Title: Genetic association of the KLK4 Locus with Risk of Prostate Cancer.

Felicity Lose,¹,⁷ Srilakshmi Srinivasan,² Tracy O’Mara,¹,² Louise Marquart,³ Suzanne Chambers,⁴,⁵,⁶ Robert A. Gardiner,⁶ Joanne F. Aitken,⁴,⁵ the Australian Prostate Cancer BioResource,⁷,⁸ Amanda B. Spurdle,¹,⁷ Jyotsna Batra,¹,* and Judith A. Clements²,*

Corresponding author: Dr Jyotsna Batra, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia 4059. Phone: +61 7 3138 6440, Fax: +61 7 3138 6030, email: jyotsna.batra@qut.edu.au.

* Equal last authors

¹. Molecular Cancer Epidemiology group, Genetics and Population Health Division, Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane, Queensland, Australia.

². Australian Prostate Cancer Research Centre – Queensland, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

³. Statistics Unit, Queensland Institute of Medical Research, Brisbane, Queensland, Australia.

⁴. Griffith Health Institute, Griffith University, Brisbane, Queensland, Australia.

⁵. Viertel Centre for Cancer Research, Cancer Council Queensland, Brisbane, Queensland, Australia.
6. University of Queensland Centre for Clinical Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia.

7. Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

# The Australian Prostate Cancer BioResource denotes the Queensland node participants in this study including –

Trina Yeadon, Pamela Saunders, Allison Eckert and Judith Clements - Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

Peter Heathcote, Glenn Wood, Greg Malone, Brisbane Urology Clinic, Brisbane Urology Clinic Central Queensland Urology Clinic, Wickham Terrace, Brisbane, Qld, Australia

Hema Samaratunga, Aquesta Pathology, Toowong, QLD, Australia

Angus Collins, Megan Turner and Kris Kerr, Sullivan and Nicolaides Pathology, Brisbane, Qld, Australia
Abstract

The Kallikrein-related peptidase, KLK4, has been shown to be significantly overexpressed in prostate tumours in numerous studies and is suggested to be a potential biomarker for prostate cancer. KLK4 may also play a role in prostate cancer progression through its involvement in epithelial-mesenchymal transition, a more aggressive phenotype, and metastases to bone. It is well known that genetic variation has the potential to affect gene expression and/or various protein characteristics and hence we sought to investigate the possible role of single nucleotide polymorphisms (SNPs) in the KLK4 gene in prostate cancer. Assessment of 61 SNPs in the KLK4 locus (±10 kb) in approximately 1300 prostate cancer cases and 1300 male controls for associations with prostate cancer risk and/or prostate tumour aggressiveness (Gleason score <7 versus ≥7) revealed 7 SNPs to be associated with a decreased risk of prostate cancer at the P_{trend}<0.05 significance level. Three of these SNPs, rs268923, rs56112930 and the HapMap tagSNP rs7248321, are located several kb upstream of KLK4; rs1654551 encodes a non-synonymous serine to alanine substitution at position 22 of the long isoform of the KLK4 protein, and the remaining 3 risk-associated SNPs, rs1701927, rs1090649 and rs806019, are located downstream of KLK4 and are in high linkage disequilibrium with each other (r²≥0.98). Our findings provide suggestive evidence of a role for genetic variation in the KLK4 locus in prostate cancer predisposition.
Introduction

The *Kallikrein (KLK)* gene family consists of 15 genes in a tightly clustered locus over 320 kilobases (kb) at 19q13.4 [1]. Many of the KLKs display altered expression in disease, in particular hormone-dependent cancers [1,2]. KLK4 is hormone-regulated and is expressed predominantly in the prostate [3,4], and to a lesser extent in other tissues [4,5]. KLK4 has gained support as a potential biomarker for several hormone-dependent cancers [2], and for prostate cancer specifically, in that numerous studies have found KLK4 to be significantly overexpressed in prostate carcinoma tissues compared to benign prostatic hyperplasia [6] and normal tissues [7,8,9,10,11]. Of note, KLK4 is known to be expressed as a variety of isoforms [12], with the full length protein (254 amino acids long) showing the potential to be a better biomarker of prostate tumour cells than the commonly expressed shorter isoform (205 amino acids) [11]. In addition, KLK4 has been proposed to play a role in prostate cancer progression through its involvement in epithelial-mesenchymal transition [13], a more aggressive phenotype, and metastases to bone [14]. KLK4 overexpression has been reported to be associated with prostate cancer stage, although the direction of effect differed for *KLK4* mRNA (associated with advanced stage) [6] versus KLK4 protein (early stage tumours) [15].

