Gene 551 (2014) 86–91

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Pathway analysis of genome-wide association study on serum prostate-specific antigen levels

Yu-Zheng Ge [a,1], Zheng Xu [a,1], Lu-Wei Xu [a,1], Peng Yu [b], Yan Zhao [c], Hui Xin [a], Ran Wu [a], Si-Jia Tan [a], Qun Song [a], Jian-Ping Wu [a], Wen-Cheng Li [a], Jia-Geng Zhu [a], Rui-Peng Jia [a,*]

[a] *Department of Urology, Nanjing First Hospital, Nanjing Medical University, 68 Changle Road, Nanjing 210006, China*
[b] *Department of Urology, The First Hospital of Nanchang, Nanchang University, 128 Xiangshan North Road, Nanchang 330008, China*
[c] *Department of Urology, Xuzhou Third People's Hospital, Jiangsu University, 131 Huancheng Road, Xuzhou 221005, China*

ARTICLE INFO

Article history:
Received 12 July 2014
Received in revised form 19 August 2014
Accepted 23 August 2014
Available online 26 August 2014

Keywords:
Prostate-specific antigen
Genome-wide association study
Pathway analysis

ABSTRACT

The wide application of prostate-specific antigen (PSA) has contributed to the early diagnosis and improved management of prostate cancer (PCa). Accumulating evidence has suggested the involvement of genetic components in regulating serum PSA levels, and several single nucleotide polymorphisms (SNPs) have been identified by genome-wide association studies (GWASs). However, the GWASs' results have the limited power to identify the causal variants and pathways. After the quality control filters, a total of 330,540 genotyped SNPs from one GWAS with 657 PCa-free Caucasian males were included for the identify candidate causal SNPs and pathways (ICSNPathway) analysis. In addition, the genotype–phenotype association analysis has been conducted with the data from HapMap database. Overall, a total of four SNPs in three genes and six pathways were identified by ICSNPathway analysis, which in total provided three hypothetical mechanisms. First, *CYP26B1* rs2241057 polymorphism (nonsynonymous coding) which leads to a Leu-to-Ser amino acid shift at position 264, was implicated in the pathways including meiosis, proximal/distal pattern formation, and M phase of meiotic cell cycle. Second, *CLIC5* rs3734207 and rs11752816 polymorphisms (regulatory region) to the 2 iron, 2 sulfur cluster binding pathway through regulating expression levels of *CLIC5* mRNA. Third, rs4819522 polymorphism (nonsynonymous coding) leads to a Thr-to-Met transition at position 350 of TBX1 and involves in the pathways about gland and endocrine system development. In summary, our results demonstrated four candidate SNPs in three genes (*CYP26B1* rs2241057, *CISD1* rs2251039, rs2590370, and *TBX1* rs4819522 polymorphisms), which were involved in six potential pathways to influence serum PSA levels.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Prostate cancer (PCa) is the second most commonly diagnosed cancer and the sixth leading cause of cancer-related deaths worldwide in males (Jemal et al., 2011). In the United States, an estimated 233,000 new cases and 29,480 PCa-related deaths are projected to occur in 2014 (Siegel et al., 2014). The routine application of prostate-specific antigen (PSA) screening has contributed to the early diagnosis and treatment of PCa, which could improve the survival and prognosis of PCa patients remarkably (Hayes and Barry, 2014; Schroder et al., 2012).

PSA, a protein produced by the prostate gland, has been widely applied for PCa screening (Carter et al., 2013; Gill and Wu, 2013). However, the limitation in both specificity and sensitivity has challenged the diagnostic performance of PSA screening test and lead to considerable debate in the medical community (Crawford and Abrahamsson, 2008). For example, a PSA of 2.6 ng/ml has a sensitivity of 40.5% and specificity of 81.1% for detecting PCa; while at 4.1 ng/ml, the values are 20.5% and 93.8%, respectively (Thompson et al., 2005). To improve the diagnostic performance of the PSA screening test, several novel parameters have been proposed, including PSA velocity, free PSA, and PSA density, and the cutoff values adjusted for age and ethnicity (Greene et al., 2009; Stephan et al., 2014). Emerging evidence has suggested that around 40–45% of the variance in serum PSA levels among men in the general

