Research Article

Identifying Liver Cancer-Related Enhancer SNPs by Integrating GWAS and Histone Modification ChIP-seq Data

Tianjiao Zhang,1 Yang Hu,2 Xiaoliang Wu,1 Rui Ma,1 Qinghua Jiang,2 and Yadong Wang1

1School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
2School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Qinghua Jiang; qhjiang@hit.edu.cn and Yadong Wang; ydwang@hit.edu.cn

Received 9 April 2016; Revised 30 May 2016; Accepted 1 June 2016

1. Introduction

Single nucleotide polymorphism (SNP) is a variation at a single nucleotide in a DNA sequence [1]. In the last decade, a large number of genome-wide association studies (GWAS) have been published, indicating that thousands of SNPs are associated with diseases. Linkage disequilibrium is the nonrandom association of alleles at different genome locations [2]. There are many SNPs in LD with the causal SNP at specific GWAS locus [3, 4]. Over 90% of these GWAS variants are located in noncoding regions, and approximately 10% are in LD with a protein-coding variant [5, 6]. In protein-coding regions, many studies have shown that some SNPs are associated with numerous diseases by affecting gene expression [7, 8]. However, in noncoding regions, the mechanism of how SNPs contribute to disease susceptibility remains unclear.

Enhancers are the core regulatory components of the genome that act over a distance to positively regulate gene expression [9]. It is estimated that 400,000 to 1 million putative enhancers exist in the human genome [10, 11]. Recently, some studies have shown that disease-related GWAS SNPs are correlated with enhancers marked with special histone modifications [12–15]. Therefore, through integrating GWAS and histone modification ChIP-seq data in a given disorder, we can identify disease-related enhancer SNPs.

We provided a method for identifying liver cancer-related enhancer SNPs through integrating liver cancer GWAS and histone modification ChIP-seq data. We identified 22 liver cancer-related enhancer SNPs, 9 of which were regulatory SNPs involved in distal transcriptional regulation. The results highlight that these enhancer SNPs may play important roles in liver cancer.
Identify genome enhancer regions

Human histone modification ChIP-seq data in HepG2

Human GWAS SNPs in liver cancer

Identify LD SNPs

Identify enhancer SNPs

Validation by rVarBase

2.2. Linkage Disequilibrium Analysis with Liver Cancer-Associated SNPs. We obtained 45 liver cancer-associated SNPs from GRASP (Table 1). These SNPs are the raw potential liver cancer-related SNPs. Then, we used LD data from HapMap to achieve liver cancer-associated LD SNPs. The total number of potential liver cancer-related SNPs is 340.

2.3. Identification of Liver Cancer-Related Enhancer SNPs. Previous studies indicated that the enhancer regions are marked by a strong H3K4me1 signal and a relatively weak H3K4me3 signal [19, 20]. Thus, we used histone modification ChIP-seq data to recognize the enhancer regions in liver cancer. Then, we mapped the liver cancer-related GWAS SNPs to the enhancer regions and obtained 22 enhancer SNPs in liver cancer (Table 2).

2.4. Validation as Regulatory SNPs. rVarBase is a database that provides reliable, comprehensive, and user-friendly annotations on variant’s regulatory features [18]. It includes regulatory SNPs (rSNPs), LD-proxies of rSNPs, and genes that are potentially regulated by rSNPs. We used rVarBase to analyze these 22 enhancer SNPs in liver cancer and found that 14 SNPs have evidence of regulatory SNPs and 9 SNPs (rs9494257, rs6903949, rs6996881, rs4739519, rs6988263, rs12156293, rs1568658, rs5994449, and rs5753816) are involved in distal transcriptional regulation (Table 3). Table 4 shows the potential target genes of these 9 SNPs.

