

南方医科大学 2013 级硕士学位论文

单核苷酸多态性北京杭州广州三地中国人群前列腺癌相关性研究

Influence of a single-nucleotide polymorphism of the  
DNA mismatch repair-related gene exonuclease-1 (rs9350)  
with prostate cancer risk among Chinese people

课题来源： 自选

学 位 申 请 人	李鹏举
导 师 姓 名	黄鹏 副教授
专 业 名 称	外科学（泌尿外）
培 养 类 型	学术型
培 养 层 次	医学硕士
所 在 学 院	第二临床医学院

答辩委员会主席

黄 健 教授

答辩委员会委员

高 新 教授

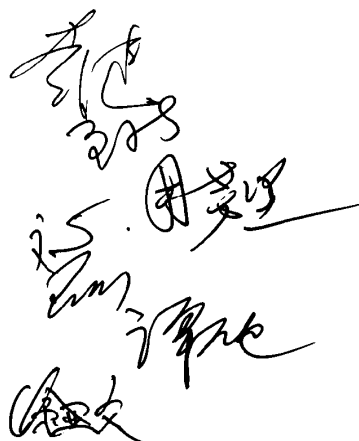
周芳坚 教授

刘久敏 教授

王树声 教授

谭万龙 教授

徐亚文 教授



2016 年 3 月 12 日 广州

# 单核苷酸多态性北京杭州广州三地中国人 群前列腺癌相关性研究

硕士研究生：李鹏举

导师：黄鹏 副教授

## 摘 要

### 研究背景：

前列腺癌是指发生在前列腺的上皮性恶性肿瘤。在 2004 年的 WHO《泌尿系统及男性生殖器官肿瘤病理学和遗传学》里面前列腺癌各种病理类型分别为腺癌（腺泡腺癌）、导管腺癌、尿路上皮癌、鳞状细胞癌、腺鳞癌等等。其中前列腺腺癌占 95%以上，因此，通常我们所说的前列腺癌就是指前列腺腺癌。2012 年我国肿瘤登记地区前列腺癌发病率为 9.92/10 万，列男性恶性肿瘤发病率的第 6 位。发病年龄在 55 岁前处于较低水平，55 岁后逐渐升高，发病率随着年龄的增长而增长，高峰年龄是 70~80 岁。家族遗传型前列腺癌患者发病年龄稍早，年龄≤55 岁的患者占 43%。在世界范围内，前列腺癌是影响男性健康的最常见恶性肿瘤之一，发病率在所有男性恶性肿瘤中位居第二，同时有着明显的地理和种族差异。澳大利亚/新西兰勒比海的纳维亚地区最高，亚洲及北非相对较低。根据美国癌症协会发布的报告，美国 2013 年前列腺癌患者数目达到 238950 人，占男性所有恶性肿瘤的 28%，死亡人数为 29720 人；在欧洲，2008 年新发前列腺癌患者数为 382000 人，90000 名患者死亡。虽然亚洲的发病率低于欧美国家，但是其近年来的增长速率远远超过欧美国家的增长速率。根据 2012 年世界普查报告，也可以看出前列腺癌在中国的发病率有了明显的增长：1998-19974 年期间发病率的增长率为 2.1%，但是在 1994-2002，前列腺癌发病率增长率高达 13.4%。

引起前列腺癌的危险因素有待研究，一些已经可以确定的危险因素包括年

龄、种族和遗传性。2个或者2个以上的一级亲属患病，患病危险性会增加5-11倍。通过全基因组相关研究，现阶段已经发现50余个单核苷酸多态性（SNP）与前列腺癌发病风险相关。有报告证实，发生5个或者5个以上的SNP的男性与无SNP的男性相比，罹患前列腺的风险比高达9.46。

SNPs是指在基因组的复制过程中，单个核苷酸变异所引起的DNA序列多态性。与突变不同的是其变异发生率，突变小于1%。发生率大于1%则被称为单核苷酸多态性。SNPs是常见人类可遗传变异，广泛存在于人类基因组中，占已知多态性的90%以上。大约每1000个基因组中就有一个SNP，故SNPs大概总共 $3 \times 10^6$ 个。

SNPs就是单个碱基变异的产物，通常我们谈到的SNP指的是转换和颠换。在我们所表述的单个基因的变异的情况中，一般都是指碱基的转换（transition）或者颠换（transversion），亦可以是碱基的插入或者缺失。并非所有的SNPs是有意义的，其中一些，例如非编码区的SNPs是没有意义的，并不影响蛋白质的功能或者基因的表达。

SNPs在人群中的分布广泛，同时作为一个相对稳定的遗传标记，能存在于从个体的出生至死亡的所有时期，使疾病的早期遗传分析成为可能。在现阶段，研究已证实SNPs能够影响基因的功能，有一部分SNPs能通过多种途径影响前列腺癌的发生、发展以及预后。所以，SNPs不但作为遗传性基因影响和证明肿瘤的发生与发展，也为肿瘤的普查提供了一个有效的思路与检测方法。

SNPs主要可以分成几部分：约95%的SNPs在非编码区内部，其余的5%分布在基因调控区里面，称为基因调控区SNPs（regulatory SNPs, rSNPs）；编码SNPs（coding SNPs, cSNPs）即是分布在基因编码区中的SNPs。若cSNPs所编码的氨基酸序列与之前没有发生变化则称为同义SNPs（synonymous SNPs, sSNPs），而编码的氨基酸序列与未变异的比较出现改变则称为非同义SNPs或错义SNPs（non-synonymous SNPs, nsSNPs）。cSNPs虽然在所有变异中所占比例不多，但是在遗传性疾病研究中却具有重要意义，因此cSNP的研究更受关

注。在 cSNPs 中，大约有一半为非同义 SNPs。

SNPs 可作为第三代遗传标志物，突变水平不及微卫星，只有两种等位型；不过因为其在人体基因组中分布频密，因而总数量巨大，所以相较微卫星来说，它的遗传性质相对更为稳定。相比第一、第二代遗传标志物，SNPs 筛查具有更简便分型检测并实现自动化分析的特点，因为在其筛查过程中，一般只需要阳性/阴性 (+/-) 分析。SNPs 已代替微卫星技术应用于基因研究领域，对于不同个体的表型差异、不同个体或者群体对疾病，尤其是复杂性疾病易感性、药物耐受性和环境适应性的差异分析，其已成为最为合适的标志。SNPs 目前应用领域包括分子诊断、临床检验、法医学、病院检测、遗传疾病和新药研发等等。

现阶段比较普遍的 SNPs 的分析方法大致有三种：第一类为遗传流行病学的相关性分析，主要通过调查 SNPs 在疾病易感性、药物反应性及其他相关遗传表型差异的手段；第二类是基于细胞水平和生化水平的分析，通过细胞通路、相关酶活性的分析来描述 SNPs 改变对于基因功能方面的影响；第三类为 SNPs 对于基因表达的影响的分子机制研究，主要是对 SNPs 在基因转录、翻译及蛋白表达相关进行分析，研究其机制。

日本的冈山大学科研小组经过与癌症相关的基因中的 48 个非同义 SNPs 的筛选，指出在 12 个 SNPs 中相关的 10 个基因对前列腺癌这种恶性肿瘤的发生与发展具有非常重要的意义。在这 10 个基因里，包含了 1 个代谢酶基因、1 个凋亡调节基因、2 个 DNA 修复基因、5 个肿瘤抑制基因和 1 个染色体分离相关基因。在这里面有 9 个 SNPs 与恶性肿瘤的发生与发展的关系是首次被发现。在预测今后 30 年里面的前列腺癌罹患风险方面，风险组最高的可达 29%，而风险组比较低的可为 0.6%，其中风险组最低的仅仅约为 0.2%。同时我们在预先评估前列腺癌罹患风险方面，发现这 12 个 SNPs 影响效应可以发生累积，风险组高与低的罹患率相比较，其 OR 值高达 47.4。所以，在肿瘤的患病风险遗传学的评估中应用上与多种恶性肿瘤有关联性的 SNPs 的整合，与现有方式相比，对于罹患前列腺癌风险较高的人群的普筛是具有更大的准确性和更高的效率，并在肿瘤的

早期诊断和预防措施的选择方面具有优势。

日本人群所在的地理位置与中国人群相近，同时两国人群也具有相似的人种起源，因此两国的遗传背景有很高的共同性。经过在日本大规模筛查得到的前列腺癌罹患风险相关的 SNPs 可能在中国人群中也会有部分适用。因此，在亚洲的四个国家，中、日、韩和新加坡联合进行了一项跨国际的多中心研究，目的是为了检验这一假说的准确与否，并将在日本的筛查结果推广到亚洲其他的国家，同时检验筛选出的前列腺癌高风险相关性 SNPs 在亚洲其他人群中的效力。本次试验中所选位点是来自日本冈山大学科研小组的筛查成果、一致于此次多中心试验的研究内容。

#### 目的和意义：

1、分析调查前列腺癌组和对照组人群中多个特定单核苷酸多态性位点的分布情况，探讨多个特定位点与中国人群前列腺癌患病风险的相关性，寻找阳性位点以期于指导前列腺癌早期诊断和预防。

2、尝试建立多个阳性单核苷酸多态性位点与中国前列腺癌罹患风险的预测模型。

#### 方法：

自 2009 年 1 月至 2013 年 4 月，在北京首钢医院吴阶平泌尿外科中心、浙江大学第一附属医院及南方医科大学附属珠江医院医院招募 253 例前列腺癌（Pca）患者和 214 例非前列腺癌及其他健康人群（对照组）。病例组为术后病理检查或者通过影像学、前列腺穿刺活检证实前列腺癌患者，初次发病或复诊均可。对照组为非前列腺癌及其他健康人群，主要条件是需与 Pca 组人员在年龄上相一致，同时也是在同一时段同一医疗单位住院或者门诊就诊的其他患者。记录患者的临床资料如发病年龄、Gleason 评分、TNM 分期、PSA 水、治疗的方法选择及治疗后疗效等等，同时发给患者问卷进行调查，其中的内容包括有饮酒吸烟史，家族史及饮食风俗等。抽血之前，需告知患者取得其同意同时签署知情同意书。使用 EDTA 抗凝管将抽取到的大概 2ml 外周静脉血保存在-20℃(或

- 80℃) 条件下。全血基因组 DNA 则通过使用 TIANamp Blood DNA Kit 血液基因组 DNA 提取试剂盒进行提取。提取到的 DNA 基因分型则通过使用多重 SNaPshot SNP 分型技术来完成, 结果所得出的基因有 8 个, 命名方式主要是根据其所在的位点来命名: RAD17 SNP、SMARCAD1 SNP、MMP27 SNP、AXIN2 SNP、PSMD8BP1 SNP、CASP9 SNP、DCLRE1B SNP、BARD1 SNP。

#### 统计学的方法:

使用 Pearson's  $X^2$  检验方法用来比较各临床指标之间的差异 (患病组与对照组之间、多种基因型之间)。各个对照组的 SNP 位点的基因型分布频率使用 Pearson's  $X^2$  检验方法, 主要是进行 Hardy-Weinberg 平衡检验同时通过 R 语言来分析。logistic 回归分析主要用于计算特定的基因型与本次研究所涉及人群的前列腺癌罹患风险比值比 (ORs)、与之相对应的 P 值以及其 95% 的可信区间 (Confidence intervals, CIs), 来分析各种基因型和前列腺癌罹患相关与否。前列腺癌罹患风险具有相关性的 SNPs 与肿瘤危险分期和年龄等因素以分层的方式进行分析。在肿瘤危险分期这一方面, 我们将局限型前列腺癌和进展型前列腺癌作为不同分层。局限型的前列腺癌的分层标准设定为: TNM 分期 T1-2; N0/NX; M0/MX; Gleason 评分 2-7 分; PSA  $\leq$  50 ng/ml。进展型前列腺癌组的分层标准则设定为: TNM 分期 T3/4; N+; M+; Gleason 评分 8-10 分; PSA  $>$  50 ng/ml。局限型入组需满足全部标准, 进展性则只需要满足标准被其中之一。在以年龄分层这一方面, 我们主要通过其中前列腺组以初发年龄、对照组以入组年龄为标准的方式, 以 72 岁年龄为分界线, 最后分成  $\leq$  72 岁组以及  $>$  72 岁组。将患者吸烟饮酒等情况作为前列腺癌罹患肿瘤相关性的多因素回归模型里面的影响因素。此次我们使用 SPSS 20.0 统计学软件, 在  $P < 0.05$  时可认定为差异有统计学意义, 所有统计检验为双侧检验。

#### 结果:

1、基因错配修复外切酶 1 相关单核苷酸多态性 (rs9350) 相关 SNPs 分析: rs9350 CC 等位基因片段在前列腺癌组与对照组中有相同百分比 (32.8%与

45.3%),rs9350 CT 和 TT 基因型在前列腺癌组与对照组中的百分比分别为 53.0% 与 44.4% (CT)、14.2%与 10.3% (TT)。与 CC 基因表型相比, CT 基因表型有更高的前列腺癌罹患风险 (OR=1.648,95%CI=1.112-2.444, p=0.013; 校正 OR-1.678,95%CI=1.176-2.500, p=0.05), C 与 T 基因表型之间具有统计学意义 (p=0.09)

2、不同年龄层的 SNPs(rs9350)分析: 在已有的数据分析中, 我们将 72 岁作为年龄分界线 (前列腺癌患者组及对照组中位年龄分别为 71.4 和 70.2), 在此年龄段上线具有近似的病例数。在大于 72 岁的年龄组中, rs9350 CT/TT 基因型具有统计学意义 (p=0.032); 携带 C 等位基因具有更高的前列腺癌罹患风险 (校正 OR=1.464,95%CI=0.998-12.147, p=0.041)。

