BMB
reports

A replication study of genome-wide CNV association for hepatic biomarkers identifies nine genes associated with liver function

*Hyo-Young Kim, Mi-Jeong Byun & Heebal Kim**

Department of Agricultural Biotechnology and the Research Institute for Agriculture and Life Science, Seoul National University, Seoul 151-742, Korea

Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are biochemical markers used to test for liver diseases. Copy number variation (CNV) plays an important role in determining complex traits and is an emerging area in the study various diseases. We performed a genome-wide association study with liver function biomarkers AST and ALT in 407 unrelated Koreans. We assayed the genome-wide variations on an Affymetrix Genome-Wide 6.0 array, and CNVs were analyzed using HelixTree. Using single linear regression, 32 and 42 CNVs showed significance for AST and ALT, respectively (P value < 0.05). We compared CNV-based genes between the current study (KARE2; AST-140, ALT-172) and KARE1 (AST-1885, ALT-773) using NetBox. Results showed 9 genes (*CIDEB, DFFA, PSMA3, PSMC5, PSMC6, PSMD12, PSMF1, SDC4,* and *SIAH1*) were overlapped for AST, but no overlapped genes were found for ALT. Functional gene annotation analysis shown the proteasome pathway, Wnt signaling pathway, programmed cell death, and protein binding. [BMB reports 2011; 44(9): 578-583]

INTRODUCTION

The biochemical liver function test (LFT) is commonly used to indicate for the extent of liver damage. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are biochemical markers used to test for liver diseases and are sensitive markers of hepatocellular disorder (1). The AST/ALT ratio is used as indicator for identifying cirrhosis in hepatitis patients (2). Whereas an AST/ALT ratio < 1 indicates mild liver damage such as nonalcoholic fatty liver disease (NAFLD), an AST/ALT ratio > 1 indicates serious liver disease such as cirrhosis, alcoholic fatty liver disease, or chronic hepatitis (3, 4).

*Corresponding author. Tel: 82-2-880-4803; Fax: 82-2-883-8812; E-mail: heebal@snu.ac.kr
http://dx.doi.org/10.5483/BMBRep.2011.44.9.578

Received 17 May 2011, Accepted 7 July 2011

Keywords: ALT, AST, Copy number, Liver, Meta-analysis

Genetic variations are shown by large-scale structural variants found in different individuals. Copy number variation (CNV) is a form of structural variation as a DNA segment ≥ 1 kb in size when compared to a reference genome. Studies on genetic variation contribute to the understanding of individual phenotypic differenceswhich can be manifested in drug dosage effects and susceptibility to disease (5, 6). Many CNVs in the human genome have been identified in various populations (7, 8). According to a CNV study from 4 populations with different ancestries in Asia, Africa, and Europe, CNVs accounted for ~12% of the genome in thesepopulations (9). CNVs have been shown to comprise 17.7% of the detected variationsin gene expression. Consequently, CNVs play an important role in determining complex traits (10, 11). Many studies on the association between CNVs and complex diseases in humans have been reported (12). Recently, a genome-wide CNV association study on AST and ALT in Koreans revealed 39 genes (13). However, association studies between CNV and diseases have been hindered due to incomplete knowledge of CNV detection criterion and lack of a reference CNV. Additionally, although most of the CNVs have been identified in various populations, the results may not directly apply to CNVs of all ethnicities (14).

In this study, we performed a genome-wide association study on CNVs and liver function biomarkers (AST or ALT) in 407 unrelated Korean subjects. We compared CNV-based genes in the current study with those in KARE1. Nine genes were overlapped for AST. This result has functional implications for CNVs associated with liver function.

RESULTS

Genome-wide association analysis between CNV and hepatic biomarkers

This study used genotype data from the 407 Korean individuals in the Korea Association Resource (KARE2) project, as by the Korean National Institute of Health. We focused on indentifying significant CNVs associated with hepatic biomarkers AST or ALT. The values of AST and ALT were transformed to 1/(y) and 1/square (y) to approximate a normal

distribution. We compared bean plots to show the frequency distributions of the AST and ALT in 407 individuals. We did not show differences in distributionsbetween men and women. CNVs were obtained from 407 individuals using the multivariate segmenting capability provided by HelixTree software. The total number of CNVs using all 407 chips as a reference was 3,046. We analyzed the impact of a single CNV for each quantitative phenotype using single linear regression. Results showed the positive β values of AST and ALT were 1,605 and 1,949 respectively, and the negative β values were 1,441 and 1,097 respectively. Of the CNVs tested in single-linear regression, 32 and 42 CNVs were significant forAST and ALT at the P value < 0.05, respectively (Fig. 1). Fig. 2 showsthe genome-wide association signals for AST and ALT on all 22 autosomes in Manhattan plots. We found 140 and 172 genes totally located within the significant CNV regions for AST and ALT, respectively.