3. Materials and Methods

3.1. GWAS and LD Datasets. We downloaded the human liver cancer-related GWAS SNPs from GRASP. The database includes 26 and 19 liver cancer-associated SNPs ($p < 10^{-5}$) from Han Chinese in Beijing, China (CHB), and Japanese in Tokyo, Japan (JPT), respectively. The URL is https://grasp.nhlbi.nih.gov/Overview.aspx. We obtained all SNPs in LD with GWAS-lead SNPs using LD blocks identified with publicly available HapMap data on the CHB and JPT populations. The LD data can be downloaded from http://hapmap.ncbi.nlm.nih.gov/index.html/.

3.2. Histone Modification Datasets. We downloaded the human histone modification ChIP-seq datasets in the HepG2 cell line from the ENCODE Production Data/Broad Institute. The URL is http://genome.ucsc.edu/ENCODE/downloads.html.

3.3. Linkage Disequilibrium Analysis. In the genome, SNPs located in close proximity tend to be in linkage disequilibrium with each other. The International HapMap Project has established linkage disequilibrium of human genome SNPs. We used LD data from HapMap to achieve liver cancer-associated LD SNPs ($R^2 > 0.8$).

3.4. Identify Enhancer Regions and Enhancer SNPs. Firstly, we downloaded the human histone modification BAM files (H3K4me1 and H3K4me3) in the HepG2 cell line from the ENCODE project. Then, we used BEDtools [21] to count read coverage for every position of the genome. Through calculating the ratio H3K4me1/H3K4me3 and picking up the regions with $\log_2(H3K4me1/H3K4me3) > 1.2$, we identified the potential enhancer regions. Finally, we mapped the potential LD SNPs to these enhancer regions and achieved liver cancer-related enhancer SNPs.
Table 1: Summary of liver cancer-associated SNPs from GRASP database.

| SNP ID    | p value | Chromosome | Populations | PMID        |
|-----------|---------|------------|-------------|-------------|
| rs17401966| 1.20E−19| 1          | CHB         | 20676096    |
| rs1249458 | 8.30E−06| 2          | CHB         | 22807686    |
| rs1714259 | 1.10E−06| 2          | CHB         | 22807686    |
| rs2396470 | 5.10E−07| 2          | CHB         | 20676096    |
| rs7424161 | 8.80E−06| 2          | CHB         | 22807686    |
| rs7574865 | 1.70E−11| 2          | CHB         | 23242368    |
| rs3905886 | 3.70E−06| 3          | CHB         | 22807686    |
| rs1073547 | 6.80E−06| 4          | CHB         | 22807686    |
| rs7821974 | 7.00E−06| 8          | CHB         | 22807686    |
| rs7898005 | 7.00E−08| 10         | CHB         | 20676096    |
| rs10160758| 6.00E−06| 11         | CHB         | 22807686    |
| rs12682266| 6.70E−06| 8          | CHB         | 22174901    |
| rs1573266 | 7.40E−06| 8          | CHB         | 22174901    |
| rs2275959 | 6.40E−06| 8          | CHB         | 22174901    |
| rs9272105 | 3.30E−23| 6          | CHB         | 22807686    |
| rs9275319 | 8.70E−19| 6          | CHB         | 23242368    |
| rs9494257 | 1.10E−14| 6          | CHB         | 20676096    |
| rs12682266| 6.70E−06| 8          | CHB         | 22174901    |
| rs1573266 | 7.40E−06| 8          | CHB         | 22174901    |
| rs2275959 | 6.40E−06| 8          | CHB         | 22174901    |
| rs9272105 | 3.30E−23| 6          | CHB         | 22807686    |
| rs9275319 | 8.70E−19| 6          | CHB         | 23242368    |
| rs9494257 | 1.10E−14| 6          | CHB         | 20676096    |
| rs12682266| 6.70E−06| 8          | CHB         | 22174901    |
| rs1573266 | 7.40E−06| 8          | CHB         | 22174901    |
| rs2275959 | 6.40E−06| 8          | CHB         | 22174901    |
| rs9272105 | 3.30E−23| 6          | CHB         | 22807686    |
| rs9275319 | 8.70E−19| 6          | CHB         | 23242368    |
| rs9494257 | 1.10E−14| 6          | CHB         | 20676096    |
| rs12682266| 6.70E−06| 8          | CHB         | 22174901    |
| rs1573266 | 7.40E−06| 8          | CHB         | 22174901    |
| rs2275959 | 6.40E−06| 8          | CHB         | 22174901    |
| rs9272105 | 3.30E−23| 6          | CHB         | 22807686    |
| rs9275319 | 8.70E−19| 6          | CHB         | 23242368    |
| rs9494257 | 1.10E−14| 6          | CHB         | 20676096    |
| rs12682266| 6.70E−06| 8          | CHB         | 22174901    |
| rs1573266 | 7.40E−06| 8          | CHB         | 22174901    |
| rs2275959 | 6.40E−06| 8          | CHB         | 22174901    |
| rs9272105 | 3.30E−23| 6          | CHB         | 22807686    |
| rs9275319 | 8.70E−19| 6          | CHB         | 23242368    |
| rs9494257 | 1.10E−14| 6          | CHB         | 20676096    |