3、不同肿瘤危险程度分层 SNPs (rs9350) 分析: 根据肿瘤侵袭性分为局限性前列腺癌组和进展性前列腺癌组。在局限性前列腺癌组中 (对照组为正常人或非肿瘤患者) rs9350 CT/TT 基因型与更高的前列腺癌罹患风险明显相关 (校正 OR=1.978,95%CI=1.070-3.022, p=0.027)。

4、其他临床指标 SNPs (rs9350) 分析: 除 PSA 水平之外 (PSA≤10 ng/ml 相比较 PSA>10 ng/ml) CT/TT 基因型存在率更高 (p=0.006), 其余临床指标与 SNPs 之间并无明显的相关性。

#### 结论:

通过本次病例对照研究, 我们发现: EXO1 SNP [rs9350]与中国人群前列腺癌具有显著相关性。在已知的研究结果中, 本次研究结果为首次发现 EXO1 SNP [rs9350 CT/TT: C/T]和前列腺癌罹患风险显著相关。尽管更大的数据样本、更多的人种仍然需要进一步试验验证, EXO1 SNP [rs9350]具有成为前列腺癌罹患风险早期筛查及评估的有效分子生物学标记。

关键词: 单核苷酸多态性, 前列腺癌, 易感性, EXO1; rs9350; 基因错配修复。

# **Influence of a single-nucleotide polymorphism of the DNA mismatch repair-related gene exonuclease-1(rs9350) with prostate cancer risk among Chinese people**

**Name: Pengju Li**

**Supervisor: Peng Huang**

## **ABSTRACT**

### **Background:**

Prostate cancer is a kind of malignant epithelial tumor which grows in prostate. The pathological pattern contains glandular cancer, ductal adenocarcinoma, urothelium carcinoma, squamous-cell carcinoma and adenosquamous carcinoma and 90% of these pathological types are glandular cancer. Thus, generally we informed about prostate cancer means that it is prostate glandular cancer. In 2012, the prostate cancer morbidity is 9.92/100 000 at the place of cancer registration area in china, takes the sixth place of all male malignant neoplasm. The morbidity of the disease is low under age 55, while it is increasing along with the age after age 55, the peak is age 70-80. An earlier morbidity occurs in hereditary familial prostate cancer patients, 43% patients are at age younger than 55. In world wide, prostate cancer is one of the most common malignant tumor which affect male's health, among all the carcinoma in male, prostate cancer ranks at the second place. There are some significant differences of morbidity when the patient come from different places and different races. It is the highest in Australia/New Zealand, Caribbean while it is lowest in Asia and North Africa. According to the report of American Cancer Society, the number of



## ABSTRACT

---

prostate cancer patients is 238950 in USA in 2013, occupies 50% in all malignant tumor, while the death number is 29720. In Europe, there are 382000 new cases in 2013 while the death number is 90000. Although the morbidity in Asia is lower than Europe and America, the rising rate of the morbidity is on the other side. According to the world census report, it is identified that the rate of prostate cancer morbidity in China is increasing these years. The rising rate of prostate cancer morbidity is 2.1% in 1998-1999 while it is 13.4% in 2000-2002.

The risk factor of prostate cancer is still unclear, a few confirmed factors are age, races and hereditary. If there is 2 or more than 2 patients in your family, the risk is increasing by 5-11 times. It is confirmed that compared with non-SNP male, the risk is 9.46 when there are 5 or more than 5 SNPs existing.

Single Nucleotide Polymorphisms (SNPs) are DNA sequence polymorphisms caused by a single nucleotide mutation during the procedure of DNA sequence duplication transcription. The difference between mutant and SNPs is the rate of variation, less than 1% is called the mutant and more than 1% is called single nucleotide polymorphism. SNPs is a common heritable variation which exists human genetic variation in the human gene widely. Among all the polymorphism, the SNPs occupies over 90%. In the human genome each of about 1000 bp is SNP and the number of all SNPs is about  $3 \times 10^6$ .

A single mutation, it could be a basic group converted caused by transition or transversion, it also could be caused by insertion or deletion of single nucleotides. Generally speaking, SNPs we talked about are mostly caused by transitions and transversions. Most of SNPs are insignificant (such as non-coding region SNPs) because of there is no effects on protein function or gene expression.

There are several kinds of SNPs. About 95% of the SNPs located in the non-coding region, a little fraction of SNPs are located in the gene regulatory region, were known

as gene control region SNPs (rSNPs). which is located in the coding region of the gene coding SNPs called SNPs (cSNPs). Furthermore, if cSNPs does not change the amino acid sequence encoded, it could be called synonymous SNPs (sSNPs). But if it change the amino acid sequence, it could be called non-synonymous SNPs or missense SNPs (nsSNPs). Although cSNPs accounts for less percentage in all variation, it means a lot in the research of hereditary disease .Hence, the research of cSNPs draw more attention. There are about 50% nsSNPs in all cSNPs.

SNPs are widely exists in human population, at the same time, SNPs could be relatively stable existing from birth to death as a biological marker. Thus, a genetic analysis in the early diagnosis of disease is reality. Nowadays, it is confirmed that SNPs is able to affect the function of DNA, some prostate cancer related SNPs could affect the occurrence , development and prognosis of prostate cancer by many methods. Thus, SNPs can be able to affect and provide the tumor occurrence and development as a biological marker, provide an effective method of common cancer screening as well.

There are two allelic type which is less than microsatellite variability while as the third generation of SNP genetic markers,. Owe to the gaint number of frequency distribution in all of the genome, compared with microsatellite genetic markers it is more stable.It is only need to analysis positive or negative (+ / -) when generally SNPs in the genomic screening process, so it is not hard to operate genotyping analysis automatically. Compared to the first and second generation of genetic markers, are more suitable for exploring population-based genetic condition and other aspects of the study in complex genetic diseases. It has been replaced microsatellite marker technology in the field of genetic research. SNPs were used to explain the phenotypic differences between individuals, different complex diseases susceptibility of different groups or individuals, drug and environmental factors

## ABSTRACT

---

response. It has been widely used in clinical diagnosis, forensics, pathogen detection, genetic diseases, new drug development and other aspects. Common methods of analysis of SNPs can be divided into three categories. The first category is association analysis based genetic epidemiology method, including the research of SNPs in disease susceptibility, drug reactions and other phenotypic differences. The second category is about the analysis in cellular and biochemical levels including enzyme activity and other aspects in cell signaling pathways to illustrate the impact of SNPs on gene function. The third category focus on SNPs impacting on molecular mechanisms of gene expression. Through the role of SNPs in the gene transcription, translation and protein expression to explore the mechanism of SNPs affecting gene function.

The research group in Okayama University had screened 48 missense SNPs (non-synonymous SNPs) in cancer-related genes. They found 12 SNPs in 10 genes and these SNPs have a significant impact on the incidence of prostate cancer. This 10 genes include five tumor suppressor genes, two DNA repair genes, a metabolic enzyme gene, a chromosome segregation gene and an apoptosis-related gene. Nine SNPs which are correlated with prostate cancer were first discovered. This 12 SNPs in prediction of prostate cancer risk also has a cumulative effect. The highest risk group's OR value can be up to 47.4 comparing with the low-risk group. The highest risk group the incidence of prostate cancer in the next 30 years was 29%, while the low-risk group was 0.6% and the lowest risk group was only about 0.2%. Thus, through the integration of multiple tumor-associated SNPs risk to conduct tumor genetics evaluation, can further improve the accuracy and efficiency of prostate cancer screening in high-risk populations. It also can improve the prevention or early diagnosis of prostate cancer.

Because of similar geographic and ethnic origin, Chinese population and Japanese

population have great similarity in genetic background. These prostate cancer risk SNPs obtained by large-scale screening of the Japanese population may also apply to the Chinese population. In order to test this hypothesis and further expand this research results from Japan into China and other Asian countries, China, Japan, South Korea and Singapore jointly launched an international multi-center study. This study was designed to test the effectiveness of these prostate cancer risk SNPs in other Asian populations. The selected SNPs in this article is derived from the achievement of Okayama University research group and is consistent with multi-center study.

### **Purpose**

1. To investigate the association between SNPs and the risk of prostate cancer in Chinese population, provide sites to guidance for prostate cancer prevention and early diagnosis. Using positive SNP for prostate cancer prevention and early diagnosis.
2. Try to establish a risk model with number of positive single-nucleotide polymorphisms in predicting the risk of prostate cancer among Chinese Han population.

### **Method**

From January 2009 to April 2013, recruited 253 cases of prostate cancer (Pca) patients and 214 cases of non- prostate cancer patients and healthy volunteers ( control group ) from Southern Medical University zhujiang Hospital and the first affiliated hospital of Zhejiang university, Peking University Shougang Hosipitol. Case group included biopsy or pathologic diagnosis of prostate cancer patients , the initial onset or referral. Control group included patients with prostate cancer and other malignancies with age matched patients maintained. The could be in the same patient hospitalization or outpatient medical units during the same period. Recorded clinical data such as age at onset, PSA, Gleason score, TNM stage, treatment, and treatment effects and distributed questionnaires to patients. The questionnaires included

smoking history, drinking history, family history, diet habit and so on.

A 2 mL blood sample was obtained from each participant, and it remained at room temperature for no more than six hours. Genomic DNA was extracted using a TIANamp Blood DNA Kit (TIANGEN Biotech, Beijing, China) according to the manufacturer's instructions, and was stored at -20°C. We analyzed the Axin2 SNP (rs2240308) and seven SNPs of the other genes (not shown in this paper) using these samples. The genetic analyses were performed using the ABI SNaPshot multiplex system (Life Technologies Corporation, Carlsbad, CA, USA).

### **Statistical analysis**

We compared the proportion (percentage) of the each genotype and allele of the SNP [rs9350: G/A] and other seven SNPs in the controls and prostate cancer cases. The association between the SNP and incidence of prostate cancer was analyzed using a logistic regression model. The odds ratio (OR), 95% confidence interval (CI) and corresponding p values for the association between the prostate cancer risk and the genotypes or alleles were calculated. The data for each genotype or allele was compared with that of the common homozygote or allele as the reference group. We also stratified our analyses by the age of the patient at diagnosis ( $\leq 72$  or  $> 72$  years) and by the aggressiveness of the disease (localized or advanced prostate cancer). Localized prostate cancer inclusion criteria are T1-2, N0, M0, Gleason score 2-7, and PSA levels  $\leq 50$  ng/mL. Advanced prostate cancer inclusion criteria are T3/4 or N+ or M+ or Gleason score 8-10 or PSA levels  $> 50$  ng/mL.

In these analyses, the data were adjusted for the age, smoking status and drinking status. The data are shown as the means  $\pm$  standard deviation (SD). The Chi-square test was used to compare the distribution of the control males and prostate cancer patients or of the clinical characteristics. The Mann-Whitney U test was also performed to analyze the statistical significance of differences in the age and PSA

level at diagnosis. All statistical analyses were conducted using the SPSS software program, version 20.0. The differences were considered to be significant for values of  $p < 0.05$ .

## Results

### 1, SNP analysis of rs9350 in prostate cancer

The allele and genotype frequencies of rs9350 among cases and control subjects are listed in Table 3. The same proportion of rs9350 CC allele (32.8 vs. 45.3 %) was observed in prostate cancer patients and control subjects. The proportions of rs9350 CT and TT genotypes in prostate cancer patients versus control subjects were 53.0 versus 44.4 % and 14.2 versus 10.3 %, respectively. Compared with the CC genotype, the CT genotype may increase prostate cancer risk (OR=1.648, 95 % CI=1.112–2.444,  $P=0.013$  and adjusted OR=1.678, 95 % CI=1.130–2.494,  $P=0.010$ ). Further, the CT/TT genotypes were significantly associated with increased prostate cancer risk (OR=1.698, 95%CI=1.166–2.473,  $P=0.006$  and adjusted OR=1.714, 95 % CI=1.176–2.500,  $P=0.005$ ), and the C allele had a significant statistically compared with T allele ( $P=0.009$ ) of EXO1 (rs9350).

### 2, SNP analysis of rs9350 for different ages at diagnosis

In stratified analysis, we selected an age at diagnosis of 72 years as the cutoff, with similar numbers of cases above and below the cutoff age. In the >72 years group, the CT/TT genotypes of rs9350 were statistically significant ( $P=0.032$ ) as well as the C allele compared with the T allele (adjusted OR=1.464, 95 % CI=0.998–12.147,  $P=0.041$ ) for prostate cancer risk.

### 3, SNP analysis of rs9350 in different prostate cancer classes

In another stratified analysis, we classified the prostate cancer patients into localized and advanced disease in terms of the aggressiveness. In the localized disease subgroup (controls: all cancer-free patients or healthy men), the CT/TT

genotypes of rs9350 were associated with significantly increased prostate cancer risk (adjusted OR=1.798, 95 % CI=1.070–3.022, P=0.027).

4, SNP analysis of rs9350 for different patient factors

Except for PSA levels at diagnosis, no significant association was observed between rs9350 and clinical factors, including aggressiveness, Gleason scores, age at diagnosis, smoking status, or drinking status . Subjects with PSA levels of  $\leq 10$  ng/mL were more likely to have the CT/TT genotype than those with PSA levels of  $>10$  ng/mL (P=0.006).

**Conclusion:**

this variation in location could have helped obtain convincing results, although the sample size was small. Other studies have demonstrated a significant association between rs9350 and development of pancreatic cancer in American and Taiwanese populations.

To our knowledge, our study provides the first evidence for the association between the EXO1 SNP rs9350 and prostate cancer risk in Chinese people. we provided evidence of independent replication for the EXO1 SNP rs9350 associated with prostate cancer risk in China. Although additional studies with larger and more diverse populations and functional analysis of the SNP are necessary to confirm and extend our findings, we believe that the EXO1 SNP rs9350 could be a useful biomarker for assessing predisposition to prostate cancer and early diagnosis of the disease.