Replication of CNV-based genes associated with AST and ALT

We asked whether any gene was replicated when comparedto the earlier study. Using GWAS meta-analysis for our genes within a CNV, we compared CNV-based genes between the

Fig. 1. Visualization of the physical distribution of significant CNVs for AST and ALT.

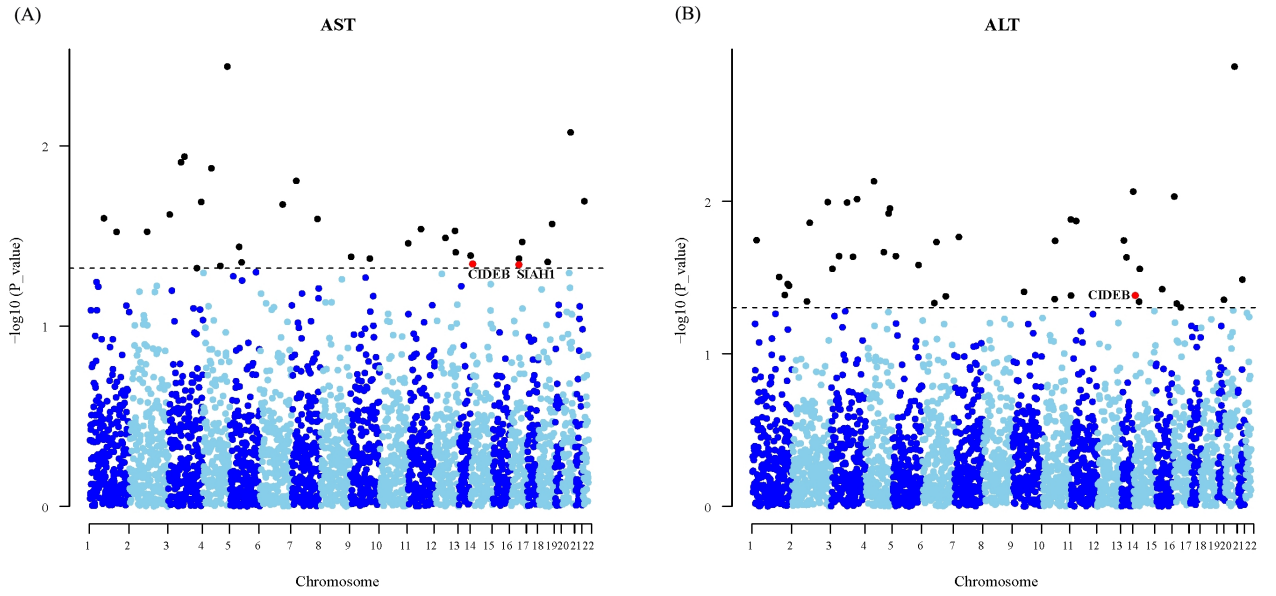

Fig. 2. Manhattan plot for the genome-wide association signals between all CNVs (n = 3,064) and AST (A) and ALT (B) on all 22 autosomes. Association was assessed using single linear regression adjusted for gender and age. The X axis shows chromosomal locations, and the P values were plotted on the Y axis using a logarithmic scale. The black dotted significant CNVs (P value < 0.05) and the red dotted genes associated with the liver were identified in previous studies.

