Int J Clin Exp Pathol 2015;8(9):11929-11933
www.ijcep.com /ISSN:1936-2625/IJCEP0012200

Original Article

Analysis of susceptible genes and chromosome loci for lung cancers by automated gene prediction tools and genome scanning meta-analysis

Yu Tian[1], Daoxin Wang[2]

[1]Department of Respiratory Medicine, Anyue People's Hospital of Sichuan Province, Ziyang 642350, Sichuan, China; [2]Department of Respiratory and Critical Care Medicine, Second Hospital Affiliated to Chongqing Medical University, Yuzhong, Chongqing 400010, China

Received July 1, 2015; Accepted August 20, 2015; Epub September 1, 2015; Published September 15, 2015

Abstract: Genome-wide scanning of susceptible loci and genes for medical diseases is important in current post-genome era. To date, a variety of studies have been focused on the experimental validation or genome-wide linkage scans across multiple populations hunting for susceptibility genes in lung cancer. In the present study, we used two gene prediction tools (PROSPECTR and SUSPECTS, Gen Wanderer) to analyze eight previously identified susceptibility loci for lung tumors, which are selected from literature searching. Our results showed that there was a subset of 26 likely candidate susceptible genes related to lung cancer in each chromosomal region. For potential susceptible chromosome loci, the genome-wide scanning meta-analysis using bins of 60 cM width predicted a group of potential regions associated with lung cancer. Locus 15q21-26 (P=0.000606) is strongly evidenced, which has been confirmed in previous work. In contrast, another potential locus 10q11.2-q23.3 (P=0.0435223) is suggestively evidenced, which was never identified before. Ac compared to previous known regions, the latter one is the new detected one in our study. In conclusion, our study may be useful to contribute to further experimental tests of susceptibility genes/loci related to lung cancer.

Keywords: Lung caner, susceptibility genes/loci, prioritization, computational strategy, meta-analysis

Introduction

Complex diseases such as lung cancer are influenced by multiple genes with small individual effects, environments and gene interaction [1]. Prolonged exposure to carcinogens found in tobacco smoke and other environmental carcinogens that interact with various genetic susceptibility factors contribute to lung cancer development in humans [2].

Now there are some promising bioinformatics tools developed for disease-gene identification. Most of these tools in disease candidate gene identification and prioritization depend on functional annotations of the genes in the genome project. These methods have been successfully used to prioritize candidate disease genes for some kinds of diseases, like type 2 diabetes and obesity [3-5].

Genome scanning meta-analysis (GSMA) is a technique aiming to identify susceptible genomic loci of diseases. The method is derived from meta-analysis to some extent, and was firstly introduced by Wise et al [6]. The GSMA is a non-parametric method, which is applicable to the results for any genome-wide linkage study [7]. The method can combine genome-scan methods across studies with different markers, different statistical tests, and is reliable to the study design and ascertainment differences [8]. The GSMA method has been successfully applied to a lot of diseases so far. For example, GSMA applications on the diseases like multiple sclerosis, schizophrenia, bipolar disorder have shown that GSMA can identify potential susceptible loci [9, 10].

In this study, we applied automated gene prediction tools to prioritize candidate genes susceptible to lung cancer and to identify potential

Table 1. Final 26 candidate genes susceptible to lung cancer, which were consistently identified by two prediction methods (PROSPECTR/SUSPECTS and Gene Wanderer)

Genes	Loci
ARNTL2	12p11.2-p12.1
DDX11	12p11.2-p12.1
DNM1L	12p11.2-p12.1
FGD4	12p11.2-p12.1
PTHLH	12p11.2-p12.1
STK38L	12p11.2-p12.1
SURB7	12p11.2-p12.1
APEX1	14q11-21
CYP1A1	15q24-25.1
CCL2	17q12-13
ERBB2	17q12-13
CAST	5p15
ELL2	5p15
GLRX	5p15
KIAA0372	5p15
LNPEP	5p15
NR2F1	5p15
PCSK1	5p15
RHOBTB3	5p15
RIOK2	5p15
LAMA4	6p21
MICAL1	6p21
REV3L	6p21
SESN1	6p21
CTGF	6q23-25
ESR1	6q23-25