**Keywords:** Prostate cancer . Single-nucleotide polymorphisms . EXO1 . rs9350 .DNA mismatch repair.

## 目 录

摘 要.....	I
ABSTRACT .....	i
前 言.....	1
方法部分.....	14
试验结果.....	25
讨 论.....	32
参考文献.....	35
附件 1.....	39
附件 2.....	40
缩略词表.....	42
攻读学位期间的发表的学术论文.....	45
致 谢.....	46
学位论文原创性声明.....	47



## 前 言

### 一、 什么叫做单核苷酸多态性。

在 2003 年 4 月 15 日, 美国, 日本, 德国, 法国, 英国等六个国家共同宣布人类基因组序列图的完成, 从而标志着人类基因组计划的成功。1990 年人类基因组计划 (Human Genome Project, HGP) 正式启动, 其宗旨主要是为了明确人类染色体的组成 (主要是指单倍体), 其中的核苷酸序列大概有三十亿个碱基对用来完成基因图谱的绘制。研究图谱所携带的基因鉴定和序列, 最终目标是为了破译人类遗传信息。同时我们还发现, 人类遗传变异的范围的程度远远超出了预期。在任何两个不相关的人的 DNA 序列中, 相同的部分可达 99.9%, 但是正是在剩余的 0.1% 里是在 DNA 序列中个体易感不同程度的疾病的, 响应于不同的药物的差异性的原因[1]。经过研究发现, 深入探究 DNA 多态性位点与一些常见的疾病相关性, 有利于发现复杂的人类疾病的原因。

SNPs 是指在基因组的复制过程中, 单个核苷酸变异所引起的 DNA 序列多态性。与突变不同的是其变异发生率, 突变小于 1%。发生率大于 1% 则被称为单核苷酸多态性[2]。SNPs 是常见人类可遗传变异, 广泛存在于人类基因组中, 占已知多态性的 90% 以上。大约每 1000 个基因组中就有一个 SNP, 故 SNPs 大概总共  $3 \times 10^6$  个[3]。

SNPs 就是单个碱基变异的产物, 通常我们谈到的 SNP 指的是转换和颠换。在我们所表述的单个基因的变异的情况中, 一般都是指碱基的转换 (transition) 或者颠换 (transversion), 亦可以是碱基的插入或者缺失。并非所有的 SNPs 是有意义的, 其中一些, 例如非编码区的 SNPs 是没有意义的, 并不影响蛋白质的功能或者基因的表达。

SNPs 主要可以分成几部分: 约 95% 的 SNPs 在非编码区内部, 其余的 5% 分布在基因调控区里面, 称为基因调控区 SNPs (regulatory SNPs, rSNPs); 编码 SNPs (coding SNPs, cSNPs) 即是分布在基因编码区中的 SNPs。若 cSNPs 所编码的氨基酸序列与之前没有发生变化则称为同义 SNPs (synonymous SNPs,

sSNPs)，而编码的氨基酸序列与未变异的比较出现改变则称为非同义 SNPs 或错义 SNPs (non-synonymous SNPs, nsSNPs)。cSNPs 虽然在所有变异中所占比例不多，但是在遗传性疾病研究中却具有重要意义，因此 cSNP 的研究更受关注。在 cSNPs 中，大约有一半为非同义 SNPs[4]。

分析 SNPs 方式可分为三类[5]，第一类是基于遗传流行病学相关，通过研究敏感性单核苷酸多态性与疾病的 SNPs 的分析常用方法来研究药物反应有关的遗传性和其它表型的差异；第二类对细胞和生化水平的分析,通过分析酶的活性，以及细胞信号传导途径的其它方面来说明 SNPs 的基因各种功能的影响;第三类是基于单核苷酸多态性的分子机制影响，由基因转录翻译和蛋白 SNPs 的基因表达进行了分析，探索了 SNPs 影响基因各种功能的机制。

## 二、三代遗传标志物的变化

在第一代遗传标记主要是建立在传统的 Southern 杂交基础上的一种遗传标志物，美国遗传学家 bostein 在 1980 年首次提出 DNA 限制性片段长度多态性 (Restriction Fragment Length polymorphisms,RFLPS) 作为分子遗传标记[6]，这使得遗传标记摆脱了孟德尔时代变异数目不足的形态学问题。RFLPS 被广泛应用于遗传图谱，基因定位和生物进化的分类的组成和建设。利用 RFLSP 的原理是通过特异性杂交的探针，来检测因为限制性核酸内切酶的限制位点发生的突变，比如说单个碱基的插入、缺失或者替换，甚至直接发生突变从而引起了各个被剪切后基因片段的大小不等，从而可以用来检测比较在 DNA 中突变（称为多态性）水平的个体差异[7]。

第二代基于 PCR 遗传标记的遗传标志物，被称为微卫星多态性(Microsatellite polymorphisms) 或短串联重复 (Short tandem repeats ,STRs)。STRs 是被布置在两个或更多个核苷酸重复序列里，以及相邻的约 2-10 个碱基对之中，在非编码内含子具有共同的长度但是具有不同形式的重复序列。由于不同的重复序列和重复序列在不同种族中的数量不同，这样一来，不同组中的分配就可以组合出不同的性质，从而构成个 STRs 基因多态性。在不同的 STRs 位点及序列不同的

个体之间，通过识别基因组重复特定位点和特定序列，使得能够我们可以根据其来创建个人遗传图谱。目前已经有超过 10,000 个 STRs 基因被披露。由于 STRs 具有易于分析，分布广，容易察觉，信息巨大，高度多态性和共显性孟德尔遗传等优点，其已广泛应用于遗传图谱，连锁分析，亲子鉴定，基因定位和疾病种类多样性的研究，尤其是现场法医学方面更是有了举足轻重的地位[8]。

第三代是以基因组序列为基础的 SNPs 的遗传标记。人类遗传变异的 SNPs 之中最常见的就是单核苷酸在 DNA 序列中的变异，占有目前已知的变体的 90% 以上。单核苷酸变异具有高达 1000 万以上的总数量，换句话说也就是每 300 至 600 个人类基因组的碱基对中就有一个 SNPs。SNPs 作为第三代遗传标记，只有两个等位基因型，变异的程度与微卫星一样好，但由于总的基因组数量庞大的缘故，因而 SNPs 分布的频率也高，因此具有比微遗传特性更稳定的特性。SNPs 在基因组中的筛选方法通常仅包含正或负 (+ / -) 分析，以此其基因分型很容易达成分析的自动化。相比第一代和第二代遗传标记更适合解析和分析人群的遗传特点和复杂的遗传疾病的研究等方面。并且 SNPs 已取代微卫星标记技术在所引入的基因研究领域里的应用。SNPs 主要可以使用来帮助分析解释复杂疾病环境因素里的个人、团体和不同的疾病之间所具备的的相关性，特别是其敏感性，药物耐受性和差异个体之间的表型差异。因此它已被广泛应用于分子诊断，临床检验，法医学，病原体检测，遗传性疾病和药物开发等诸多方面[9]。

### 三、单核苷酸多态性的研究意义

#### 1、多基因疾病的突变位点和相关分析

过去遗传学家在研究导致疾病因素多采取多基因连锁分析，即以家庭为基础的 (family based linkage analysis) 方法搜索相关的基因。但由于该方法需要分析两代或家族的更多代，以及复杂的疾病 (或多基因遗传病) 在实际情况下遗传模式，并且需要设定继承相关参数，因此通常其模式不能够得到确定。主要原因是常缺乏其他参数，如外显率的连锁分析，所需基因频率等等，因而容易产生较大的不确定性的结果[10]。所以人们不再局限于定位克隆的方法，以及基于

人口的 (population based association analysis) 的相关转向不平衡关联分析。关联分析是比较病例组和对照组之间更具体更明确的遗传标记的等位基因变异频率的差异。如果该组中一般情况下较高频率的个体中遗传标记基因出现, 我们有理由认为是标记基因和此类疾病的关联。此种研究方法的对象包括正常人群和患者, 由于不需要获得家庭数据, 所以很容易得到样品。标记具有相关性的基因不一定是致病遗传性疾病相关基因, 而可能是由于风险位点和相邻多态性标记之间的连锁不平衡引起的。因此, 基于连锁不平衡的多态性标记通常需要的高密度的基因组, 以便可以系统地进行扫描。用于发现与疾病相关联的区域分析的该方法, 通常是双等位基因的 SNPs, 由于其分布广, 突变低和易于实现自动化而成为首选筛选方法[11].

## 2、单基因疾病的未知基因确定及连锁相关。

目前单基因疾病基因定位克隆研究思路可以概括如下: 首先, 具备研究样本 (如家庭, 同胞或者人群等) 可供选择; 然后, 基于连锁不平衡原理, 在全基因组, 染色体片段或候选基因的水平样本里进行大规模扫描遗传标记; 最后, 对于获得的数据分析以确定哪些基因或染色体片段和疾病的研究之间具备相关性, 并最终明确疾病易感基因位置靠近哪一个遗传标记或目的基因区域相关数据的最适合统计方法。自 1980 年以来, 全基因组关联分析方法表明这种方法是有效果的, 这种方法已使用对超过 1300 个疾病基因进行了鉴定。我们通过连锁不平衡的 SNPs 位点的研究, 可以促进我们对人类进化的规律的认识, 也可以针对疾病相关的位点, 帮助找出致病基因[12]。

## 四、在法医学领域 SNPs 的应用

因为单核苷酸多态性是人群广泛性的二等位基因等位标记物, 很容易确定其类型和基因出现的频率, 因此在分析大数据样品方面具有优势。同时 SNPs 位点的突变率很低, 很适合亲子鉴定的工作, 再加上高通量技术可以使用, 因此在法医学领域的 SNPs 受到越来越多人们的关注。另外, 根据 X-SNPs 单倍型分析还可以用来识别祖母和孙女与父母姐妹关系, 识别罪犯的身份[13]。

Y 染色体的 SNPs 的应用在法医学中相当广泛, 由于性别决定是 Y 染色体的功能, 这意味着 Y 遗传就等同于父系遗传, 并向下传递单倍型, 不会发生重组置换等。除了突变的积累以外, 其性状会从父亲稳定的传递给他的儿子。Y 染色体的 SNPs 单倍型分析, 可以进行单亲父亲和儿子的亲子鉴定, 也可鉴定和其他兄弟关系和隔辈的爷孙的关系。Y 型 SNPs 是强有力的取证工具, 可用来识别男性嫌疑人。由 Y 型的 SNPs 多重扩增, 有利于小块痕迹的雄性物质含量(是指两个或更多个血液或体液, 排泄物混合以形成痕迹个体)的鉴定。法医学在检测时可以使用一些与 DNA 检测类似的简化样本, 比如性犯罪男性嫌疑人的 Y 染色体 SNPs 和混合成分(例如精液)可直接通过男性的 Y 染色体 SNP 位点的样本来检测, 测试结果可以具有高度的特异性, 这使得测试工作更容易、更准确得到明确的结果.[14]

线粒体 DNA (mtDNA) 是在存在于细胞核外的唯一的 DNA, 约有 1/3 的具有比核 DNA 高出约 10 倍的突变率。我们可以通过通过序列分析线粒体 DNA 来确定是否是母系遗传。法医线粒体测序主要是由于测序线粒体 DNA 序列可以在样品的 SNPs 和与之相比较的其他样品之间有可直接检测到的序列多态性, 因此具有灵敏度高的优点。特别是对于骨头, 指甲, 毛发等不适合的核 DNA 检测项目是非常有效的[15]。

## 五、多种单核苷酸多态性的研究模式

现阶段关于 SNPs 的研究方法有多种, 以下是几种重要研究模式的简单介绍:

### 1、直接测序模式

直接测序是一种简单精确测序模式。在一般情况下, 在测序峰纯合的 SNPs 呈单峰, 并在显示峰杂合的 SNPs 被设置双峰, 这就比较容易区分开来。直接测序方法检测单核苷酸多态性可达到接近 100% 的检出率。由于近年来越来越少的成本损耗, 同时直接测序和高通量自动测序仪的问世, 使新的 SNPs 位点通过直接测序来发现成为可能。常用的两种策略性直接测序包括 PCR 扩增产物进行测序和随机测序[16]。期间我们发现和验证的 cSNPs PCR 产物测序的方法, 通常使

用的，是通过在编码区不同的个体以及 PCR 扩增产物进行测序和序列分析，倚靠潜在的调节区调节来达到序列扩增的目的。这种研究方法通常用于研究仅已知序列区的 SNPs。由于相对高的成本和复杂的操作，它并不适合的 SNP 的大规模筛选。有两种常用的随机测序方法：基因组比对（genomic alignment）和不完全鸟枪（reduced representative shotgun）。其中的基因组中的比较方法是把重点放在随机测序整个基因组，同时与人类基因组序列的测定序列作比较，这样可以为新发现的 SNPs 在人类的全基因组序列图中明确定位。不完全的鸟枪通过混合相同量的基因组 DNA 样品，其中主要为不同个体的 DNA 样品，通过该部分随机测序覆盖率来达到验证出只有全基因组微片段部分的目的，以构建基因文库中随机全基因组核苷酸序列片段资料库。从而比较其序列比在重叠区域中的不同的染色体的质量之间的差异，成为一组单核苷酸多态性差异评价标准。

## 2、荧光共振能量传递模式

荧光共振能量转移（Fluorescence resonance energy transfer, FRET）是指在足够接近的 2 个荧光发色团（约 1nm 两者的距离），当供体（donor）分子激发到一个较高的电子能量状态后吸收光子的某些频率，在电子返回到基态之前，通过偶极相互作用，实现能量转移到邻近的受体分子（acceptor）的现象（即，共振能量转移发生）。这种现象，称为荧光共振能量转移，简写 FRET。当两种染料的供体 - 受体形式彼此接近，使得都成为 FRET 供体因此不发出荧光，但是当这两个供体分离时又重新发出荧光[17]。