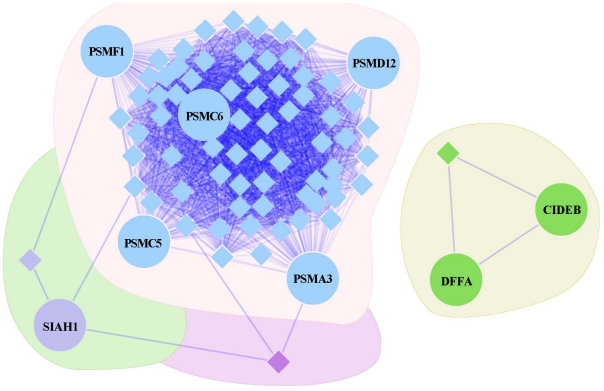

Fig. 3. The 4 largest modules were identified with a network modularity score of 0.004. Linker genes, showed as diamond shape, were not included in the original gene list, but are significantly connected with list-altered genes.

current study (KARE2; AST-140, ALT-172) and KARE1 (13) (AST-1885, ALT-773) using NetBox software. Results showed that 9 genes (*CIDEB, DFFA, PSMA3, PSMC5, PSMC6, PSMD12, PSMF1, SDC4,* and *SIAH1*) were overlapped for AST, but no overlapped gene was found for ALT. A total of 8 genes appeared within the network modules, but SDC4 was not present within the network at the shortest path threshold of 2, and the linker P value cut-off of 0.05. We discovered 4 large modules, with a network modularity score of 0.004. Fig. 3 shows visualized networks as determined using Cytoscape.

Proteasome pathway is enriched in AST

To probethe functional implications of structural variants, we analyzed the functional annotation of the 9 genes using the DAVID tool. Four genes (*PSMF1, PSMC6, PSMD12,* and *PSMA3*) were enriched in the proteasome biochemical pathway (P = 2.20E-07). The proteasome play a role in inhibiting cytokine production by liver cells. A decrease of proteasome activity develops during alcoholic liver injury and leads to inhibition of cell death. Therefore, chronic ethanol consumption suppresses proteasome activity in the liver (15, 16). Although not enriched in the KEGG pathway, *SIAH1* was found in the Wnt signaling pathway, which plays a role in liver development and regeneration (17). Okabe *et al*. found that the expression of *SIAH1* was downregulated in all hepatoma cells lines examined when compared with normal liver cells by semiquantitative RT-PCR. The decreased expression of *SIAH1* plays an important role in the development of hepatocellular carcinoma (HCC) (18).

Programmed cell death and protein binding are enriched in AST

Three genes (*DFFA, CIDEB* and *SIAH1*) were enriched in the programmed cell death (P = 0.04). Programmed cell death (PCD) plays an important role in liver development (19). Inohara *et al*. identified *CIDEB*, which is a subunit of the DNA fragment factor (DFF) (20). The *CIDEB* (cell death-inducing

Table 1. Significant genes associated with liver in the genome-wide meta-analysis

Gene*	Liver-associated phenotype	Enriched term	References
CIDEB	Programmed cell death	GO_BP	Saad *et al.* (2009)
	High expression in liver		Li *et al.* (2007); Ye *et al.* (2009)
DFFA	Programmed cell death	GO_BP	Saad *et al.* (2009)
PSMA3	Proteasome	KEGG	Donohue (2002); Donohue *et al.* (2007)
PSMC5	Protein binding	GO_MF	Jahnchen *et al.* (1981)
PSMC6	Proteasome	KEGG	Donohue (2002); Donohue *et al.* (2007)
	Overexpressed in human hepatocytes		Richert *et al.* (2006)
PSMD12	Proteasome	KEGG	Donohue (2002); Donohue *et al.* (2007)
	Overexpressed in human hepatocytes		Richert *et al.* (2006)
PSMF1	Proteasome	KEGG	Donohue (2002); Donohue *et al.* (2007)
SDC4	Abundant in liver		Rioux *et al.* (2002)
SIAH1	Wnt signaling pathway		Armengol *et al.* (2009)
	Programmed cell death	GO_BP	Saad *et al.* (2009)

*There are significant 9 genes for liver disease using GWAS meta-analysis

DFFA-like effector B) is expressed at high levels and plays an important role as a regulator of lipid metabolism in the liver (21-23). All 9 genes demonstrated enriched molecular functions, including protein binding (P = 0.0071). Liver disease can affect protein binding and causes impaired plasma protein binding of azapropazone (24, 25). Kojima *et al*. isolated *SDC4* from a rat endothelial cell (26). Rioux *et al*. identified *SDC4* (Syndecan-4) expressed at high levelsin mouse liver tissue by Northern blot analysis (27). The *SDC4* gene is comprised of 5 exons, and located in human chromosome 20q12. Table 1 shows that in previous studies, all 9 genes were reported to be associated with liver function.