Approximately 40% of prostate cancer is estimated to have a genetic component (http://www.genome.gov/gwastudies/) [16], and to date single nucleotide polymorphisms (SNPs) in over 40 loci have been identified by genome-wide association studies (GWAS) to be associated with prostate cancer risk [17]. One of these SNPs is located in the *KLK* locus,
downstream of the \textit{KLK3} gene [18,19,20], and is thought to be a marker for a potentially functional non-synonymous SNP within the \textit{KLK3} gene [21]. Although no SNPs in \textit{KLK4} have been reported by GWAS to be associated with prostate cancer at genome-wide significance levels to date, commonly used GWAS chips only capture 22% [22] - 44% [23] of validated genetic variation in the locus with $r^2 \geq 0.80$. Hence we sought to comprehensively investigate the role of \textit{KLK4} in prostate cancer risk and tumour aggressiveness by genotyping the majority of validated genetic variation (±10 kb) around the \textit{KLK4} locus in a large prostate cancer study group and male controls not screened for PSA levels.
Materials and Methods

Study subjects

Study subjects have been described elsewhere [24,25]. Briefly, from 2004 onwards, 1349 histopathologically-confirmed prostate cancer cases were recruited through private and public urologists in Queensland, Australia via three prostate cancer studies or resources: the Retrospective Queensland Study (N=154; [26]), the Prostate Cancer Supportive Care and Patient Outcomes Project (ProsCan, N=857; [25]) and from the Australian Prostate Cancer BioResource (APCB, N=338; http://www.apccbioresource.org.au/index.html). Men presented to urologists with lower urinary tract symptoms and/or abnormal serum Prostate Specific Antigen (PSA), and 72% of cases possessed prostate tumours of Gleason score 7 or above. Cases ranged in age at diagnosis from 40-88 years (median 63 years). Male controls (N=1405) with no self-reported personal history of prostate cancer were randomly selected from the Australian Electoral Roll and age-matched (in 5 year groups) and post-code matched to cases (N=569), or recruited through the Australian Red Cross Blood Services in Brisbane (N=836). Controls were not screened for PSA levels and analyses excluded 50 controls with age at interview <40 years (the age of the youngest case); included controls ranged in age at interview from 40-89 years of age (median 62 years). All participants had self-reported European ethnicity and gave written informed consent. The study protocol was approved by the Human Research Ethics Committees of the Queensland University of Technology, Queensland Institute of Medical Research, the Mater Hospital (for Brisbane Private Hospital), the Royal Brisbane Hospital, Princess Alexandra Hospital and the Cancer Council Queensland.
SNP selection and genotyping

The KLK4 gene region used for SNP selection was chr19:56091420...56115806 (hg18), which encompasses the longest KLK4 isoform ±10 kb. All SNPs in this region were extracted from the National Center for Biotechnology Information (NCBI) dbSNP build 130 [27], CHIP SNPper [28] and the “ParSNPs” database [29] and duplicates removed. SNPs not classified as validated were removed and validated SNPs were further investigated for occurrence in Europeans using SPSmart [30] and 1000 Genomes [23]. Additional SNPs excluded from investigation included all SNPs on the Illumina 550K, 610K and Omni1 genome-wide genotyping chips and SNPs assessed in the Cancer Genetics Markers of Susceptibility (CGEMS) project [31], unless there was evidence of association with prostate cancer by CGEMS (P<0.05). SNPs in high linkage disequilibrium (LD; r^2 ≥ 0.80) with these excluded Illumina and CGEMS SNPs were also removed, determined by the SNP Annotation and Proxy Search program (SNAP) version 2.1 [32] using HapMap release 22 (1000 Genomes data was not available at the time of initiation of this study). We then prioritised for genotyping all independent SNPs (r^2 < 0.80) according to SNAP using HapMap release 22 data (N=74). An additional 8 KLK4 tagSNPs (selected using HapMap data release 24/phase II, Nov 2008, NCBI build 36, dbSNP b126, using the Tagger program within Haplovie v4.1 [33]), genotyped as part of a previous study, were also included (N=82 overall).