Table 2. Top-ranked loci based on GSMA weighted analysis

BIN	SR	*P*-value	-log (P)
15.2	306	0.000606*	3.217527
10.2	244.333	0.043522*	1.361288
12.2	235.667	0.066966	1.174144
2.2	230.667	0.085505	1.068011
5.1	229	0.091865	1.036852

*denotes the significant bins in the analysis (P<0.05).

susceptible loci by GSMA method. Our aim is to explore possible susceptibility genes/loci based on the data mining of available databases and literature so as to deeper our understanding of the role of gene-environment interaction on lung cancer incidence. In gene identification, four tools were used for forming consensus gene prioritization results. After the final candidate list was generated, the selected genes were subject to pathway analysis and network construction. In loci identification, GSMA method is implemented based on the collection of previous literature on the genome-wide linkage association of lung tumors.

Materials and methods

Susceptibility loci selection

To date, the genome-wide linkage scans across multiple populations have been performed to identify quantitative trait loci underlying. Firstly, we used the key words “susceptible loci”, “lung cancer” to search the online literatures. The purpose is to identify the possible susceptible loci for lung caner for our study. Finally, in total the selected loci were 5p15.33, 6q23-25, 6p21.33, 15q24-25.1, 17q12-13, 14q11-21, 12p11.2-p12.1, and 19q13.3.

Gene identification methods

Two online tools are used to identify disease and gene, including PROSPECTR/SUSPECTS [11], and Gene Wanderer [12]. The setup is to provide the start and end locations in the base pair of the corresponding loci. A gene was considered to be interesting as a candidate gene if it was indicated by both tools. For each analysis, only the top 10 candidate genes were retained for subsequent analysis.

Genome scanning meta-analysis

To identify published studies for inclusion in the GSMA of lung cancer and related phenotypes, we completed a systematic literature review. We used PudMed and the following search string (“lung cancer OR lung tumor AND linkage AND genome scanning OR genome linkage OR genome-wide”).

The relevant characteristics of each study included in the GSMA are summarized. For each study, the following information was extracted: first author, journal and year of publication, ethnicity of study population, number of families, number of markers, linkage statistic, and software for linkage analysis. If the genome-wide linkage results were available for the published graphs, the required linkage sta-

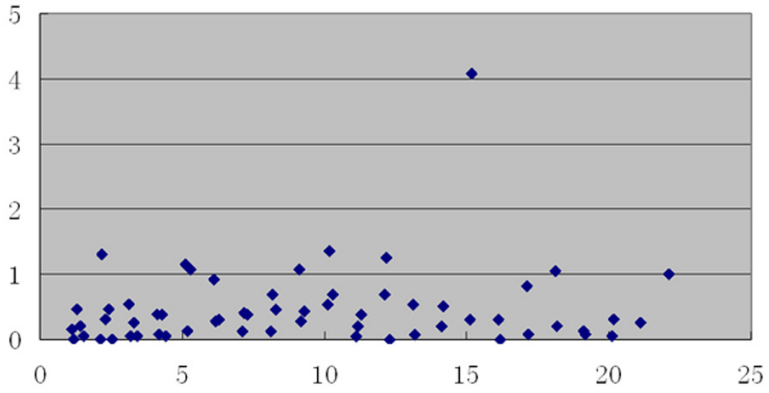

Figure 1. GSMA weighted analysis based on 60 cM cytogenic bands. The Y-axis is the -log (*P*) value. Only values >1.3 means the significance level at P<0.05.

tistics for SMA were extracted from the Figures by digital software g3 data (www.frantz.fi/software/g3data.php). The genetic map positions and marker locations were unified on the basis of Marshfield genetic map.

The GSMA method was used to synthesize the evidence for linkage across multiple genome scans. In the primary GSMA, chromosomes are divided into approximately equal-length bins-traditionally approximately 30 centiMorgan (cM), generating a total number of 118 bins on the autosomes based on the Marshfield genetic map. For each study, each bin was assigned a within-study rank by its highest logarithm of odds, nonparametric linkage, or Z score or minimum *p* value. So the bin with the highest linkage score or minimum *p* value is assigned a rank 118, and other binds are ranked in the descending order of their strength of linkage. The ranks were then summed across the studies for each bin to obtain the summed rank, which forms the basic test statistic for assessing linkage within the bin. Bins with high ranks might show significant evidence for linkage. We used GSMA software [7] to evaluate empirically the significance of the summed rank. We performed both unweighted and weighted GSMA analysis [10].

