 769-775 (2011).
21. Davern, T. J. Drug-induced liver disease. *Clinics in Liver Disease* **16**:231 (2012).
22. Björnsson, E. S. & Chalasani, N. Drug-Induced Liver Injury. Practical Gastroenterology and Hepatology: *Liver and Biliary Disease* 235-242 (2010).
23. Farazi, P. A. & DePinho, R. A. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nature Reviews Cancer* **6**:674-687 (2006).
24. Warrell, D. & Anderson, S. eds. Expedition medicine. London: Profile Books, 1998.
25. Steinkühler, C. *et al*. Product inhibition of the hepatitis C virus NS3 protease. *Biochemistry* **37**:8899-8905 (1998).
26. Lee, W. M. Drug-induced hepatotoxicity. *New England Journal of Medicine* **349**:474-485 (2003).
27. Davidson, D. & Eastham, W. Acute liver necrosis following overdose of paracetamol. *British Medical Journal* **2**:497-499 (1966).
28. Kullai Reddy Ulavapalli, J., Pyreddy, U. R. M. V. R. & Bobbarala, V. Indian Journal of Novel Drug Delivery. *Indian Journal of Novel Drug delivery* **3**:134-142 (2011).
29. Miller, M. J., Krupp, I. M., Little, M. D. & Santos, C. Mebendazole. *JAMA: The Journal of the American Medical Association* **230**:1412-1414 (1974).
30. Have, I. Previous version-Roundworm. *Hospitals* **121**: 0357.
31. De Clercq, D. *et al*. Failure of mebendazole in treatment of human hookworm infections in the southern region of Mali. *American Journal of Tropical Medicine and Hygiene* **57**:25-30 (1997).
32. Yim, S. H. *et al*. Copy number variations in East-Asian population and their evolutionary and functional implications. *Human Molecular Genetics* **19**:1001-1008 (2010).
33. Kim, H. Y., Byun, M. J. & Kim, H. A replication study of genome-wide CNV association for hepatic biomarkers identifies nine genes associated with liver function. *BMB reports* **44**: 578-583 (2011).
34. Rasmussen, H. Cell communication, calcium ion, and cyclic adenosine monophosphate. *Science* **170**:404-412 (1970).
35. Williamson, J. R., Cooper, R. H. & Hoek, J. B. Role of calcium in the hormonal regulation of liver metabolism. *Biochim Biophys Acta* **639**:243 (1981).
36. 山口正義 & 山本丈夫. Purification of calcium binding substance from soluble fraction of normal rat liver. *Chem Pharm Bull* **26**:1915-1918 (1978).
37. Mori, S. & Yamaguchi, M. Hepatic calcium-binding protein regucalcin decreases Ca^{2+}/calmodulin-dependent protein kinase activity in rat liver cytosol. *Chem Pharm Bull* **38**:2216 (1990).
38. Leese, T., Farges, O. & Bismuth, H. Liver cell adenomas. A 12-year surgical experience from a specialist hepato-biliary unit. *Annals of Surgery* **208**:558 (1988).
39. Edmondson, H. A., Henderson, B. & Benton, B. Liver-cell adenomas associated with use of oral contraceptives. *New England Journal of Medicine* **294**:470-472 (1976).
40. Beasley, R. P., Lin, C. C., Hwang, L. Y. & Chien, C. S. Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22 707 men in Taiwan. *The Lancet* **318**:1129-1133 (1981).
41. Bressac, B., Kew, M., Wands, J. & Ozturk, M. Selective G to T mutations of p 53 gene in hepatocellular carcinoma from southern Africa. *Nature* **350**:429-431 (1991).
42. Sommadossi, J. P. Comparison of Metabolism and in vitro antiviral activity of stavudine versus other 2′, 3′-Dideoxynucleoside Analogues. *J Infect Dis* **171**:S88 (1995).
43. Santra, S. & Baumann, U. Experience of nitisinone for the pharmacological treatment of hereditary tyrosinaemia type 1 informahealthcare (2008).
44. Lambert, C. HelixTree® Genetics Analysis Software. Golden Helix. Inc. http//wwwgoldenhelixcom (2005).
45. Karolchik, D. *et al*. The UCSC genome browser database. *Nucleic Acids Research* **31**:51-54 (2003).
46. Dennis, Jr, G. *et al*. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**:P3 (2003).