DISCUSSION

An association study of CNV is important to understand the effect of variations on complex diseases. We identified 3,046 CNVs from 407 unrelated Korean subjects. CNV sizes ranged from 1,002 to 24,744 kb, with a median size of 547 kb. The NetBox software is based on copy number alteration and sequence mutation data, and assemblesaltered genes. It identifies linker genes, connects all altered genes, and then identifies network modules and calculatesnetwork modularity (28, 29). Although many meta- analysis methods have been reported (30), none were appropriate for our gene-based CNV data. For GWAS meta-analysis of our genes within CNV, we compared CNV-based genes between the current study and KARE1 using NetBox for meta-analysis. Results showed that 9 genes (*CIDEB, DFFA, PSMA3, PSMC5, PSMC6, PSMD12, PSMF1, SDC4,* and *SIAH1*) were overlapped for AST, but none were overlapped for ALT.

Regarding functional implications of the 9 genes, we analyzed functional annotation using the DAVID tool. Four genes (*PSMF1, PSMC6, PSMD12,* and *PSMA3*) were enriched in the proteasome biochemical pathway, which inhibition cytokine production by liver cells. And *SIAH1* was shown Wnt signaling pathway, which plays a role in liver development and regeneration (17). The decrease of proteasome activity causes alcoholic liver injury and inhibits liver cell death. Therefore, chronic ethanol consumption suppressed proteasome activity in the liver (15, 16). Richert *et al*. reported that PSMC6 (ATPase activity subunit) and PSMD12 (a non-ATPase subunit) were significantly overexpressed in human hepatocytes (31). The enriched Gene Ontology terms included programmed cell death (*DFFA, CIDEB* and *SIAH1*; P = 0.04) and protein binding (all 9 genes; P = 0.0071). The *PSMC6* and *PSMD12* genes encode a 403 and 397 amino-acid protein, and are located on chromosome 14q22.1and 17q24.2, respectively. Okabe *et al*. found the expression of *SIAH1* was downregulated in all hepatoma cells lines. The decreased expression of *SIAH1* plays an important role in the development of hepatocellular carcinoma (18).

In conclusion, we identified CNVs associated with the liver biomarkers of AST and ALT in 407 unrelated Koreans using the Affymetrix Genome-Wide 6.0 array. Four genes (*PSMF1, PSMC6, PSMD12,* and *PSMA3*) are involved in the proteasome biochemical pathway, and *SIAH1* was shown to be active in the Wnt signaling pathway. The 3 genes (*DFFA, CIDEB,* and *SIAH1*) were active in programmed cell death, and all 9 genes showed significant enrichment in protein binding, based on Gene Ontology. The enrichment of these genes suggests susceptibility or resistance mechanisms for liver disease. The CNV-based genes identified in this study will provide a valuable resource for further investigations of liver diseases. Additionally, our results require validation for candidate genes using quantitative PCR (qPCR).

MATERIALS AND METHODS

Research subjects

This study genotyped 407 unrelated Korean subjects (154 men, 253 women) under the Korea Associated Resource 2 (KARE2) project approved by the Korean National Institute of Health in 2009. All 407 participants signed informed-consent documents. Subject ages ranged from 35 to 80 years (mean 62.13; standard deviation 6.9). A 500 ng sample of genomic

DNA isolated from the peripheral blood of each participant was assayed on the Affymetrix Genome-Wide Human 6.0 array with 945,806 CN markers and 932,979 SNP markers. Two hepatic biomarkers (AST and ALT) were used in a genome-wide CNV association study of liver function.

Association analysis between CNV and hepatic biomarkers

We assayed the genome-wide variations on an Affymetrix Genome-Wide Human 6.0 array (Affymetrix, Santa Clara, CA, USA). CNVs were analyzed using the copy-number analysis module (CNAM) in the HelixTree software (Golden Helix, USA) (32). The HelixTree analysis software reads the intensity files, and runs normalization on probe intensities against reference samples, creating normalized log2 ratios. CNVs require a reference genome to be compared with samples. If a reference consists of imported chips run in different labs or using ethnicities, systematic differences represented variability. Hence, we used all 407 chips as reference sets to decrease the variability as much as possible. CNAM module analysis was done using a multivariate algorithm, a moving window of 5,000 markers, a maximum of 40segments per window, a minimum of 1 marker per segment, and a significance level of $P = 0.01$ for pairwise permutations ($n = 1,000$). Single linear regression was performed to test for association between single CNV as a dependent variable and adjusted for each phenotype associated with continuous response variables. We adjusted for the effects of gender and age of individuals in this analysis. For multiple testing, we used a P value of < 0.05 to signify statistical significance. All statistical analyses and parsing were performed using R and Python software.