Abbreviations: PSA, prostate-specific antigen; PCa, prostate cancer; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; ICSNPathway, the identify candidate causal SNPs and pathways; CYP26B1, cytochrome P450, family 26, subfamily B, polypeptide 1; CISD1, CDGSH iron sulfur domain 1; TBX1, T-box 1; TERT, telomerase reverse transcriptase; MSMB, microseminoprotein beta; FGFR2, fibroblast growth factor receptor 2; TBX3, T-box 3; HNF1B, HNF1 homeobox B; KLK3, kallikrein-related peptidase 3; ATF7IP, activating transcription factor 7 interacting protein; SLC45A3, solute carrier family 45, member 3; LD, linkage disequilibrium; PBA, pathway-based analysis; NCBI, National Center for Biotechnology Information; HWE, Hardy–Weinberg equilibrium; QC, quality control; CEU, Utah residents with ancestry from northern and western Europe; YRI, Yoruba in Ibadan; KEGG, the Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology; SPES, significant proportion based enrichment score; FDR, false discovery rate; SD, standard deviation.

* Corresponding author.
E-mail address: urojiarp@126.com (R.-P. Jia).

[1] Yu-Zheng Ge, Zheng Xu, and Lu-Wei Xu contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2014.08.044
0378-1119/© 2014 Elsevier B.V. All rights reserved.

population could be explained by genetic components, such as single nucleotide polymorphisms (SNPs) (Bansal et al., 2000; Pilia et al., 2006). The discovery of genetic variants influencing the serum PSA levels could lead to better understanding of the molecular mechanisms and utility of PSA test (Helfand et al., 2013).

To date, several genome-wide association studies (GWASs) have been conducted to address the genetic components of serum PSA levels (Gudmundsson et al., 2010; Jin et al., 2013; Pilia et al., 2006; Sun et al., 2013; Terao et al., 2014), and various SNPs associated with serum PSA levels in different genes were found: rs2736098 in *TERT* (5p15.33), rs10993994 in *MSMB* (10q11), rs10788160 in *FGFR2* (10q26), rs11067228 in *TBX3* (12q24), rs4430796 in *HNF1B* (17q12), rs17632542, rs2735839 and rs1058205 in *KLK3* (19q13.33), rs3213764 in *ATF7IP* (12p13), and rs12409639 and rs16856139 in *SLC45A3* (1q32.1) (Gudmundsson et al., 2010; Jin et al., 2013; Sun et al., 2013; Terao et al., 2014). Among these genes, *KLK3* is the gene encoding PSA, with different SNPs to impact the serum PSA levels (Jesser et al., 2008; Nobata et al., 2012; Penney et al., 2011). However, the biological roles of other SNPs remain poorly understood, which posed as one of the key challenges in interpreting GWAS data. The identification of causative SNPs and mechanisms responsible for the observed traits is of vital importance in the interpretation of GWAS data (Lee et al., 2013; Wang et al., 2010). The identify candidate causal SNPs and pathways (ICSNPathway) was developed to identify candidate causal SNPs and the corresponding pathways using GWAS data, through integration of linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) (Zhang et al., 2011). In the present study, we applied bioinformatics methods based on the ICSNPathway analysis to identify causative SNPs and mechanisms, with the aim to generate SNP–gene–pathway hypotheses about the serum PSA levels.

2. Materials and Methods

2.1. Selection of Data Sources

To identify all eligible GWASs on serum PSA levels for further bioinformatics analysis, the National Center for Biotechnology Information (NCBI) dbGap (http://www.ncbi.nlm.nih.gov/gap/), the National Human Genome Research Institute GWAS catalog (http://www.genome.gov/26525384), and the GWAS central (http://www.gwascentral.org/) were comprehensively searched. Furthermore, both PUBMED and EMBASE databases were electronically searched with the various combinations of the following medical subheadings and key words: genome-wide association study or GWAS and prostate-specific antigen or PSA. All electronic searches were conducted up to July 1st, 2014 with no limitations on the language. The studies without detailed summary data were excluded, after the efforts to extract data from the original paper (including supplementary data) and contact the corresponding authors failed.