TaqMan 探针技术是基于 FRET 原理的通用方法，在反应体系由基于特异性 SNPs 位点设计的两个常规 PCR 引物和探针。探针的 5'和 3'末端标记有荧光团和淬灭基团，非荧光探针是完整的。探针 3'末端磷酸化，以防止自扩增的引物。PCR 延伸反应使用的 Taqman DNA 聚合酶具有 5'-3'外切酶活性，这样使得探针 5'从探测荧光团被切割下来，从而使荧光团和淬灭剂分离得到的荧光释放信号。在反应中，这样的 SNPs 能够通过检测反应混合物的荧光强度来确定探针和靶序列。因为其紧密度和荧光团在反应后将下降，所以最终影响荧光强度相关的是具有

释放裂解功能的核酸外切酶活性，以及影响两者的组合之间的核苷酸错配。这种方法的优点是，该反应可以通过 PCR 来进行，而不需要分离或洗脱过程，从而进一步降低的 PCR 污染的可能性，扩增产物与所述操作分析具有简单，快速，准确和未经电泳的优点，提高了实验的效率。同样的方法也有一些缺点，因为只有有一个探针设计的基极之间的差异，有假阳性测试结果。

### 3、构象检测模式

单链构象多态性 (Single-strand conformation polymorphism, SSCP) 是现阶段通用的方法，它是一个分离核酸的技术，可以分离长度相同，但序列不同的核酸序列。其检测原理如下：单链 DNA 片段可以形成复杂的空间折叠构象，同时保持分子的三维结构主要是依靠内部碱基配对的相互作用等方式来实现。如果其构象受到影响，从而使发生构象变化的基础也产生变化，因此在聚丙烯酰胺凝胶电泳里不同的构象迁移率的单链 DNA 分子也不同。因此，非变性聚丙烯酰胺凝胶电泳 (PAGE)，可以在分子构象具有区别的差异情况下检出。值得注意的是：(1) SSCP 只是一个突变检测模式，为了最后确定突变类型和位置还需要进一步测序。(2) 由于 SSCP 由三维构象变化，根据上述单链 DNA 分子指向导致的电泳迁移率来区分不同突变造成的，因此，当一个单链 DNA 分子，当在某些位置点突变的三维构象的变化影响不大或没有影响时，或受影响时主要是其他条件，聚丙烯酰胺凝胶电泳不能准确区分由未被发现的位置。(3) SSCP 技术简单，低成本的测试，但近年来因通量不够高已很少使用。温度梯度凝胶电泳 (TGGE) 和 SSCP 原理较为相似，但已很少使用[18]。

DHPLC (变性高效液相色谱, DHPLC) 也是一种快速的、自动化检测方式来检测高通量 SNP 位点。近年来，该技术最早是等于斯坦福大学 Oefner 和 Underhill 在 1995 年的报道。其原理是杂合个体异源双链在部分热变性的条件下，在塔内停留时间与纯合子相比是不同的。DNA 的分离柱是高效液相色谱法的关键，分离柱的化合物结合正电荷，带负电荷的是 DNA 分子，通过将 DHPLC 调节温度到在相近 DNA 分子的  $T_m$  值条件运行。随后通过亲和 DNA 分子大小

的溶出分离柱的能力与电荷成正比，根据亲和力更小更容易分离柱中洗脱出来的原则得到较小的 DNA 分子，反之亦然[19]。DNA 通过 PCR 扩增，由于错配碱基导致异源双链的情况下，单链 DNA 比双链 DNA 携带更少电荷，它会先被洗掉，正常配对双链最终从中分离出来，这样我们可以通过分析色谱峰的类型和数量来确定 SNPs 的存在或不存在。这种方式测量未知的 SNP 具有明显的优势，因为它的自动化程度高，灵敏度高，更大的 DNA 片段的分析也适用。但它只能检测样品的 SNP 的突变型存在或不存在，并且不能在新的突变或多态性的筛选领域来确定突变位置从而得到准确的结果。

#### 4、杂交的模式

基因芯片 (Gene chip)，也被称为 DNA 芯片，生物芯片，是一种高通量的 SNPs 检测方法，通过固相-支持培养基来检测原位杂交和检测荧光。根据互补核苷酸的碱基配对 (A 和 T, C 和 G 配对) 原理设计多个探针，通过优化操作，探针仅与其完全互补的，并且不含有单个碱基错配的序列进行杂交。在由诸如玻璃，硅等特殊载体表面，然后制作成的具有 DNA 特定序列探针芯片的基因芯片。此方法通过荧光化学提取和扩增，杂交固定化的探针标记来检测之后需要被检测的基因。通过洗涤没有发生杂交反应来得到杂交样品。由于目标基因和探针杂交的荧光强度和荧光种类的程度相关，因此，荧光强度或荧光种类的程度可用来检测核苷酸序列的类型[20]。

动态等位基因特异性杂交(DASH)是利用热力学检测在 DNA 变性中的应用。其原理是，检测所含有的 SNPs 使用生物素标记的引物之一，扩增产物是一些固定在抗生蛋白链菌素 - 生物素包被的微量管中扩增的 DNA 样品，无标记产物用 NaOH 则可以洗脱。单链产物和经过特异性寡核苷酸探针标记的等位基因杂交以释放荧光，然后继续加热样品并且实时检测荧光信号。由于双链体具有错配的热稳定性在与没有错配的情况下相比来说较差，所以可通过  $T_m$  值时检测判断[21]。

#### 5、质谱模式



质量检测图谱 (MS) 分析是根据离子移动的方法来进行测定的方法, 主要是检测离子经过充电的电场和磁场分离后比率的方法 (比率主要为携带的带电粒子的质量和电荷的比)。

基质辅助激光解吸电离 - 飞行时间质谱 (MALDI-TOF MS)。该方法使用共结晶薄膜, 激光的能量从基质吸收并从基质形成, 基质样品得到一定强度的激光照射, 使得分子离子化样品中的电场电离。样品之间发生电荷转移并通过飞行加速管道, 根据检测样品的飞行到达时间 (即测量离子的质荷比和离子的飞行时间) 来达到目的。其结果正比于检测到的离子大小。越小离子质量越高, 经过负电荷区时, 到达检测器的时间更快[22]。通过这种方法, 可以只分离包含的两个不同的碱基序列区域的基因中的一个。MS 不需要检测荧光标记物, 其可以是样品的 PCR 产物, 引物延伸反应产物, 等位基因特异性终止产物, 入侵者消化物等等, 我们可以实现按质量直接检测不同质量样品或探针的办法, 然后根据推导出的 SNPs 样品来找到已知 SNPs 或者未知的 SNPs 的基因分型。

## 6、SNPs 的高通量研究

全基因组关联分析 (GWAS) 是一种基于人类基因组计划和人类基因组单体型图谱计划的研究工具。为大群体 DNA 样本、高密度的全基因组遗传标记 (如 SNPs 位点) 的方法来进行基因分型, 以便于找到与复杂疾病遗传因素相关联的方法。这种方法的进一步研究方式主要是快速基因型分析。

从候选基因关联研究与 GWAS 研究的发展, 从根本上是一致的。但它相对于候选基因关联研究是极具成本效益的, 因为其能够评估整个基因组的遗传资料, 同时并不需要在实验前做出假设, 也不需要相关基因的背景资料[23]。

目前 GWAS 想法基本相同: 首先通过 GWAS, 找到一些疾病相关性状, 例如以第一组的人来说疾病相关的 SNPs, 这些位点然后再于第二和第三组的人口中分析类似 SNPs 位点, 以证明与 SNPs 位点性状相关的疾病。第一组人的分析, 几乎所有都基于全基因组芯片技术来进行高密度 SNP 分析, 人群的第二和第三组, 可以使用各种不同的 SNPs 分析方法, 如 TaqMan 探针, 质谱, 焦磷酸测序, 低

密度的 SNP 芯片定制方法, Sanger 测序方法[24]。

## 六、引物延伸

单碱基引物延伸 (SBE), 也被称为微测序 (Minisequencing) 或模板介导的荧光染料终止子掺入反应 (TDI)。首先包含 SNP 的扩增和 DNA 序列的微测序。通过设计一条 3'端碱基与 SNP 位点只差一个碱基的引物, 加入 DNA 聚合酶和带有不同荧光标记的 ddNTP 进行延伸反应, 当加入的 ddNTP 和 SNP 位点的碱基互补时, 延伸反应才可能进行, 最后通过检测延伸碱基发出的荧光信号以判断 SNP 类型。

焦磷酸测序技术是一个测量焦基于生物发光测定法,对于已知短序列分析,可重复性和精确度性能与 SangerDNA 方法相类比。这种技术的原理是, 当在 DNA 聚合酶 (DNA 聚合酶) 退火的引物和模板 DNA, 三磷酸腺苷酸酶 (ATP 硫酸), 萤光素酶 (萤光素酶) 和腺苷三磷酸双磷酸酶 (Apyrase) 4 种酶共同作用, 接着加入四种不同的脱氧核糖核苷酸 (dNTPs)。如果模板与新加入碱基是相对应的, 等摩尔量释放的 PPi 则可以在延伸反应中出现。PPi 由 ATP sulfurytase 催化, 结合腺苷 5'-磷酸硫酸生成三磷酸腺苷 (ATP)。ATP 的产生萤光素酶进一步激活发射荧光, 荧光强度与结合的碱基的数目是正相关的。相反, 如果添加的碱基不是与模板互补, 它们不能产生荧光, 这样就以便检测其他引物的核苷酸序列[26]。

## 七、药物基因组学分析

药物基因组学 (Pharmacogenomics), 也称为基因组药理学, 属于基因组学和药理学其中一个分支, SNPs 多态型标志物可以通过检测遗传根源来揭示在不同个体对不同药物的人群敏感性的差异。药物基因组学主要集中在药物吸收, 转运代谢的遗传特征和清除药物以及靶标分子基因突变的个体差异和药物 (包括疗效和不良反应) 之间的关系。药物功效和对身体的不良反应是由多种因素影响, 主要是包括药物代谢酶, 转运体, 受体, 其它药物靶标蛋白和不同个体的遗传多态性之间蛋白质编码基因的影响, 这是多态性 SNPs 分析的基本形式。

在基于大量的临床研究中，基因多态性是导致在体内的个体差异的主要因素。其影响会出现在药效和副作用方面。通过药物基因组学研究 SNPs 的领域，主要是用于明确在不同的药代动力学和药效学方面的个体差异的遗传基础，以及寻求开发有针对性的新的基因治疗方案，此外，我们可以根据个体的遗传背景，反向调整药物治疗方案和达到“个体化治疗”的目的。

## 八、人类进化历史的分析

SNP 可用于研究不同种群间的 DNA 序列变异，因为变化水平的演变是可以使用 SNP 反映的。大部分 SNPs 位于基因组中的非蛋白编码区域，其分布不受选择压力的影响，能够长期得以保留。对位于蛋白编码区域并且影响表现型的变异可能受到自然选择的压力。如果变异能够在长时间内被保留下来，那么该变异很可能对个体后代繁衍有重要影响。经过长期自然选择最终能够得以保存的变异可能代表着进化史上的重要事件。因此通过计算不同物种间蛋白质家族中编码区、非编码区与变异数目的比值，可以追溯进化树上的分支点。在该分支点处，那些对物种进化有正面影响的变异才能被固定并保留在基因池 (gene pool) (所有个体的集合，包含所有的基因相同的人群)

## 九、前列腺癌与 SNPs 之间的关系

前列腺癌是指发生在前列腺的上皮性恶性肿瘤。2004 年 WHO《泌尿系统及男性生殖器官肿瘤病理学和遗传学》中前列腺癌病理类型上包括腺癌 (腺泡腺癌)、导管腺癌、尿路上皮癌、鳞状细胞癌、腺鳞癌。其中前列腺腺癌占 95%以上，因此，通常我们所说的前列腺癌就是指前列腺腺癌[27]。2012 年我国肿瘤登记地区前列腺癌发病率为 9.92/10 万，列男性恶性肿瘤发病率的第 6 位。发病年龄在 55 岁前处于较低水平，55 岁后逐渐升高，发病率随着年龄的增长而增长，高峰年龄是 70~80 岁。家族遗传型前列腺癌患者发病年龄稍早，年龄≤55 岁的患者占 43%[28]。在世界范围内，前列腺癌是影响男性健康的最常见恶性肿瘤之一，发病率在所有男性恶性肿瘤中位居第二，同时有着明显的地理和种族差异。澳大利亚/新西兰、加勒比海的纳维亚地区最高，亚洲及北非相对较低[29]。根据

美国癌症协会发布的报告，美国 2013 年前列腺癌患者数目达到 238950 人，占男性所有恶性肿瘤的 28%，死亡人数为 29720 人；在欧洲，2008 年新发前列腺癌患者数为 382000 人，90000 名患者死亡。虽然亚洲的发病率低于欧美国家，但是其近年来的增长速率远远超过欧美国家的增长速率。根据 2012 年世界普查报告，也可以看出前列腺癌在中国的发病率有了明显的增长：1998-19974 年期间发病率的增长率为 2.1%，但是在 1994-2002，前列腺癌发病率增长率高达 13.4%[30]。

引起前列腺癌的危险因素有待研究，一些已经可以确定的危险因素包括年龄、种族和遗传性。2 个或者 2 个以上的一级亲属患病，患病危险性会增加 5-11 倍[31]。通过全基因组相关研究，现阶段已经发现 50 余个单核苷酸多态性 (SNP) 与前列腺癌发病风险相关。有报告证实，发生 5 个或者 5 个以上的 SNP 的男性与无 SNP 的男性相比，罹患前列腺的风险比高达 9.46[32]。