Enrichment analysis of CNV-based genes

We assembled the genes totally located within the CNV region associated with the phenotypes. The genes were annotated using the RefGene (UCSC genome browser ver. hg18). NetBox software (http://cbio.mskcc.org/tools/netbox.html) (28) was used for GWAS meta-analysis, and networkmodules were visualized using Cytoscape (33). Gene enrichment analysis was performed using a total of 255 genes included in our CNVs. For that, we adopted 2 function sets from the DAVID tool (http://david.abcc.ncifcrf.gov/) (34), including Gene Ontology (35) and KEGG pathway (36).

Acknowledgements

This work was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ008116), Rural Development Administration, Republic of Korea, and the genotyping data was supported by the Korea Association Resource 2 (KARE2) project funded by the Korean National Institute of Health, Republic of Korea.

REFERENCES

1. Yap, C. and Aw, T. (2010) Liver Function Tests (LFTs). *Proceedings of Singapore Healthcare* **19**, 80-82.
2. Sheth, S., Flamm, S., Gordon, F. and Chopra, S. (1998) AST/ALT ratio predicts cirrhosis in patients with chronic hepatitis C virus infection. *Am. J. Gastroenterol.* **93**, 44-48.
3. Cohen, J. and Kaplan, M. (1979) The SGOT/SGPT ratio–an indicator of alcoholic liver disease. *Dig. Dis. Sci.* **24**, 835-838.
4. Clemenz, M., Frost, N., Schupp, M., Caron, S., Foryst-Ludwig, A., Bohm, C., Hartge, M., Gust, R., Staels, B. and Unger, T. (2008) Liver-Specific Peroxisome Proliferator-Activated Receptor ¥á Target Gene Regulation by the Angiotensin Type 1 Receptor Blocker Telmisartan. *Diabetes* **57**, 1405-1413.
5. Feuk, L., Carson, A. and Scherer, S. (2006) Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85-97.
6. Estivill, X. and Armengol, L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* **3**, 1787-1799.
7. Perry, G., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C., Scheffer, A., Steinfeld, I., Tsang, P. and Yamada, N. (2008) The fine-scale and complex architecture of human copy-number variation. *Am. J. Hum. Genet.* **82**, 685-695.
8. de Stahl, T., Sandgren, J., Piotrowski, A., Nord, H., Andersson, R., Menzel, U., Bogdan, A., Thuresson, A., Poplawski, A. and von Tell, D. (2008) Profiling of copy number variations (CNVs) in healthy individuals from three ethnic groups using a human genome 32 K BAC clone based array. *Hum. Mutat.* **29**, 398-408.
9. Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A. and Chen, W. (2006) Global variation in copy number in the human genome. *Nature* **444**, 444-454.
10. Stranger, B., Forrest, M., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C., de Grassi, A. and Lee, C. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853.
11. Beckmann, J., Estivill, X. and Antonarakis, S. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat. Rev. Genet.* **8**, 639-646.
12. Hastings, P., Lupski, J., Rosenberg, S. and Ira, G. (2009) Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551-564.
13. Kim, H., Cho, S., Yu, J., Sung, S. and Kim, H. (2010) Analysis of copy number variation in 8,842 Korean individuals reveals 39 genes associated with hepatic biomarkers AST and ALT. *BMB Rep.* **43**, 547-553.
14. Yim, S., Kim, T., Hu, H., Kim, J., Kim, B., Lee, J., Han, B., Shin, S., Jung, S. and Chung, Y. (2009) Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum. Mol. Genet.* **19**, 1001-1008.
15. Donohue, Jr, T., Cederbaum, A., French, S., Barve, S., Gao, B. and Osna, N. (2007) Role of the Proteasome in Ethanol Induced Liver Pathology. *Alcohol. Clin. Exp. Res.* **31**, 1446-1459.