Results and discussion

Candidate susceptible genes

Twenty-six genes were selected as potential lung caner susceptibility genes by using the online analysis tools (**Table 1**). Some interesting genes identified are the ones being found empirically to be associated with the incidence of lung cancers. For example, ESR1 is a tumor suppressor gene, and the inactivation of it by promoter lesion methylation in lung cancer has been reported [13]. FGD4, an activator of the small GTPase CDC42, is one of six genes screened by the work of Whitehurst et al. [14], which can sensitize non-small lung cancer cells to paclitaxel.

Another gene in our interest is the CCL2 gene. The protein derived from CCL2 gene is secreted from tumor cells and associated tumor stromal cells. Blockade of CCL2 mediated by neutralizing antibodies has been shown to reduce tumorigenesis of several tumors. However, the role of CCL2 in lung cancers is controversial so far. Recently, the work of Fridlender et al. [15] identified that CCL2 blockade can inhibit lung cancer tumor growth by altering macrophage phenotype and activating CD8+ cells.

Candidate susceptible loci

Based on the online searching of literature, we found four genome-wide association studies on lung cancer and can be used in our GSMA analysis. The literatures were Amos et al. [16], Broderick et al. [17], Landi et al. [18], and Li et al. [19]. Based on the collected literatures, the GSMA results from the weighted and unweighted analyses for the defined 60 cM bin width are similar, and the weighted results are shown in **Table 2** and **Figure 1**.

We identified two significantly potential loci in lung cancer susceptibility as compared to previously eight loci (the ones used for susceptible gene identification above). Locus 15q21-26 (P=0.000606) is strongly evidenced, while the locus 10q11.2-q23.3 (P=0.0435223) is suggestively evidenced. Actually, the locus 15q21-26 has been confirmed before. Amos et al. [16] confirmed that 15q25.1 is a susceptible locus for lung cancers, which is located within the region we identified. As compared to previous works, the locus 10q11.2-q23.3 was firstly identified in our study.

Traditionally, 30 cM bin width is commonly used in previous literatures. However, peak linkage scores in individual studies may map to adjacent bins, and dilute the evidence for linkage in SMA, reducing both power and precision [9]. This is also of problem where genes are located close to bin boundaries, and shifting bins to span the region from the mid-points of the original bins should increase the evidence of linkage [10].

Hermanowksi et al. [10] used 20 cM and 40 cM bin widths in their GSMA analysis. Our study followed a broad-scale bin boundary, by defining 60 cM width, resulting in a total number of 60 bins (only autosomes are considered).

Conclusions

In this brief work, we identified susceptible chromosome loci and genes related to lung cancer occurrence. The methods we used are the automated gene prediction online tools and meta-analysis of genetic linkage studies. Twenty-six genes were found related to lung cancers, among which some have been confirmed in previous experiments and some were firstly predicted. One potentially new susceptible locus was found, i.e., the chromosome location of which was 10q11.2-q23.3. Our work provides insights into the works on the genetic associations and genetic markers of lung tumors.

Disclosure of conflict of interest

None.

Address correspondence to: Daoxin Wang, Department of Respiratory and Critical Care Medicine, Second Hospital Affiliated to Chongqing Medical University, No. 74, Linjiang Road, Yuzhong, Chongqing 400010, China. Tel: +86-023-63693222; E-mail: tianyu898@yeah.net

References

[1] Huang QY, Li GHY, Cheung WMW, Song YQ. Prediction of osteoporosis candidate genes by computational disease-gene identification strategy. J Hum Genet 2008; 53: 644-655.

[2] Liu PY, Vikis H, James M, Lu Y, Wang DL, Liu HB, Wen WD, Yang Y, You M. Identification of Las2, a major modifier gene affecting the Pas1 mouse lung tumor susceptibility locus. Cancer Res 2009; 69: 6290.