4. Discussion

Through integrating liver cancer GWAS SNPs from GRASP, LD data from HapMap, and histone modification ChIP-seq data from ENCODE, we explored liver cancer-related enhancer SNPs. We compared our results with rVarBase and found that 9 SNPs (rs9494257, rs6903949, rs6996881, rs4739519, rs6988263, rs12156293, rs1568658, rs5994449,
Table 2: Summary of predicted enhancer SNPs in liver cancer.

| SNP ID     | Chromosome | Start   | End     | Chain | Populations |
|------------|------------|---------|---------|-------|-------------|
| rs12751375 | chr1       | 10291873| 10291874| +     | CHB         |
| rs6700866  | chr1       | 10306037| 10306038| +     | CHB         |
| rs9494257  | chr6       | 13582747| 13582747| +     | CHB         |
| rs17064474 | chr6       | 13568013| 13568013| +     | CHB         |
| rs17721919 | chr6       | 13574892| 13574892| +     | CHB         |
| rs17721931 | chr6       | 13574977| 13574978| +     | CHB         |
| rs6903949  | chr6       | 13582106| 13582106| +     | CHB         |
| rs6996881  | chr8       | 37407919| 37407920| +     | CHB         |
| rs4739519  | chr8       | 37412858| 37412859| +     | CHB         |
| rs6988263  | chr8       | 37414659| 37414660| +     | CHB         |
| rs12156293 | chr8       | 37419921| 37419922| +     | CHB         |
| rs6928810  | chr6       | 31410523| 31410524| +     | JPT         |
| rs3869132  | chr6       | 31410947| 31410948| −     | JPT         |
| rs2596562  | chr6       | 31354594| 31354595| −     | JPT         |
| rs2523475  | chr6       | 31361709| 31361710| −     | JPT         |
| rs2523467  | chr6       | 31362929| 31362930| −     | JPT         |
| rs9501387  | chr6       | 31364458| 31364459| +     | JPT         |
| rs1568658  | chr7       | 29141557| 29141558| −     | JPT         |
| rs1794304  | chr16      | 12625394| 12625395| +     | JPT         |
| rs5994449  | chr22      | 32304178| 32304179| +     | JPT         |
| rs5753816  | chr22      | 32312841| 32312842| +     | JPT         |
| rs5749339  | chr22      | 32315734| 32315735| +     | JPT         |

Table 3: Summary of liver cancer-related regulatory SNPs validated by rVarBase.