日本的冈山大学科研小组经过与癌症相关的基因中的 48 个非同义 SNPs 的筛选，指出在 12 个 SNPs 中相关的 10 个基因对前列腺癌这种恶性肿瘤的发生与发展具有非常重要的意义。在这 10 个基因里，包含了 1 个代谢酶基因、1 个凋亡调节基因、2 个 DNA 修复基因、5 个肿瘤抑制基因和 1 个染色体分离相关基因。在这里面有 8 个 SNPs 与恶性肿瘤的发生与发展的关系是首次被发现。在预测今后 30 年里面的前列腺癌罹患风险方面，风险组最高的可达 29%，而风险组比较低的可为 0.6%，其中风险组最低的仅仅约为 0.2%。同时我们在预先评估前列腺癌罹患风险方面，发现这 12 个 SNPs 影响效应可以发生累积，风险组高与低的罹患率相比较，其 OR 值高达 47.4。所以，在肿瘤的患病风险遗传学的评估中应用上与多种恶性肿瘤有关联性的 SNPs 的整合，与现有方式相比，对于罹患前列腺癌风险较高的人群的普筛是具有更大的准确性和更高的效率，并在肿瘤的早期诊断和预防措施的选择方面具有优势。

日本人群所在的地理位置与中国人群相近，同时两国人群也具有相似的人种起源，因此两国的遗传背景有很高的共同性。经过在日本大规模筛查得到的

前列腺癌罹患风险相关的 SNPs 可能在中国人群中也会有部分适用。因此，在亚洲的四个国家，中、日、韩和新加坡联合进行了一项跨国际的多中心研究，目的是为了检验这一假说的准确与否，并将在日本的筛查结果推广到亚洲其他的国家，同时检验筛选出的前列腺癌高风险相关性 SNPs 在亚洲其他人群中的效力。本次试验中所选位点是来自日本冈山大学科研小组的筛查成果、一致于此次多中心试验的研究内容。

## 方法部分

### 一、仪器和试剂

#### 1、主要实验仪器

BIQRAD电泳仪	美国BIQRAD公司
Gilson移液器	法国Gilson公司
超净台	上海浦东物理光学仪器厂
DYY-III286型电泳槽	北京六一仪器厂
NanoDrop 2000 超微量分光光度计	美国Thermo Scientific公司
- 80℃超低温冰箱	日本SANYO公司
凝胶成像系统	珠海黑马医学仪器有限公司
扩增仪	ABI公司, 美国
漩涡振荡器	浙江乐清仪器厂
MICROCL 21R台式高速离心机	德国Thermo公司
高压消毒锅	日本SANYO公司
电热恒温鼓风干燥箱	上海一恒科技有限公司
PRISM 3730 DNA分析仪	美国ABI公司
Genemapper 4.1软件	美国应用生物系统公司

#### 2、主要实验试剂

四乙酸乙二胺	国产, 分析纯
脱氧核糖三磷酸 2.5 mM	TianGen, 北京
Hi-Di™ formamide	美国应用生物系统公司
核酸标记物I	TianGen 北京
多重SNPs DNA提取试剂盒	美国应用生物系统公司
Phosphatase alkaline shrimp(SAP)	美国普洛麦格 (Promega)
琼脂糖	美国英骏 (Invitrogen)
hotmaster Taq 2.5U/ 1	TianGen北京

---

EDTA基因分析缓冲液	美国应用生物系统公司
标准GeneScan™-120 LIZ Size	美国应用生物系统公司
标准基质液DS-02	美国应用生物系统公司
SDSPAGE电极缓冲液	美国应用生物系统公司
POP™-4	美国应用生物系统公司

---

## 二、受测试人员的入组

自2009年1月至2013年4月,在北京首钢医院吴阶平泌尿外科中心、浙江大学第一附属医院及南方医科大学附属珠江医院招募253例前列腺癌(Pca组)患者和214例非前列腺癌及其他健康人群(对照组)

Pca组人员招募:病例组为术后病理检查或者通过影像学、前列腺穿刺活检证实前列腺癌患者,初次发病或复诊均可。记录患者的临床资料如发病年龄、Gleason评分、TNM分期、PSA水、治疗的方法选择及治疗后疗效等等,同时发给患者问卷进行调查,其中的内容包括有饮酒吸烟史,家族史及饮食风俗等。在确定入组后,床位主管医生与患者进行沟通,告知患者抽取血液作为样本步骤,血液样本的去向、处置,此次研究的主要目标等。获得患者同意之后,记录保存患者的临床信息并且与患者签署知情同意书。临床资料记录表及调查表里面的具体内容见后文附件(附件1)。

对照组人员招募:对照组为非前列腺癌及其他健康人群,主要条件是需与Pca组人员在年龄上相一致,同时也是在同一时段同一医疗单位住院或者门诊就诊的其他患者。确定入组之后,主管医生同样告知患者抽取血液作为样本步骤,血液样本的去向、处置,此次研究的主要目标等。获得患者同意之后,记录保存患者的临床信息并且与患者签署知情同意书,发给患者问卷进行调查,其中的内容包括有饮酒吸烟史,家族史及饮食风俗等并记录。知情同意书的具体内容可见后文附件(附件2)。

## 三、血液样本的抽取保存

在获得患者同意并且签署知情同意书之后,床位主管医生联系主要相关责

任护士，患者入院常规检查项目如血常规、生化验血用血液抽取后，需额外再抽取 2ml 外周静脉血，以便进行后续的 DNA 提取、分析等等研究。使用 EDTA 抗凝管将抽取到的大概 2ml 外周静脉血保存，同时贴上患者信息条形码、记录患者姓名、住院号等等信息。在 6 小时以内保存在在 -20℃（或 -80℃）条件下。为了防止收集到的血液样本在外界的温度变化下所导致的反复冻融而造成的 DNA 变化影响最后试验结果，操作过程因此需严格按照实验室相关规定，血液样的的 -80℃ 冰箱保存与提取需定期由实验室人员操作，平常运输、使用过程中注意干冰的使用，主要是为了防止温度过高。

#### 四、DNA 提取定量及定纯

此次试验主要通过由北京 TianGen 生物化学科技有限公司的全血 DNA 抽提试剂盒(TIANamp Blood DNA Kit)来进行 DNA 的提取。产品信息如下: TIANamp Blood DNA Kit, 北京 TianGen 公司, 其号码为 DP318-02, 产品为离心柱型。

##### 1、主要组成成分:

试剂	规格
Proteinase K	1ml
GD Buffer	13ml
GB Buffer	15ml
GS Buffer	15ml
漂洗液 PW	15ml
TB	15ml
CL Solution	60ml
CB3 absorbing-c	50 个
2ml 收集管	50 个
1.5 ml 离心管	50 个

##### 2、试验所涉及设备及仪器

水浴箱、高速离心机、Gilson 移液枪，低温冰箱（-20℃和 -80℃）



3、提取步骤如下文，主要参考操作说明书以及依据实际情况操作：

- 1) 取 1.5ml 离心管，加入 1000 $\mu$ L CL 细胞裂解液、400 $\mu$ L 血液样品，56 $^{\circ}$ C 水浴，颠倒震荡混合均匀后，高速离心机离心 2min，12,000 rpm。得到样本后，将上清去除，沉淀的物质即为细胞核（此操作可重复，以得到更纯的细胞核沉淀物质）。
- 2) 200 $\mu$ L GS Buffer 加入到经过离心所得到的细胞核沉淀中，室温震荡混合均匀。
- 3) 200 $\mu$ L GB Buffer、20 $\mu$ L 蛋白酶 K 加入混合物中，室温震荡混合均匀，56 $^{\circ}$ C 水浴静置十分钟，期间可将溶液颠倒震荡，可得清亮溶液。假如所得溶液仍有浑浊，可以适当再延长裂解时间。
- 4) 200  $\mu$ L 无水酒精加入所得溶液中，室温震荡混合均匀后所得溶液可能出现飘絮样沉淀。
- 5) 取出一个 CB3 吸附柱，放置入 2ml 收集管中，将所得的飘絮样沉淀及溶液加入至收集管中，高速离心机离心 30s，12, 000rpm，之后保留收集管中的吸附柱，将收集管里面的废液倒出。
- 6) 500 $\mu$ LGD Buffer 加入吸附柱中，高速离心机离心 30 s，12,000rpm，之后保留收集管中的吸附柱，将收集管里的废液倒出。
- 7) 700  $\mu$ L 漂洗液 PW 加入吸附柱中，高速离心机离心 30 s，12,000rpm，之后保留收集管中的吸附柱，将收集管里面的废液倒出。
- 8) 500  $\mu$ L 漂洗液 PW 加入吸附柱中，高速离心机离心 30s，12,000 rpm，之后保留收集管中的吸附柱，将收集管里面的废液倒出。
- 9) 高速离心机离心 2 min，12,000 rpm，将收集管里的废液倒出。吸附柱盖子打开后，室温静置 5 min。
- 10) 取出一个 1.5ml 离心管，在开盖吸附柱吸附膜的中央位置隔空加入 TB 洗脱缓冲液 100 $\mu$ L，室温静置，2 min 后高速离心机离心 2 min，12,000rpm，保留离心管中所得到的的洗脱液。
- 11) 上述离心管中的洗脱液再次加入到吸附柱中，室温静置，2min 后高速离心

机离心 2min, 12,000rpm。

12) 所得到的 DNA 沉淀物 - 80℃条件下保存。

#### 4、所得 DNA 定量及定纯

所得的 DNA 的纯度和浓度主要通过核酸蛋白定量仪器来测定，其中主要用 Concentration（浓度）和 OD 值（吸光度）来作为定量单位。

#### 5、DNA 的琼脂糖凝胶电泳验证

1) 琼脂糖凝胶的配置：20ml 0.5×Tris 硼酸加入定量琼脂糖凝胶后加热，酒精灯加热琼脂糖融化完全后冷却，在温度降至 50~60℃时适量溴化乙锭加入其中，使其最终浓度为 0.5μg/ml。小心得到溶液倒入胶版之中，凝固后将梳子拔出，得到 0.8%琼脂糖凝胶。

2) 样品添加：2 μL 上样缓冲液加入至 10μL DNA 样品溶液中，混合均匀后加入样品槽；

3) 电泳验证：电压：140V；电泳时间：8 min；

4) 拍照保留所得结果。

### 五、SNP 位点的选择

日本的冈山大学科研小组经过与癌症相关的基因中的 48 个非同义 SNPs 的筛选，指出在 12 个 SNPs 中相关的 10 个基因对前列腺癌这种恶性肿瘤的发生与发展具有非常重要的意义。在这 10 个基因里，包含了 1 个代谢酶基因、1 个凋亡调节基因、2 个 DNA 修复基因、5 个肿瘤抑制基因和 1 个染色体分离相关基因。在这里面有 9 个 SNPs 与恶性肿瘤的发生与发展的关系是首次被发现。在预测今后 30 年里面的前列腺癌罹患风险方面，风险组最高的可达 29%，而风险组比较低的可为 0.6%，其中风险组最低的仅仅约为 0.2%。同时我们在预先评估前列腺癌罹患风险方面，发现这 12 个 SNPs 影响效应可以发生累积，风险组高与低的罹患率相比较，其 OR 值高达 47.4。所以，在肿瘤的患病风险遗传学的评估中应用上与多种恶性肿瘤有关联性的 SNPs 的整合，与现有方式相比，对于罹患前列腺癌风险较高的人群的普筛是具有更大的准确性和更高的效率，并在肿瘤的

早期诊断和预防措施的选择方面具有优势。

在此次的中国人群前列腺癌罹患风险因素相关性研究中，我们从 48 个非同义 SNPs 位点中选取了 8 个。其主要原则为：

- 1) 此次 8 个位点已经在日本人群验证其与前列腺癌罹患风险相关性 SNPs。
- 2) 最小等位突变频率 (MAF) 大于 5%;
- 3) 此次所选的基因位点都与 DNA 错配修复基因、抑癌基因等肿瘤发生发展相关基因
- 4) 此次我们所选定的 48 个位点均为非同义 SNPs。

此次我们所选位点主要是来源于日本岗山大学清水教授课题组，鉴于数据仍处于保密状态，故本论文所涉及的其他 SNPs 位点及引物信息省去，我们使用其所属位点来作为命名方式，其中本文研究位点为 EXO 1 称为 rs9350，其余七个位点分别为：RAD17 SNP、SMARCAD1 SNP、MMP27 SNP、PSMD8BP1 SNP、CASP9 SNP、DCLRE1B SNP、BARD1 SNP。

## 六、多重 PCR (SNsPshot) 分型基因

此次试验，我们主要应用多重 PCR 的方法为该基因进行分型，同时此次试验在北京、广州、浙江、日本所用方法相同，其主要涉及到两次的 PCR 步骤，分别为 SNsPshot (多重 PCR) 和二次 PCR (单个碱基延伸)，主要的引物设计方法见下文：

### SNsPshot 引物

PCR 引物设计遵循一般原则，比如错误发生率低，避免引物及引物之间、自身与引物之间发生二次结构形成，随机分布的碱基等等，满足上述条件，其自身设计原则主要是：

- 1) length 引物：18 ~ 25 bp;
- 2) length 产物：50 ~ 100 bp;
- 3) 退火 T：55~65℃;
- 4) CG 所占百分比：40 ~ 60%;

## 1、二次 PCR 引物

- 1) 在试验 SNPs 位置的 3'上游大约 1bp 的地方插入引物,有效引物(模板可与引物完美配对)的长度大约为  $24\pm 6\text{bp}$ , 退火 T 可在  $47^{\circ}\text{C}\sim 53^{\circ}\text{C}$ 。
- 2) dT (poly) 加入有效引物的 5'末端以及相同反应管中的引物长度可有 4-6bp 的差别用来检测相同位点, 其中主要目的是为了完成引物长度的控制, 完成不同位点的区分。
- 3) 正反双链都可以用来进行引物设计, 倘若其中一条无法设计出最理想的引物, 则可以使用另外一条进行设计。