[3] Elbers CC, Onland-Moret NC, Franke L, Niehoff AG, van der Schouw YT, Wijmenga C. A strategy to search for common obesity and type 2 diabetes genes. Trends Endocrin Met 2006; 18: 19-26.

[4] Teber ET, Liu JY, Ballouz S, Fatkin D, Wouters MA. Comparison of automated candidate gene prediction systems using genes implicated in type 2 diabetes by genome-wide association studies. BMC Bioinformatics 2009; 10: S69.

[5] Tiffin N, Adie E, Turner F, Brunner HG, van Driel MA, Oti M, Lopez-Bigas N, Ouzounis C, Perez-Iratxeta C, Andrade-Navarro MA, Adeyemo A, Patti ME, Semple CA, Hide W. Computational disease gene identification: a concert of methods prioritises type 2 diabetes and obesity candidate genes. Nucleic Acids Res 2006; 34: 3067-81.

[6] Wise LH, Lanchbury JS, Lewis CM. Meta-Analysis of Genome Searches. Ann Hum Genet 1999; 63: 263-272.

[7] Pardi F, Levinson DF, Lewis CM. GSMA: software implementation of the genome search meta-analysis method. Bioinformatics 2005; 21: 4430-4431.

[8] Cooper ME, Goldstein TH, Maheer BS, Marazita ML. Identifying genomic regions for fine-mapping using genome scan meta-analysis (GSMA) to identify the minimum regions of maximum significance (MRMS) across populations. BMC Genet 2005; 6 Suppl 1: S42.

[9] Levinson DF, Levinson MD, Segurado R, Lewis CM. Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: methods and power analysis. Am J Hum Genet 2003; 73: 17-33.

[10] Hermanowski J, Bouzigon E, Forabosco P, Ng MY, Fisher SA, Lewis CM. Meta-analysis of genome-wide linkage studies for multiple sclerosis, using an extended GSMA method. Eur J Hum Genet 2007; 15: 703-710.

[11] Adie EA, Adamns RR, Evans KL, Porteous DJ, Pickard BS. SUSPECTS: enabling fast and effective prioritization of positional candidates. Bioinformatics 2006; 22: 773-774.

[12] Köhler S, Bauer S, Horn D, Robinson PN. Walking the Interactome for Prioritization of Candidate Disease Genes. Am J Hum Genet 2008; 82: 949-958.

[13] Suga Y, Miyajima K, Oikawa T, Maeda J, Usuda J, Kajiwara N, Ohira T, Uchida O, Tsuboi M, Hirano T, Kato H, Ikeda N. Quantitative p16 and ESR1 methylation in the peripheral blood of patients with non-small cell lung cancer. Onco Rep 2008; 20: 1137-42.

[14] Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, Minna JD, Michnoff C, Hao W, Roth MG, Xie XJ, White MA. Synthetic lethal screen identification of chemo-

sensitizer loci in cancer cells. Nature 2007; 446: 815-819.
[15] Fridlender ZG, Kapoor V, Buchlis G, Cheng G, Sun J, Wagn LCS, Singhal S, Snyder LA, Albelda SM. CCL12 blockade inhibits lung caner tumor growth by altering macrophage phenotype and activating CD8+ cells. Am J Respir Cell Mol Biol 2011; 44: 230-7.
[16] Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnana J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008; 40: 616-622.
[17] Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, Amos CI, Houlston RS. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer Res 2009; 69: 6633-41.
[18] Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Wang Y, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metspalu A, Vooder T, Nelis M, Välk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeböller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S, Caporaso NE. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. Am J Hum Genet 2009; 85: 679-691.
[19] Li Y, Sheu CC, Ye Y, de Andrade M, Wang L, Chang SC, Aubry MC, Aakre JA, Allen MS, Chen F, Cunningham JM, Deschamps C, Jiang R, Lin J, Marks RS, Pankratz VS, Su L, Li Y, Sun Z, Tang H, Vasmatzis G, Harris CC, Spitz MR, Jen J, Wang R, Zhang ZF, Christiani DC, Wu X, Yang P. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. Lancet Oncol 2010; 11: 321-30.