| SNP ID     | Regulatory SNP | Distal regulation | Chromatin state | Related regulatory elements                        |
|------------|----------------|-------------------|-----------------|-----------------------------------------------------|
| rs12751375 | Yes            | No                | Inactive region | n/a                                                  |
| rs6700866  | Yes            | No                | Weak transcription; ZNF genes and repeats; strong transcription; enhancers | n/a |
| rs9494257  | Yes            | Yes               | Enhancers; flanking active TSS; weak transcription | Chromatin interactive region |
| rs17064474 | Yes            | No                | Weak transcription; active TSS; flanking active TSS; enhancers | n/a |
| rs17721919 | Yes            | No                | Weak transcription | n/a |
| rs17721931 | Yes            | No                | Weak transcription | n/a |
| rs6903949  | Yes            | Yes               | Weak transcription; enhancers | TF binding region; chromatin interactive region |
| rs6996881  | Yes            | Yes               | Weak transcription; enhancers | Chromatin interactive region |
| rs4739519  | Yes            | Yes               | Enhancers; weak transcription | Chromatin interactive region |
| rs6988263  | Yes            | Yes               | Enhancers; weak transcription; genic enhancers; bivalent enhancer; flanking active TSS | Chromatin interactive region |
| rs12156293 | Yes            | Yes               | Enhancers; weak transcription; bivalent enhancer; genic enhancers | Chromatin interactive region |
| rs1568658  | Yes            | Yes               | Weak transcription; enhancers; strong transcription | Chromatin interactive region |
| rs5994449  | Yes            | Yes               | Weak transcription; strong transcription; ZNF genes and repeats | Chromatin interactive region |
| rs5753816  | Yes            | Yes               | Weak transcription; enhancers; flanking active TSS | Chromatin interactive region |
Table 4: Summary of liver cancer-related regulatory SNPs and potential target genes validated by rVarBase.

| SNP ID     | Gene symbol | Ensemble ID           | Regulation type                        |
|------------|-------------|-----------------------|----------------------------------------|
| rs9494257  | BCLAF1      | ENSG00000029363       | Distal transcriptional regulation       |
| rs9494257  | AHI1        | ENSG000000135541      | Distal transcriptional regulation       |
| rs9494257  | LINC00271   | ENSG000000231028      | Distal transcriptional regulation       |
| rs6903949  | MYB         | ENSG00000018513       | Distal transcriptional regulation       |
| rs6903949  | BCLAF1      | ENSG00000029363       | Distal transcriptional regulation       |
| rs6903949  | AHI1        | ENSG000000135541      | Distal transcriptional regulation       |
| rs6903949  | LINC00271   | ENSG000000231028      | Distal transcriptional regulation       |
| rs6996881  | ZNF703      | ENSG000000183779      | Distal transcriptional regulation       |
| rs6996881  | ERLIN2      | ENSG00000147475       | Distal transcriptional regulation       |
| rs6996881  | Null        | ENSG000000183154      | Distal transcriptional regulation       |
| rs6996881  | Null        | ENSG000000231361      | Distal transcriptional regulation       |
| rs4739519  | ZNF703      | ENSG000000183779      | Distal transcriptional regulation       |
| rs4739519  | Null        | ENSG000000254290      | Distal transcriptional regulation       |
| rs6988263  | ZNF703      | ENSG000000183779      | Distal transcriptional regulation       |
| rs6988263  | Null        | ENSG000000254290      | Distal transcriptional regulation       |
| rs12156293 | ZNF703      | ENSG000000183779      | Distal transcriptional regulation       |
| rs12156293 | Null        | ENSG000000254290      | Distal transcriptional regulation       |
| rs12156293 | ERLIN2      | ENSG000000147475      | Distal transcriptional regulation       |
| rs12156293 | Null        | ENSG000000183154      | Distal transcriptional regulation       |
| rs1568658  | Null        | ENSG000000228421      | Distal transcriptional regulation       |
| rs1568658  | TRIL        | ENSG000000176734      | Distal transcriptional regulation       |
| rs1568658  | Null        | ENSG000000255690      | Distal transcriptional regulation       |
| rs5994449  | DEPDC5      | ENSG000000100150      | Distal transcriptional regulation       |
| rs5994449  | FBXO7       | ENSG000000100225      | Distal transcriptional regulation       |
| rs5994449  | SYN3        | ENSG000000185666      | Distal transcriptional regulation       |
| rs5994449  | PRRI4L      | ENSG000000183530      | Distal transcriptional regulation       |
| rs5994449  | PISD        | ENSG000000241878      | Distal transcriptional regulation       |
| rs5994449  | EIF4ENIF1   | ENSG000000184708      | Distal transcriptional regulation       |
| rs5994449  | RNU6-28     | ENSG000000199248      | Distal transcriptional regulation       |
| rs5994449  | SFI1        | ENSG000000198089      | Distal transcriptional regulation       |
| rs5753816  | YWHAH       | ENSG000000128245      | Distal transcriptional regulation       |
| rs5753816  | C22orf24    | ENSG000000128254      | Distal transcriptional regulation       |
| rs5753816  | PISD        | ENSG000000241878      | Distal transcriptional regulation       |
| rs5753816  | DEPDC5      | ENSG000000100150      | Distal transcriptional regulation       |
| rs5753816  | RNU6-28     | ENSG000000199248      | Distal transcriptional regulation       |
| rs5753816  | SFI1        | ENSG000000198089      | Distal transcriptional regulation       |
| rs5753816  | EIF4ENIF1   | ENSG000000184708      | Distal transcriptional regulation       |
| rs5753816  | RPL3S       | ENSG000000205853      | Distal transcriptional regulation       |
| rs5753816  | Null        | ENSG000000230736      | Distal transcriptional regulation       |
| rs5753816  | Null        | ENSG000000231361      | Distal transcriptional regulation       |
| rs5753816  | Null        | ENSG0000002243519     | Distal transcriptional regulation       |
| rs5753816  | Null        | ENSG000000241954      | Distal transcriptional regulation       |
| rs5753816  | SYN3        | ENSG000000185666      | Distal transcriptional regulation       |