## 2、引物序列的特异性验证

在对于已经设计出来的引物序列特异性以及其普通属性的比较和验证主要是通过 oligo6.0 软件来实现。其中为了验证引物的特异性及稳定性, 我们可以登陆 <http://blast.ncbi.nlm.nih.gov/Blast.cgi> 或 Human BLAT Search (UCSC Genome Bioinformatics Site) 进行比较。

## 七、首次 PCR (SNsPshot)

### 1、PCR primer mix

反应体系见下表所示:

试剂	用量
dNTPs(10mM)	0.5 $\mu\text{L}$
氯化镁	0.8 $\mu\text{L}$
10 $\times$ PCR 缓冲液	2.5 $\mu\text{L}$
Platinum Taq(5U)	0.2 $\mu\text{L}$
扩增模板	1 $\mu\text{L}$
混合引物	1 $\mu\text{L}$
ddH <sub>2</sub> O	19 $\mu\text{L}$
Total	25 $\mu\text{L}$

反应体系总用量为 25 $\mu\text{L}$ , 基因组 DNA10~50ng 包含在 DNA template 中, PCR

循环条件为：95℃ 5 min→ (95℃30 sec → 56℃30 sec → 72℃30 sec) × 32 cycle→72℃ 5 min→ 4℃。主要通过 0.2ml 毛细微管在 ABI 9700 Thermocycler 分析仪上完成。

2、同 DNA 提取物验证，PCR 所得产物主要通过琼脂糖凝胶电泳来验证。

3、PCR 所得产物提纯。

产物提纯主要目的是为了去除多余的反应物，如引物、dNTPs 等等，从而得到更纯的扩增反应产物。主要使用虾碱酶方法，反应体系见下表：

试剂	用量(μL)
蒸馏水	7.5μL
ExoI (20U/μL)	0.2μL
PCR 产物	2μL
SAP (1U/μL)	0.3μL
Total	10μL

室温下混合均匀，保持温度 37℃90 min，EXO 1 及 SAP 的灭活主要是在 75℃条件下保持 15 min，在 4℃条件下可保持 24h。

## 八、二次 PCR（单个碱基延伸）

二次 PCR 主要在冰面上进行操作。冰箱取出试验所需要试剂，室温融化，震荡后离心。室温下混合均匀。

### 1、二次 PCR 设计引物以及配备模板

在进行二次 PCR（单个碱基延伸）之前，采用虾碱酶的方法先对 PCR 产物进行纯化。PCR 产物浓度一般为 0.01-0.04pmol 来作为单个碱基延伸试验的 DNA 模板。模板纯化后，可在-80℃冰箱长期保存，也可在-20℃冰箱保存 24h。

设计引物时需注意的事项：

1) 不同位点的引物在相同的反应体系中，设计的长度不能相同，其中 4-6bp 的长度差异为最佳。

2) 可采用 HPLC 方法纯化合成后碱基长度大于 30bp 的 DNA 引物;

3); dT (poly) 加入有效引物的 5'末端以及相同反应管中的引物长度可有 4-6bp 的差别用来检测相同位点, 其中主要目的是为了完成引物长度的控制, 完成不同位点的区分。

4) 有效引物 (与模板完美配对部分) 至少需要 50°C 的 T<sub>m</sub> 值。

5) 正反双链都可以用来进行引物设计, 倘若其中一条无法设计出最理想的引物, 则可以使用另外一条进行设计。

EXO 1 SNPs所指的位置为位于EXO1基因上的位点, 我们将其命名为rs9350。此次两次PCR试验共需要三条引物: 其中首次扩增需要两条引物, 分别如下: 上游引物为: 5'-CAGAATGGTCTTAAATGGGTGT-3' 下游引物为: 5'-TTCAGAATAAGAAACAAGGCAAC-3'。而用于二次PCR所需要的产物延伸引物则为:

5'-TTTTTTTTTTTTTTTTTTAGACTCTCTTTCTACAACCAAGATCAAAC-3'.

## 2、二次 PCR (SNaPshot PCR 延伸)

DNA 模板混合: 将纯化完成后 PCR 产物各种分别取 22 μL 混合。

延伸引物混合: 引物混合后确保其在反应体系中保持 0.2 μM 的终浓度。

二次 PCR 延伸反应体系见下表所示:

试剂	用量 μL / Sample
Reaction Mix	1.5μL / Sample
PCR 产物	1μL / Sample
Probe Mix	0.5μL / Sample
总体积	3 μL

二次循环, 也被成为 SBE 扩增 (即各 SNPs 的但碱基扩增), 在此每个循环的主要步骤有:

- 1) 变性, 打开模板双链;
- 2) 退火, 模板连打开后与延伸引物相配对;
- 3) 延伸, 在引物 3'端、目的 SNP 位点处延伸一个 ddNTP;
- 4) 大概有约 25 个 cycle。

根据多重 PCR 试剂盒说明书，实验步骤循环主要如下所示：

95°C 5 min → (95°C 30 sec → 56°C 30 sec → 72°C 30 sec) × 30 cycle → 4°C。

3、清除未参加反应的 ddNTPs，将延伸后的 DNA 产物进行纯化。反应体系如下表所示：

试剂	用量
SAP (1U/μL)	0.3μL
蒸馏水	1.7μL
延伸产物	3μL
Total	5μL

将上述试剂震荡混匀，37°C保温 1h，75°C保温 15min 灭活 SAP 酶，4°C保存 24h。

## 九、检测结果的判读分析

1、在此次实验中，我们主要使用了四种脱氧核苷酸进行标记，bp 与荧光的对应关系如下图所示：

Bp 与荧光的对应惯性

脱氧核苷酸	Dye 标签	颜色
T	dROX	Red
G	dR110	Blue
C	dTAMRA	Black
A	dR6G	Green

2、纯合子的判读主要表现为单峰，杂合子的判读主要表现为双峰。

## 十、统计学方法

计算人口学特征时，病例组和对照组之间的年龄、PSA 水平的比较采用两独立样本 t 检验。符合正态分布的数据采用平均值±标准差 (Standard Deviation, SD) 形式统计，不符合正态分布的数据采用中位数形式统计。使用 Pearson's  $\chi^2$  检验方法用来比较各临床指标之间的差异 (患病组与对照组之间、多种基因型

之间)。各个对照组的 SNP 位点的基因型分布频率使用 Pearson's  $X^2$  检验方法，主要是进行 Hardy-Weinberg 平衡检验同时通过 R 语言来分析。计算 Gleason 评分和不同基因型关系（或等位基因）时采用的是 Mann-Whitney U 检验。logistic 回归分析主要用于计算特定的基因型与本次研究所涉及人群的前列腺癌罹患风险比值比(ORs)、与之相对应的 P 值以及其 95%的可信区间(Confidence intervals, CIs)，来分析各种基因型和前列腺癌罹患相关与否。前列腺癌罹患风险具有相关性的 SNPs 与肿瘤危险分期和年龄等因素以分层的方式进行分析。在肿瘤危险分期这一方面，我们将局限型前列腺癌和进展型前列腺癌作为不同分层。局限型的前列腺癌的分层标准设定为：TNM 分期 T1-2; N0/NX; M0/MX; Gleason 评分 2-7 分; PSA $\leq$ 50 ng/ml。进展型前列腺癌组的分层标准则设定为：TNM 分期 T3/4; N+; M+; Gleason 评分 8-10 分; PSA $>$ 50 ng/ml。局限型入组需满足全部标准，进展性则只需要满足标准被其中之一。在以年龄分层这一方面，我们主要通过其中前列腺组以初发年龄、对照组以入组年龄为标准的方式，以 72 岁年龄为分界线，最后分成 $\leq$ 72 岁组以及 $>$ 72 岁组。将患者吸烟饮酒等情况作为前列腺癌罹患肿瘤相关性的多因素回归模型里面的影响因素。此次我们使用 SPSS 20.0 统计学软件，在  $P < 0.05$  时可认定为差异有统计学意义，所有统计检验为双侧检验。



## 试验结果

### 一、样本的采集以及 SNPs 试验情况

此次试验前列腺癌患者血液样本 253 例,对照组血液样本 214 例。其余 SNPs 检测失败可能由于:

1. 样本收集、保存过程出现错误操作(例如血液样本暴露时间过长)或者 DNA 提取是操作失误;
2. 由于 PCR 产物不纯,软件、仪器原因,SNPs 分型读数差,出现低峰值;
3. 多对引物扩增是出现相互影响的情况;

### 二、样本其余临床指标统计

253 例前列腺癌患者与其余 214 例对照组临床指标统计分析,结果见表 1。除平均 PSA 结果统计后具有统计学意义外( $p < 0.05$  符合对于前列腺癌病理分析),前列腺癌组与对照组在抽烟、喝酒、平均年龄等临床指标均无统计学意义,证明样本在其余指标情况吻合,符合此次试验要求。

表 1 受试者的临床指标

Table 1 Demographic and clinical characteristics of study subjects

Variables	Control	Patients	Localized Pca	Advanced Pca	p value
No. of subjects	214	253	91	162	
Mean age(SD)	70.2(8.0)	71.4(7.2)	72.0(6.6)	71.1(7.5)	0.174
Mean PSA(SD)	7(5.5)	60.4(64.1)	13.7(8.2)	74.3(64.4)	< 0.01
Smoking status no.(%)					
Never	66(30.8)	74(29.2)	26(28.6)	48(29.6)	0.708
Ever	148(69.2)	179(70.8)	65(71.4)	114(70.4)	
Drinking status no.(%)					0.975
Never	40(18.7)	47(18.6)	16(17.6)	31(19.1)	
Ever	174(81.3)	206(81.4)	75(82.4)	131(80.9)	

### 试验结果

<b>Gleason score</b>			
No. of subjects(%)	212(100)	84(100)	158(100)
2-6	52(24.5)	42(50.0)	10(7.8)
7	68(32.1)	42(50.0)	26(26.3)
8-10	92(43.4)	0(0)	92(71.9)
<b>Tumor stage</b>			
No.(%)	241(100)	83(100)	158(100)
T1	33(13.7)	25(30.1)	8(5.1)
T2	101(41.9)	58(69.9)	43(27.2)
T3	75(31.1)	0(0)	75(47.5)
T4	32(13.3)	0(0)	20.2
<b>Nodal stage</b>			
No.(%)	221(100)	82(100)	139(100)
N0	163(73.8)	82(100)	81(58.3)
N1	58(26.2)	0(0)	58(41.7)
<b>Metastasis stage</b>			
No(%)	240(100)	83(100)	157(100)
M0	165(68.8)	83(100)	82(52.2)
M1	75(31.2)	0(0)	75(47.8)

### 三、rs9350 的基因频率的比较

在中国不同地区的人群中，rs9350 基因频率的区别并不具备统计学意义，如表 2.基因分布符合 Hardy-Weinberg 原则 ( $p$  大于 0.05)，可认为该对照组具有一定的代表性。

表 2 rs9350 基因频率分布

Area	Comparison of genotype frequencies of rs9350							
	Control				Cases			
	Guangzhou	Hangzhou	Beijing	p value	Guangzhou	Huangzhou	Beijing	p value
EXO1 rs9350	n=100	n=20	n=94	0.105	n=103	n=58	n=92	0.309
CC	41(41.0)	7(35.0)	49(52.1)		38(37.0)	18(31.0)	27(29.3)	
CT	52(52.0)	9(45.0)	34(36.2)		56(54.3)	29(50.0)	49(53.3)	
TT	7(7.0)	4(20.0)	11(11.7)		9(8.7)	11(19.0)	16(17.4)	

#### 四、EXO1 (rs9350) 相关 SNPs 分析

分析结果见表 3.rs9350 CC 等位基因片段在前列腺癌组与对照组中有相同百分比 (32.8%与 45.3%), rs9350 CT 和 TT 基因型在前列腺癌组与对照组中的百分比分别为 53.0%与 44.4% (CT)、14.2%与 10.3% (TT)。与 CC 基因表型相比, CT 基因表型有更高的前列腺癌罹患风险 (OR=1.648,95%CI=1.112-2.444, p=0.013; 校正 OR-1.678,95%CI=1.176-2.500, p=0.05), C 与 T 基因表型之间具有统计学意义 (p=0.09)

试验结果

表 3rs9350 相关 SNPs 分析

Table3 Association of rs9350 prostate cancer risk

SNP	Gene	Genotype or allele	Controls(%)	Cases(%)	Crude OR (95% CI)	p value	Adjusted ORa (95% CI)	p value
			n=214	n=253				
rs9350	EXO1	CC	97(45.3)	83(32.8)	1.00(ref.)		1.00 (ref.)	
		CT	95(44.4)	134(53.0)	1.648 (1.112-2.444)	0.013	1.678 (1.130-2.494)	0.01
		TT	22(10.3)	36(14.2)	1.912 (1.043-3.505)	0.036	1.955 (1.062-3.598)	0.031
		CT/TT	117(54.7)	170(67.2)	1.648 (1.166-2.473)	0.006	1.714 (1.176-2.500)	0.005
		C	289(67.5)	300(59.3)	1.00 (ref.)		1.00 (ref.)	
		TT	139(32.5)	206(40.7)	1.428 (1.091-1.868)	0.009	1.435 (1.096-1.879)	0.009

五、不同年龄层的 SNPs(rs9350)分析

分析结果见表 4.在已有的数据分析中,我们将 72 岁作为年龄分界线(前列腺癌患者组及对照组中位年龄分别为 71.4 和 70.2),在此年龄段上线具有近似的病例数。在大于 72 岁的年龄组中,rs9350 CT/TT 基因型具有统计学意义(p=0.032);携带 C 等位基因具有更高的前列腺癌罹患风险(校正 OR=1.464,95%CI=0.998-12.147, p=0.041)。