SNPs were genotyped using iPLEX Gold assays on the Sequenom MassARRAY platform (Sequenom, San Diego, CA), as described previously [34]. There were 4 negative (H_2O) controls per 384-well plate, and quality control parameters included genotype call rates
>95%, a combination of cases and controls on each plate, inclusion of 20 duplicate samples per 384-well plate (>5% of samples) with ≥98% concordance between duplicates and Hardy-Weinberg Equilibrium P values >0.05. Of a total of 82 KLK4 SNPs selected for investigation, 11 could not be designed for Sequenom assays, and after application of quality control parameters, 61 SNPs were successfully genotyped. After the study was completed, 1000 Genomes data became available and revealed that 6 KLK4 SNPs not genotyped directly in our study (rs2659108, rs1654556, rs1090648, rs11881373, rs2569531 and rs73598979) were actually tagged by our genotyped SNPs (r²>0.80).

Statistical methods

Predictive Analytics Software (PASW) Statistics version 17.0.2 (SPSS Inc., Chicago, IL) was used for all analyses. Genotype and allele frequencies were calculated for the patient and control groups. SNP allele and genotype distributions were compared using χ² and their association with prostate cancer susceptibility and clinical data were performed under codominant and linear models using logistic regression analysis. Prostate cancer cases with tumour Gleason scores ≥7 were classified as aggressive. All analyses were adjusted for age (as a continuous variable).

SNP function prediction

Alibaba (http://labmom.com/link/alibaba_2_1_tf_binding_prediction), TFsearch (http://www.cbrc.jp/research/db/TFSEARCH.html) and MatInspector
were used to predict the transcription factor binding sites. The program SignalP was used to predict the signal peptide. miRNADA (http://www.microrna.org/microrna/home.do), Patrocles and miRBase (http://www.mirbase.org/) were used to determine the effect of the SNP alleles on miRNA binding. JASPAR (http://jaspar.binf.ku.dk/), CISTER (http://zlab.bu.edu/~mfrith/cister.shtml) and NHRScan were used for the prediction of nuclear hormone receptor response elements. Splicing effects (using splice-finder), protein structure and stability (Polyphen, SIFT, SNP3D) were determined through the SNPinfo web server (http://snpinfo.niehs.nih.gov). Histone marks, DNase hypersensitive sites and conservation scores were obtained from HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php), which extracts data from the UCSC Browser (http://genome.ucsc.edu/). The F-SNP web server (http://compbio.cs.queensu.ca/F-SNP) was used to determine the functional score and putative effect of each SNP.
Results

Seven SNPs were found to be monomorphic in our sample group (Table S1). Results of analyses of the remaining 54 KLK4 SNPs and risk of prostate cancer are displayed in Table 1. Although no KLK4 SNPs were statistically significantly associated with prostate cancer risk after Bonferroni correction \( (P<9\times10^{-4}) \), 7 SNPs were associated at the \( P_{\text{trend}}<0.05 \) significance level and the majority of these displayed a modest decrease in prostate cancer risk of around 20%. Two of these SNPs, rs268923 (Odds Ratio (OR) 0.89, 95% Confidence Interval (CI) 0.79-1.00, \( P_{\text{trend}}=0.045 \)) and rs56112930 (OR 0.37, 95% CI 0.14-0.96, \( P_{\text{trend}}=0.040 \); Minor Allele Frequency (MAF) 0.006), are located several kilobases upstream of the long isoform of KLK4, 8.2 kb and 6.5 kb, respectively. The KLK4 tagSNP rs7248321 (OR 0.77, 95% CI 0.60-0.98, \( P_{\text{trend}}=0.033 \); MAF 0.060), also represented on several of the genome-wide chips including the Illumina 550K, 610K and Omni1 chips, is located \(~4.5\) kb upstream of KLK4, and rs7248321 tags two nearby SNPs rs13345980 and rs7246794 with \( r^2 \) 1.00. rs1654551 (OR 0.79, 95% CI 0.65-0.97, \( P_{\text{trend}}=0.023 \); MAF 0.093) is a non-synonymous SNP in the full length KLK4 protein coding for a serine to alanine amino acid (aa) substitution at position 22. In the more commonly expressed 205 aa KLK4 isoform, this SNP is located in the 5’ untranslated region [11]. The remaining three risk-associated SNPs, rs1701927, rs1090649 and rs806019, we determined to be in high LD with each other \( (r^2\geq0.98) \) and accordingly all display ORs of around 0.85 (95% CI range 0.73-1.00, \( P_{\text{trend}} \) range 0.030-0.044; MAF 0.175). Results were similar for rs1701926 that is also part of this high LD block (OR 0.86, 95% CI 0.74-1.00, \( P_{\text{trend}}=0.058 \)). All four SNPs are located downstream of the KLK4 gene from 750 base pairs (bp) to 3.6 kb past the 3’ untranslated region (UTR).
Only one *KLK4* SNP, rs198968, was associated with prostate tumour aggressiveness (Gleason score <7 vs. ≥7: OR 0.76 (95% CI 0.60-0.95, \(P_{\text{trend}}=0.016\); Table 2). However, this result was not reflected in a more robust Gleason score analysis comparing “extreme” Gleason categories, ≤6 (N=329) vs. ≥8 (N=173), with an odds ratio of 0.95 (95% CI 0.68-1.32; \(P_{\text{trend}}=0.752\)).