and rs5753816 were regulatory SNPs involved in distal transcriptional regulation. The results highlight that these enhancer SNPs may play important roles in liver cancer.

Compared with protein-coding regions in the human genome, noncoding regions contain much more genetic variations. Some important regulation regions, such as enhancers, have great influence on target gene expression. SNPs located in these regions may disturb gene expression and even cause diseases. Thus, the identification of SNPs in enhancer regions is helpful to understand the mechanism of association between SNPs and diseases.

We presented a method to identify disease-related SNPs located in enhancer regions that gives a new solution to investigate the relationship between SNPs and diseases. The
presented method can also be applied to other diseases and will enable biologists to investigate the mechanism of disease risk associated with SNPs.

**Competing Interests**

The authors declare that there are no competing interests regarding the publication of this paper.

**Authors’ Contributions**

Tianjiao Zhang collected the data, designed the computational experiments, carried out the statistical analysis, and wrote the paper. Qingshua Jiang participated in the design of the study. Yang Hu, Xiaoliang Wu, and Rui Ma participated in the revision of this paper. Yadong Wang gave comments and revisions to the final version of this paper. All authors read and approved the final paper. Tianjiao Zhang and Yang Hu equally contributed to this paper.

**Acknowledgments**

This work was partially supported by the National High-Tech Research and Development Program (863) of China (2012AA02A601, 2012AA02A602, 2012AA020404, 2012AA020409, 2012AA02A604, 2014AA021505, 2015AA020101, and 2015AA020108), National Science and Technology Major Project [no. 2013ZX03005012], and the National Natural Science Foundation of China (2012AA02A601, 2012AA02A602, 2012AA020404, 2012AA02A604, 2014AA021505, 2015AA020101, and 2015AA020108), National Natural Science Foundation of China (61571152, 31301089).

**References**

[1] R. Sachidanandam, D. Weissman, S. C. Schmidt et al., “A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,” Nature, vol. 409, no. 6822, pp. 928–933, 2001.

[2] M. Slatkin, “Linkage disequilibrium—understanding the evolutionary past and mapping the medical future,” Nature Reviews Genetics, vol. 9, no. 6, pp. 477–485, 2008.