表 4 不同年龄层 SNPs 分析

Table 4 rs9350 genotype and allele frequencies [n (%)] and adjusted OR stratified by age at diagnosis

Genotype or allele	Age at diagnosis $\leq 72$ years				Age at diagnosis $> 72$ years			
	controls n=108	Cases n=108	Adjusted OR (95% CI)	p value	controls n=106	cases n=131	Adjusted OR(95% CI)	p value
CC	45(41.7)	37(30.3)	1.00(ref.)		52(49.0)	46(35.1)	1.00(ref.)	
CT	52(48.1)	67(54.9)	1.537(0.869-2.720)	0.139	43(40.6)	67(51.2)	1.758(1.008-3.064)	0.047
TT	11(10.2)	18(14.8)	2.043(0.848-4.926)	0.112	11(10.4)	18(13.7)	1.830(0.781-4.290)	0.164
CT/TT	63(58.3)	85(69.3)	1.627(0.943-2.805)	0.08	54(51.0)	85(64.9)	1.776(1.051-3.002)	0.032
C	142(65.7)	141(57.8)	1.00(ref.)		147(69.3)	159(60.7)	1.00(ref.)	
T	74(34.3)	103(42.2)	1.410(0.965-2.061)	0.076	65(30.7)	103(39.3)	1.464(0.998-12.147)	0.041

### 六、不同肿瘤危险程度分层 SNPs (rs9350) 分析

分析结果见表 5.根据肿瘤侵袭性分为局限性前列腺癌组和进展性前列腺癌组。在局限性前列腺癌组中（对照组为正常人或非肿瘤患者）rs9350 CT/TT 基因型与更高的前列腺癌罹患风险明显相关（校正 OR=1.978,95%CI=1.070-3.022，

试验结果

p=0.027)。

表 5 不同肿瘤危险程度分层 SNPs 分析

Table 5		rs9350 genotype and allele frequencies [n (%)] and adjusted OR stratified by the aggressiveness of the cancer							
		localized				Advanced			
genotype	Controls	cases	Adjusted OR (95%		Controls	cases	Adjusted OR (95%		
or allele	n=214	n=91	CI)	P value	n=214	n=162	CI)	p value	
CC	97(45.3)	29(31.9)			97(45.3)	54(33.3)			
CT	95(44.4)	43(47.3)	1.550(0.890-2.699)	0.122	95(44.4)	91(56.2)	1.749(1.124-2.723)	0.013	
TT	22(10.3)	19(20.8)	3.077(1.440-6.573)	0.004	22(10.3)	17(10.5)	1.409(0.685-2.899)	0.351	
CT/TT	117(54.7)	62(68.1)	1.798(1.070-3.022)	0.027	117(54.6)	108(66.7)	1.350(0.981-1.857)	0.065	
C	289(67.5)	101(65.9)	1.00(ref.)		289(67.5)	199(61.4)	1.00(ref.)		
T	139(32.5)	81(34.1)	1.678(1.175-2.396)	0.004	139(32.5)	125(38.6)	1.313(0.970-1.776)	0.077	

七、其他临床指标 SNPs (rs9350) 分析

分析结果见表 6.除 PSA 水平之外 (PSA≤10 ng/ml 相比较 PSA>10 ng/ml) CT/TT 基因型存在率更高 (p=0.006), 其

余临床指标与 SNPs 之间并无明显的相关性。

表 6 临床指标与 rs9350 SNPs 分析

Table 6	Association of rs9350 with clinical characteristics										
	Aggressiveness of disease no.(%)		Gleason score [n(%)]		PSA level at diagnosis mean(ng/ml)		Mean age at diagnosis (years)	Smoking status		Drinking status	
	localized	advanced	≤7	>7	≤10	>10		Never	ever	Never	ever
No.of sub.	91	162	121	91	248	219	70.8(n=467)	140	327	87	380
reference CC	29(31.9)	54(33.3)	35(28.9)	35(38.5)	111(44.8)	71(32.4)	71.5(n=180)	51(36.4)	131(40.1)	30(34.5)	152(40.0)
Association											
CT/TT	62(68.1)	108(66.7)	86(71.1)	56(61.5)	137(55.2)	148(67.6)	70.5(n=287)	89(63.6)	196(59.9)	57(65.5)	
p value	0.812		0.184		0.006		0.526	0.461	0.341		

### 讨 论

引起前列腺癌的危险因素有待研究，一些已经可以确定的危险因素包括年龄、种族和遗传性。2个或者2个以上的一级亲属患病，患病危险性会增加5-11倍。通过全基因组相关研究，现阶段已经发现50余个单核苷酸多态性（SNP）与前列腺癌发病风险相关。有报告证实，发生5个或者5个以上的SNP的男性与无SNP的男性相比，罹患前列腺的风险比高达9.46。

日本的冈山大学科研小组经过与癌症相关的基因中的48个非同义SNPs的筛选，指出在12个SNPs中相关的10个基因对前列腺癌这种恶性肿瘤的发生与发展具有非常重要的意义。在这10个基因里，包含了1个代谢酶基因、1个凋亡调节基因、2个DNA修复基因、5个肿瘤抑制基因和1个染色体分离相关基因。在这里面有9个SNPs与恶性肿瘤的发生与发展的关系是首次被发现。在预测今后30年里面的前列腺癌罹患风险方面，风险组最高的可达29%，而风险组比较低的可为0.6%，其中风险组最低的仅仅约为0.2%。同时我们在预先评估前列腺癌罹患风险方面，发现这12个SNPs影响效应可以发生累积，风险组高与低的罹患率相比较，其OR值高达47.4。所以，在肿瘤的患病风险遗传学的评估中应用上与多种恶性肿瘤有关联性的SNPs的整合，与现有方式相比，对于罹患前列腺癌风险较高的人群的普筛是具有更大的准确性和更高的效率，并在肿瘤的早期诊断和预防措施的选择方面具有优势。

日本人群所在的地理位置与中国人群相近，同时两国人群也具有相似的人种起源，因此两国的遗传背景有很高的共同性。经过在日本大规模筛查得到的前列腺癌罹患风险相关的SNPs可能在中国人群中也会有部分适用。因此，在亚洲的四个国家，中、日、韩和新加坡联合进行了一项跨国际的多中心研究，目的是为了检验这一假说的准确与否，并将在日本的筛查结果推广到亚洲其他的国家，同时检验筛选出的前列腺癌高风险相关性SNPs在亚洲其他人群中的效力。本次试验中所选位点是来自日本冈山大学科研小组的筛查成果、一致于此次多中心试验的研究内容。



为了评估 DNA 的变异积累是否与前列腺癌相关，我们深入探究了中国人群对于前列腺癌易患性与 rs9350 之间的关系。我们的研究结果提示 EXO1 遗传性表型与前列腺癌易患性有明显相关。我们所调查的样本包含了中国北部、中部和南部的北京、杭州和广州的前列腺癌患者和健康人血液，来作为实验组和对照组。这些不同地区的数据结果可以帮助我们理解在美国[33]与台湾[34]地区 rs9350 与前列腺癌发病率的显著相关性。此次试验是已知首次例证了中国地区 rs9350 与前列腺癌罹患之间的关系。

EXO1 是特定结构的 5'核苷酸总科蛋白，可以识别单链 DNA 与双链 DNA 结合位点并将单个核苷酸插入双链 DNA 中。EXO1 是一个 5'到 3'端的核算外切酶，在基因错配修复、双链断裂修复和端粒酶维持中起着重要作用。EXO1 的主要作用是在基质，比如单链 DNA、RNA 和双链 DNA 中识别出特定结构损伤[35]。在前期研究中，我们已经阐述过了 EXO1 的基质和产物范围，同时它也和许多种类人体癌症，比如胃肠道[36]、子宫[37]和肺部癌症[38]及头颈部鳞癌[39]。

暴露在相关环境因素中被认为是肿瘤发展的重要病理因素。例如烟草中的致癌物质苯并芘，可以同上消化道的上皮细胞 DNA 相结合，形成 DNA 共价并引起复制错误，造成肿瘤发生。近期一个巴西的研究表明，在 DNA 错配修复系统中 EXO1 的亚正常状态是头颈鳞状细胞癌的主要决定因素，特别是在吸烟人群中，这个情况可以用于患者结果预测。

在我们的分析中，为了更好的比较 rs9350 和前列腺癌之间的关系，我们加入了环境暴露因素和家族病史。结果表示年纪（大于 72 岁），癌症类型（局限性或者进展性），PSA 水平（大于 10ng/ml）与前列腺癌高风险因素相关，但是与 gleason 评分、抽烟水平、喝酒水平并无太大联系。

我们的研究仍然有它的局限性。首先两组数据的量会限制统计学的意义，特别是不同因素不同年级等亚组。其次，对照组是随机选择的医院非肿瘤患者，他们是否形成了典型对照群组仍然有待商榷。最后，这些结果来自中国三个不同的地区，并不是确切的全国性数据。

总的来说, 我们提供了一份有关于中国地区的 EXO1 rs9350 与前列腺癌风险相关性的有力证明。尽管仍需后续研究包括更大样本更多人种和 SNPs 的功能性分析来充实和丰富我们的发现, 但是我们相信 EXO1 SNP rs3950 能够作为生物标记物为前列腺癌的评价和早期诊断的有效手段。

## 参考文献

1. Brown P , Hartwell L. Genomics and human disease-variations on variation.Nat Genet , 1998, 18:91~ 93.5
2. 刘炬, 徐兵河等. 单核苷酸多态性与乳腺癌患者的预后[ J] .癌症进展, 2007 , 5(3):242-248 .
3. Sherry ST, Ward MH, KholodovM, Baker J, Phan L, Smigielski EM, Sirotkin K.dbSNP:the NCBI database of genetic variation.Nucleicacids Res , 2001, 29:308~ 311.
4. 张凤珍, 翟静, 蒋汉明,deng 等.人类单核苷酸多态性及其应用[ J] .预防医学文献信息, 2004 , 10(5):569-573 .
5. 刘越, 吕社民. 单核苷酸多态性影响基因功能的机制[J]. 生命的化学,2008,(02): 214-216.
6. Bou ziqon E , Corda E .Effect of 17q21 variants and smoking exposure in early-onset asthma[ J] .N Engl J Med , 2008 , 359(19):1985-1994.
7. Rafal ski A.Appli cations of single nucleot ide polym orphism s incrop genetics[ J] .C urr Opin Plant Biol , 2002 , 5(2):94-100 .
8. HAMADA H,Petrino M G,A novel repeated element with z-DNA forming potential is widely found in diverse eukaryotic genomes.Proc NatlA cad SciUSA,1982,79:6465-6469
9. 赵春霞, 石先哲, 吕申, 等. 人类基因组的单核苷酸多态性及其研究进展[J]. 色谱, 2003, (02): 110-114.
10. E.H. Cook Jr., R.C. Arora, G.M. etc Platelet serotonin studies in hyperserotonemic relatives of children with autistic disorder Life Sci, 52 (1993), p. 2005
11. N. Yirmiya, T. Pilowsky,l etc.Evidence for an association with the serotonin transporter promoter region polymorphism and autism Am J Med Genet, 105 (2001), p.381

12. Bianchi,D.W. . From prenatal genomic diagnosis to fetal personalized medicine: Progress and challenges [J]. 2012,7
13. Underhill PA, Passarino G, Lin AA, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human population [j]. *Ann Hum genet.* 2001, 65(Pt 1); 43-62.
14. Onofri V, Alessandrini F, et al. Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs [J]. *Forensic Sci Int.*, 2006, 157(1): 23-25.
15. Hutcllin T. Haworth I, et al. A molecular basis for human hypersensitivity to aminoglycoside antibiotics. *Nucleic Acids Res*, 1993, 21: 4174-4179.
16. Altshuler D, Pollara V J, Cowles C R, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing [J]. *Nature*, 2000, 407(6803): 513-516.
17. Ru Q H, Jing H E, Luo G A, et al. Single-strand conformation polymorphism analysis to detect the p53 mutation in colon tumor samples by capillary electrophoresis [J]. *J Chromatogr A*, 2000, 894(1-2): 171-177
18. Sugano, K., Kyogoku, A., Fukayama, N., Ohkura, H., Shimosato, Y., Sekiya, T. and Hayashi, K. (1993) *Lab. Invest.* 68, 361-366.
19. XIAO W, OEFNER P J. Denaturing high performance liquid chromatography: a review [ J ]. *Hum Mutat*, 2001, 17( 6): 439- 474.
20. SE Johnston, M Lindqvist, E Niemelä, P Orell, J Erkinaro, . Fish scales and SNP chips: SNP genotyping and allele frequency estimation in individual and pooled DNA from historical samples of Atlantic salmon (*Salmo salar*) *BMC Genomics*, 2013, 14(27):3306-3310
21. Grimm RH Jr, Neaton JD, Elmer PJ, et al. The influence of oral potassium chloride on blood pressure in hypertensive men on a low-sodium diet. *N Engl J*