2.2. Data Extraction and Pre-procession

The primary GWAS data was extracted and pre-processed by two independent authors (Yu-Zheng Ge and Zheng Xu). To reduce the potential impact of genotyping errors, the data set was filtered to exclude the individual SNPs with genotype call rate below 95%, minor allele frequency below 0.01, and deviating from the Hardy–Weinberg equilibrium (HWE) test ($P < 1.0 \times 10^{-3}$). After the quality control (QC) filtering, the full list of GWAS SNP *P*-values was prepared for the subsequent analysis. During the process of data extraction and pre-procession, all discrepancies were resolved through discussing with a third author (Lu-Wei Xu), and consensus on each item was achieved eventually.

2.3. Identification of Candidate SNPs

During the first stage of ICSNPathway analysis, the candidate causal SNPs were pre-selected by LD analysis and functional SNP annotation based on the most significant SNPs (Zhang et al., 2011). Of note, the *P*-value threshold for the most significant SNP was defined as $<1.0 \times 10^{-3}$. In the LD analysis, the SNPs in LD with ($r^2 > 0.8$) and located in the flanking region (with distance up to 200 kb upstream and downstream) of the most significant SNPs from GWAS were searched. To capture more possible candidate causal SNPs, the extended data set was applied by including HapMap (http://hapmap.ncbi.nlm.nih.gov) data (Altshuler et al., 2010). In addition, SNAP (http://www.broadinstitute.org/mpg/snap/) was adopted to plot regional views of associations or LD structures (Johnson et al., 2008). Subsequently, the functional annotation on these SNPs was conducted by querying the international SNP function annotation databases, namely, Ensembl database (http://www.ensembl.org) (Flicek et al., 2010), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2013), SIFT (http://sift.jcvi.org) (Kumar et al., 2009), and SNPs3D (http://www.snps3d.org) (Yue et al., 2006).

2.4. Genotype–Phenotype Correlation Analysis

We have extracted the allelic and genotypic frequencies of the selected SNPs from the International HapMap Project (phase II, release 23), which consists of 3.96 million SNP genotypes from 270 individuals (He et al., 2013; International HapMap Consortium, 2003). These 270 subjects were recruited from three ethnic populations, including 90 Caucasians (Utah residents with ancestry from northern and western Europe, CEU), 90 Asians (45 unrelated Han Chinese in Beijing and 45 unrelated Japanese in Tokyo), and 90 Yoruba in Ibadan (YRI) subjects (International HapMap Consortium, 2003). Meanwhile, the mRNA expression data were derived from the lymphoblastic cell lines of the same 270 subjects (Stranger et al., 2007), and available online from SNPexp (http://app3.titan.uio.no/biotools/help.php?app=snpexp/) (Holm et al., 2010).

2.5. Identification of Candidate Pathways

In the second stage of ICSNPathway analysis (Zhang et al., 2011), PBA was conducted to annotate biological mechanisms to pre-selected candidate causative SNPs, by integrating data from four pathway databases including the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg), BioCarta (http://www.biocarta.com), gene ontology (GO, http://www.geneontology.org), and MSiDB (http://www.broadinstitute.org/gsea/msigdb). In brief, SNP label permutation and normalization were applied to generate the distribution of significant proportion based enrichment score (SPES) and to correct gene variation. Based on the distributions of SPESs, a nominal *P*-value and a false discovery rate (FDR, cutoff value: 0.05) were computed. Of note, the cutoff number of genes in the pathway was set as: minimum 5 and maximum 100, to avoid the inclusion of too general biological process.

2.6. Statistical Analysis

The gene expression levels were presented as mean ± standard deviation (SD), and the difference in expression levels between two strata was evaluated by two-side Student's *t* test. In addition, the general linear regression modeling was used to evaluate the trend of transcript expression levels by genotypes. The statistical analysis was performed with SPSS software (version 20.0; SPSS Institute, Chicago, IL, USA), and $P < 0.05$ was considered significant.

Download English Version:

https://daneshyari.com/en/article/2816150

Download Persian Version:

https://daneshyari.com/article/2816150

Daneshyari.com