[3] J. Peng, S. Uygur, T. Kim, Y. Wang, S. Y. Rhee, and J. Chen, “Measuring semantic similarities by combining gene ontology annotations and gene co-function networks,” BMC Bioinformatics, vol. 16, no. 1, article 44, 2015.

[4] J. Peng, T. Wang, J. Wang, Y. Wang, and J. Chen, “Extending gene ontology with association networks,” Bioinformatics, vol. 32, no. 8, pp. 1185–1194, 2016.

[5] M. T. Maurano, R. Humbert, E. Rynes et al., “Systematic localization of common disease-associated variation in regulatory DNA,” Science, vol. 337, no. 6099, pp. 1190–1195, 2012.

[6] M. A. Schaub, A. P. Boyle, A. Kundaje, S. Batzoglou, and M. Snyder, “Linking disease associations with regulatory information in the human genome,” Genome Research, vol. 22, no. 9, pp. 1748–1759, 2012.

[7] L. A. Hindorff, P. Sethupathy, H. A. Junkins et al., “Potential etiologic and functional implications of genome-wide association loci for human diseases and traits,” Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 23, pp. 9362–9367, 2009.

[8] Q. Zou, J. Li, Q. Hong et al., “Prediction of microRNA-disease associations based on social network analysis methods,” BioMed Research International, vol. 2015, Article ID 805514, 9 pages, 2015.

[9] J. Banerji, S. Rusconi, and W. Schaffner, “Expression of a β-globin gene is enhanced by remote SV40 DNA sequences,” Cell, vol. 27, no. 2, pp. 299–308, 1981.

[10] C. Buecker and J. Wysocka, “Enhancers as information integration hubs in development: lessons from genomics,” Trends in Genetics, vol. 28, no. 6, pp. 276–284, 2012.

[11] W. Xie and B. Ren, “Developmental biology. Enhancing pluripotency and lineage specification,” Science, vol. 341, no. 6143, pp. 245–247, 2013.

[12] S. Heinz, C. Benner, N. Spann et al., “Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities,” Molecular Cell, vol. 38, no. 4, pp. 576–589, 2010.

[13] J. Ernst, P. Kheradpour, T. S. Mikkelsen et al., “Mapping and analysis of chromatin state dynamics in nine human cell types,” Nature, vol. 473, no. 7345, pp. 43–49, 2011.

[14] B. Akhtar-Zaidi, R. Cowper-Sallari, O. Corradin et al., “Epigenomic enhancer profiling defines a signature of colon cancer,” Science, vol. 336, no. 6082, pp. 736–739, 2012.

[15] G. Trynka, C. Sandor, B. Han et al., “Chromatin marks identify critical cell types for fine mapping complex trait variants,” Nature Genetics, vol. 45, no. 2, pp. 124–130, 2013.

[16] R. Leslie, C. J. O’Donnell, and A. D. Johnson, “GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database,” Bioinformatics, vol. 30, no. 12, pp. i85–i94, 2014.

[17] The International HapMap Consortium, “The international HapMap project,” Nature, vol. 426, no. 6968, pp. 789–796, 2003.

[18] L. Guo, Y. Du, S. Qu, and J. Wang, “rVarBase: an updated database for regulatory features of human variants,” Nucleic Acids Research, vol. 44, no. 1, pp. D888–D893, 2016.

[19] F. de Santa, I. Barozzi, F. Mietton et al., “A large fraction of extragenic RNA Pol II transcription sites overlap enhancers,” PLoS Biology, vol. 8, no. 5, article e1000384, 2010.

[20] A. C. Marques, J. Hughes, B. Graham, M. S. Kowalczyk, D. R. Higgs, and C. P. Ponting, “Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs,” Genome Biology, vol. 14, no. 11, article R31, 2013.

[21] A. R. Quinlan, “UNIT 11.12 BEDTools: the swiss-army tool for genome feature analysis,” Current Protocols in Bioinformatics, vol. 47, pp. i11.2.1–i11.2.34, 2014.