Med 1990;322:569-74

22. Wise C A, Paris M, Morar B, et al. A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry[J]. Rapid Commun Mass Spectrom, 2003, 17(11): 1195-1202.
23. Sauer S, Lehrach H, Reihardt R.MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging.Nucleic Acids Res , 2003, 31:e63.
24. Stanssens P, Zabeau M, Meersseman G, Remes G, Gamsemans Y, Storm N, Honisch C, Codi C P, Bocker S, van den Boom D . High-throughput MALDI-TOF discovery of genomic sequence polymorphisms.Genome Res , 2004, 14:126~133.
25. Kim S, Shi S, Bonome T, Ulz ME, Edwards JR, Fodstad H, Russo J J , Ju J Y. Multiplex genotyping of the human  $\beta$  2 adrenergic receptor gene using solid-phase capturable dideoxynucleotides and mass spectrometry. Anal Biochem, 2003, 316:251~ 258.
26. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate[J]. Science, 1998, 281(5375): 363, 365.
27. JOHN N EBLE.泌尿及男性生殖器官肿瘤病理学和遗传学.人民卫生.2006
28. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011; 61:69-90.
29. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004[J]. Ann Oncol, 2005, 16(3): 481-488.
30. Murray CJ, Atkinson C, Bhalla K, et al. The state of US health, 1990-2010:burden of diseases, injuries, and risk factors. JAMA 2013; 310:591-608.
31. 范天勇.前列腺癌病因及肿瘤标志物的研./华西医学.2003(1);27-29

32. Zheng S L, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer[J]. *N Engl J Med*, 2008, 358(9): 910-919
33. Tang H, Wei P, Duell EJ, Risch HA, Olson SH, Bueno-de-Mesquita HB, et al. Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data. *Carcinogenesis*.2014;35(5):1039-45.
34. Dong X, Li Y, Hess KR, Abbruzzese JL, Li D. DNA mismatch repair gene polymorphisms affect survival in pancreatic cancer. *Oncologist*. 2011;16(1):61-70.
35. Tsutakawa SE, Tainer JA. Double strand binding-single strand incision mechanism for human flap endonuclease: implications for the superfamily. *Mech Ageing Dev*. 2012;133(4):195-202.
36. Kim YR, Yoo NJ, Lee SH. Somatic mutation of EXO1 gene in gastric and colorectal cancers with microsatellite instability. *Acta Oncol*. 2010;49(6):859-60.
37. Tsai MH, Tseng HC, Liu CS, Chang CL, Tsai CW, Tsou YA, et al. Interaction of EXO1 genotypes and smoking habit in oral cancer in Taiwan. *Oral Oncol*. 2009;45(9):e90-44.
38. Bayram S. The exonuclease 1 Glu589Lys gene polymorphism and cancer susceptibility: evidence based on a meta-analysis. *Asian Pac J Cancer Prev*. 2014;15(6):2571-6.
39. Nogueira GA, Lourenço GJ, Oliveira CB, Marson FA, Lopes-Aguiar L, Costa EF, et al. Association between genetic polymorphisms in DNA mismatch repair-related genes with risk and prognosis of head and neck squamous cell carcinoma. *Int J Cancer*. 2015;137(4):810-8. *Tumor Biol*.

附件 1

调查问卷

1.姓名	2.ID号	2.出生日期	年 月 日		
4.入院日期	年 月 日	5.确诊日期	年 月 日		
6.初诊 PSA	total PSA: ng/ml	free PSA: ng/ml	f/t:	%	
7.病理诊断	1.前列腺癌 2.(非前列腺癌)肿瘤		3.其他		
8.TNM 分期 (若为前列腺癌)					
T 分类	0.T0	1.T1a	2.T1b	3.T1c	4.T2a 5.T2b
	6.T3a	7.T3b	8.T4	X.T <sub>x</sub>	
N 分类	0.N0	1.N1	2.N <sub>x</sub>		
M 分类	0.M0	1.M1a	2.M1b	3.M1c	X.M <sub>x</sub>
9.Gleason 评分 (若为前列腺癌)	2.2 3.3	4.4	5.5	6.6	7.7
	8.8	9.9	10.10	X. 不明	G= +
10.治疗方法	1. 手术 (a. TUERP+双侧睾丸切除术 b. TUERP+内分泌治疗 c. 根治术 d. TURP e. TUERP) 2. 激素治疗 3. 药物治疗 4. 放射线治疗 5. 其他				
11. 复发:	0. 无 1. 复发				
12. 预后	A. 生存 年 月 日 0. 无癌生存 1.癌生存 X. 不明 (其他肿瘤部位: ) B. 死亡 年 月 日 0.手术死亡 1.癌死亡 2.其他原因死亡 X.不明 (其他肿瘤部位: ) C. 随访不能				
13.家族史	若有肿瘤家族史, 则具体为:				
0.无肿瘤家族史	1.父母 2.兄弟 3.子女 4.其他 X.不明				
1.有肿瘤家族史	部位 ( ) ( ) ( )				
14.吸烟史:	1. 目前吸烟 1天 ( )支 X( )年 2. 过去吸烟,( )年前戒烟 1天 ( )支 X( )年 3. 不吸烟				
15.饮酒史:	1. 目前饮酒 1天 ( )g 1.白酒 2.啤酒 3.红酒 4.其他 X( )年 2. 过去饮酒,( )年前戒酒 1天 ( )g 1.白酒 2.啤酒 3.红酒 4.其他 X( )年 3. 不饮酒				

## 附件 2

### 知情同意书

本临床研究旨在从遗传流行病学的角度寻找前列腺癌的致病因素以及早期发现该肿瘤的方法。

您作为前列腺癌患者/与前列腺癌患者进行对照的非恶性肿瘤患者，已被邀请参与本项目。医生认为您是本项目合适的候选者，但是否进入本项目由您自己决定。请在决定是否参加本项目之前，阅读完以下的所有内容，并可以向医生询问任何相关的问题。

#### 操作过程的解释与描述

如果您同意参与该研究，我们将联系您的责任护士，在采集您的血液进行其他临床项目检查时多采集 2ml 静脉血，抽提 DNA 进行化验，以进行基因多态性分布影响前列腺癌风险的研究，血液的抽取只有一次。

#### 风险和不适，注意事项

对您本人没有任何风险，因为不会影响您的治疗。

#### 参加本项目可获得的益处

对您本人没有直接的好处，但对于今后前列腺癌的预防有很多帮助。

#### 保密问题

您的所有资料都是保密的，因为所有的医疗信息都将是医生医疗记录的一部分。你的治疗者可能在科学会议上使用这些医疗结果，或者在科学医疗杂志发表。在所有这些过程中，都不会出现您的名字。



### 参加和退出项目的权利

参加本项目是自愿的，您可以随时退出。如果您决定提前终止，您应与医生联系，以便医生能为您安排一次随访，以确定您下一步的治疗选择。

如果您不能遵守规定，不能按时随访以及完成各项检查或中途退出本项目，您将失去以上参与本项目应获得的益处。

### 合法权利

签署同意后书后，您拥有保护您健康的合法权益。如果您受到与本项目相关的伤害，您可以得到法律援助。

本知情同意书包括有关您决定参与项目的重要信息。如果您有任何本知情同意书未提到的问题，请询问您的医生。

我已经阅读和了解（或者有人向我讲解）以上有关的内容和治疗的危险及益处，我自愿加入本项目。

日期 年 月 日	参与者姓名 (正楷)	参与者签名
日期 年 月 日	见证人姓名 (正楷)	见证人签名
日期 年 月 日	医生姓名 (正楷)	医生签名

## 缩略词表

缩略词	英文全称	中文全称
A	Adenine	腺嘌呤
APC	adenomatous polyposis colli	肠腺瘤性息肉病蛋白
AR	Androgen receptor	雄激素受体
bp	Base pairs	碱基对
C	Cytosine	胞嘧啶
CI	Confidence intervals	可信区间
cSNP	Coding SNP	编码区内 SNP
CYP1A1	Cytochrome P450 1A1	细胞色素 P450 酶 1A1
DASH	Dynamic allele specific hybridization	动态等位基因特异性杂交
DHPLC	Denaturing high performance liquid chromatography	变性高效液相色谱
DHT	Dihydrotestosterone	双氢睾酮
DNA	Deoxyribonucleic acid	脱氧核糖核酸
dNTP	Deoxynucleotide triphosphate	脱氧核糖三磷酸腺苷
G	Guanine	鸟嘌呤
GSK-3	Glycogen synthase kinase-3	糖原合成酶激酶-3
GSTs	Glutathione-S-transferases	谷胱甘肽转硫酶
GWA	Whole genome association	全基因组相关性分析
H	Heterozygote	杂合子
HapMap	The International Haplotype Map Project	人类基因组单体型图计划
HGP	The Human Genome Project	人类基因组计划
HWD	Hardy-Weinberg Disequilibrium	Hardy-Weinberg 不平衡
HWE	Hardy-Weinberg Equilibrium	Hardy-Weinberg 平衡
LD	Linkage disequilibrium	连锁不平衡

MAF	Minor allele frequency	稀有等位基因频率
MM	Major homozygote	主要等位基因纯合子
MMPs	Matrix metalloproteinases	基质金属蛋白酶
MnSOD	Manganese superoxide dismutase	锰超氧化物歧化酶
mRNA	Messenger RNA	信使 RNA
ms-SNP	Missense SNP	错义单核苷酸多态性
mt DNA	Mitochondrial DNA	线粒体 DNA
OD	Optical density	光密度
OR	Odds ratio	比值比
ORF	Open reading frame	开放阅读框
PAGE	Polyacrylamide gel electrophoresis	聚丙烯酰胺凝胶电泳
PCR	Polymerase chain reaction	多聚酶链式反应
PCR-SSP	Polymerase chain reaction-sequence specific primer	多聚酶链反应-序列特异性引物
RFLP	Restriction fragment length polymorphism	限制性片段长度多态性
rpm	Round per minute	转每分
PSA	Prostate Specific-Antigen	前列腺特异性抗原
RR	Rare homozygote	稀有等位基因纯合子
SBE	Single nucleotide primer extension	单碱基引物延伸
SNPs	Single nucleotide polymorphisms	单核苷酸多态性
SRD5A2	Steroid 5-alpha reductase	5- $\alpha$ 还原酶
SSCP	Single strand conformation polymorphism	单链构象多态性
STR	Short tandem repeat	短串联重复序列
T	Thymine	胸腺嘧啶

缩略词表

---

T	Testosterone	睾酮
Taq	Thermus aquaticus	嗜热菌
Tm	Melting temperature	DNA 解链温度
tSNP	Tag SNP	标签 SNP
VDR	Vitamin D receptor	维生素 D 受体

---

## 攻读学位期间的发表的学术论文

- 1、Influence of a single-nucleotide polymorphism of the DNA mismatch repair-related gene exonuclease-1 (rs9350)with prostate cancer risk among Chinese people (共同一作, 发表于Tumor Biology)
- 2、Remarkable Pathologic Change in Advanced Prostate Cancer Patient Using Dendritic CelleCytokine-Induced Killer Combined Therapy: A Case Report (共同一作, 发表于Clin Genitourin Cancer)
- 3、New antibacterial isocoumarin glycosides from a wetland soil derived fungal strain *Metarhizium anisopliae*.(共同一作, 发表于BMCL)
- 4、Increased expression of metastasis-associated in colon cancer-1 in renal cell carcinoma is associated with poor prognosis (第五作者, 发表于Int J Clin Exp Pathol)
- 5、A potent inhibitor for *P.Aeruginosa* virulent factors by inhibiting Quorum-sensing signals. (撰写投稿中)

## 致 谢

弹指之间，三年时光已过，我的硕士研究生生涯已经逐渐接近了尾声。在这个时间里，首先我要感谢我的导师黄鹏教授对我的知遇之恩。三年中，黄鹏教授在我的生活、学习及工作给予的无微不至的关怀、指导，带给了我极大的帮助与启迪。跟随导师学习的这段经历，也将成为我人生之中一笔无法忘怀的珍贵记忆。感谢黄鹏教授在毕业论文的实验设计、实验过程、论文撰写以及发表过程中，给予的大力支持与帮助。

感谢珠江医院泌尿外科刘春晓主任在临床上对于我的训练与指导。跟随主任门诊、查房、病例讨论、课题讨论以及手术的经历，在我今后的从医生涯中，是一笔宝贵的人生财富。能够跟随刘主任学习是我的荣幸。

感谢南方医科大学珠江医院泌尿外科申海燕护士长在学习生活过程中给予的呵护和关心；感谢泌尿外科郑少波主任、徐亚文副主任、李虎林教授、李传印老师及徐阿白副主任医师，李炳坤副主任医师，许凯副主任医师，郭凯主治医师，杜伟主治医师在我参加科室学习期间给予的指导和帮助，感谢泌尿外科温勇、赵朋朋、邹勇、许鹏等师兄弟在生活及学业上给以的指点与支持。

感谢南方医科大学珠江医院泌尿外科全体护士在标本收集过程中给予的大力支持。

感谢南方医科大学所有关心和帮助过我的老师和同学。

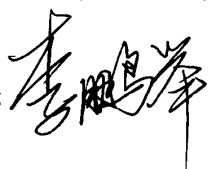
在这三年的时间里，有欢乐，有辛酸。漫漫求学路上，家人与朋友一直在背后默默的支持着我，给予了我充分的理解、关心、爱护与包容。不论我的今后的人生如何，家人永远是我背后的倚靠。这份谢意与感动难以通过言语来表达，只能通过这短短的篇幅来聊表万一。

只希望我今后通过努力与汗水所得到的渺渺成绩能够不辜负诸位的指导与关怀，能够不辜负这研究生时光，能够得到诸位的点头与肯定。谢谢大家。

## 南方医科大学 学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。除与外单位合作项目将予以明确方式规定外，本研究已发表与未发表成果的知识产权均归属南方医科大学。

本人承诺承担本声明的法律效果。

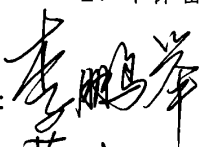
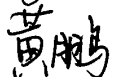
作者签名： 日期：2016年5月30日

## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权南方医科大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于（请在以下相应方框内打“√”）：

- 1、保密 ，在\_\_\_年解密后适用本授权书。
- 2、不保密

作者签名： 日期：2016年5月30日  
导师签名： 日期：2016年5月10日