16. Donohue, Jr, T. (2002) The ubiquitin proteasome system and its role in ethanol induced disorders. *Addict. Biol.* **7**, 15-28.
17. Armengol, C., Cairo, S., Fabre, M. and Buendia, M. (2009) Wnt signaling and hepatocarcinogenesis: The hepatoblastoma model. *Int. J. Biochem. Cell Biol.* **43**, 265-270.
18. Okabe, H., Satoh, S., Furukawa, Y., Kato, T., Hasegawa, S., Nakajima, Y., Yamaoka, Y. and Nakamura, Y. (2003) Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. *Cancer Res.* **63**, 3043-3048.
19. Saad, A., Aziz, A., Yehia, I., El-Ghareeb, A. and Ismail, H. (2009) Programmed Cell Death in the Liver of Different Species of Anuran Amphibians During Metamorphosis. *Australian Journal of Basic and Applied Sciences* **3**, 4644-4655.
20. Inohara, N., Koseki, T., Chen, S., Wu, X. and Nunez, G. (1998) CIDE, a novel family of cell death activators with homology to the 45 kDa subunit of the DNA fragmentation factor. *EMBO J.* **17**, 2526-2533.
21. Li, J., Ye, J., Xue, B., Qi, J., Zhang, J., Zhou, Z., Li, Q., Wen, Z. and Li, P. (2007) Cideb regulates diet-induced obesity, liver steatosis, and insulin sensitivity by controlling lipogenesis and fatty acid oxidation. *Diabetes* **56**, 2523.
22. Ye, J., Li, J., Liu, Y., Li, X., Yang, T., Ma, X., Li, Q., Yao, Z. and Li, P. (2009) Cideb, an ER-and Lipid Droplet-Associated Protein, aMediates VLDL Lipidation and Maturation byaInteracting with Apolipoprotein B. *Cell Metab.* **9**, 177-190.
23. Li, Z. (2007) To study the physiological role of cideb protein by using cideb knockout mice as a model system.
24. Blaschke, T. (1977) Protein binding and kinetics of drugs in liver diseases. *Clin. Pharmacokinet.* **2**, 32-44.
25. Jahnchen, E., Blanck, K., Breuing, K., Gilfrich, H., Meinertz, T. and Trenk, D. (1981) Plasma protein binding of azapropazone in patients with kidney and liver disease. *Br. J. Clin. Pharmacol.* **11**, 361-367.
26. Kojima, T., Shworak, N. and Rosenberg, R. (1992) Molecular cloning and expression of two distinct cDNA-encoding heparan sulfate proteoglycan core proteins from a rat endothelial cell line. *J. Biol. Chem.* **267**, 4870-4877.
27. Rioux, V., Landry, R. and Bensadoun, A. (2002) Sandwich immunoassay for the measurement of murine syndecan-4. *J. Lipid Res.* **43**, 167-173.
28. Cerami, E., Demir, E., Schultz, N., Taylor, B. and Sander, C. (2010) Automated network analysis identifies core pathways in glioblastoma. *PLoS One* **5**, 1-10.
29. Ding, L., Wendl, M., Koboldt, D. and Mardis, E. (2010) Analysis of Next Generation Genomic Data in Cancer: Accomplishments and Challenges. *Hum. Mol. Genet.* **1**, R188-R196.
30. Bax, L., Yu, L., Ikeda, N., Tsuruta, H. and Moons, K. (2006) Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Med. Res. Methodol.* **6**, 50.
31. Richert, L., Liguori, M., Abadie, C., Heyd, B., Mantion, G., Halkic, N. and Waring, J. (2006) Gene expression in human hepatocytes in suspension after isolation is similar to the liver of origin, is not affected by hepatocyte cold storage and cryopreservation, but is strongly changed after hepatocyte plating. *Drug Metab. Dispos.* **34**, 870-879.
32. Lambert, C. (2005) HelixTree¢ç Genetics Analysis Software. Golden Helix. Inc. http//www.goldenhelix.com.
33. Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504.
34. Dennis Jr, G., Sherman, B., Hosack, D., Yang, J., Gao, W., Lane, H. and Lempicki, R. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**, P3.
35. Harris, M., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B. and Mungall, C. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258-D261.
36. Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30**, 42-46.