Results of bioinformatic prediction of functions of the associated SNPs are provided in Table S2. SNPs rs268923, rs198968, rs1654551, rs1701926, rs1090649 and rs806019 were found to alter transcription factor binding sites as predicted by at least one prediction tool. rs198968 and rs1654551 also lie within promoter histone marks as well as DNAse hypersensitive sites (Table S2), and hence are better candidate for functional follow up studies.

SNP rs1654551 leads to a serine to alanine amino acid change, but is predicted to be benign using the FASTSNP web server, although the SNP is predicted to effect o-glycosylation. In addition, PsortII prediction (http://urgi.versailles.inra.fr/Tools/PsortII) predicted the serine variant to be only 44.4% extracellularly localised as compared to the alanine variant which is predicted to be 55.6% extracellular. This is backed by SignalP predicting alteration of the *KLK4* signal peptide sequence for the serine variant. Further, this SNP is also predicted to be involved in differential splicing.
Discussion

We performed a comprehensive investigation of the role of variation in the *KLK4* gene in prostate cancer risk and/or tumour aggressiveness by assessing the majority of SNPs that have not been covered by previously performed GWA studies. Our study of approximately 1,300 cases and 1,300 male controls provided suggestive evidence that several *KLK4* SNPs may be associated with decreased risk of prostate cancer, and bioinformatic analysis provides evidence that some of these have potential biological relevance in prostate cancer.

None of the nominally risk-associated SNPs were located in known *KLK4* hormone response elements [35]. Three SNPs lay several kb upstream of the *KLK4* gene. rs7248321 is a tagSNP that has not been previously reported to be associated with prostate cancer risk in any GWAS, including CGEMS, and given the large numbers of samples assessed in previous studies [17], it is likely to be a false-positive result. Bioinformatic analyses of the rare rs56112930 SNP did not reveal any predicted effects on transcription factor binding sites [36,37,38]. SNP rs268923 was calculated by three different transcription factor binding site prediction programs to possibly have an effect [36,37,38,39], and although each program predicts different transcription factor binding sites to be altered by the SNP; one example of a prostate cancer relevant result is the predicted gain of an Oct-1 site [36]. Oct-1 is a known co-regulator of the androgen receptor [40], regulates growth of prostate cancer cells and is associated with poor prognosis [41].
The only SNP located in the KLK4 coding region found to be marginally associated with prostate cancer risk was rs1654551. Since splicing of the KLK4 locus is complex and results in several KLK4 mRNA forms being produced [12], there are several possible functional consequences of this substitution. The two protein isoforms identified to date, in order of expression in normal prostate, are an intracellular 205 amino acid (aa) protein which lacks the classical KLK signal peptide and is localised to the nucleus (“short” isoform) [9,11], and a secreted 254 aa protein that is cytoplasmically localised [11,13]. rs1654551 codes for a serine to alanine substitution at amino acid 22 of the long isoform, or is located in the 5’ UTR of the 205 aa KLK4 protein. Although both the short and long isoforms have been found to be overexpressed in prostate cancer cells, the “long” 254 aa KLK4 protein is better able to discriminate between tumour and normal cells [11] and hence may be the more biologically relevant isoform in prostate cancer. Amino acid 22 is located within the signal peptide region of KLK4, which is cleaved off between aa 26 and 27 to result in secretion. It is unknown what the potential functional effects of an amino acid substitution are within the signal peptide. However, a recent study has shown that this cleaved peptide may be a useful target in prostate cancer immunotherapy, with the KLK4 signal peptide successfully inducing and expanding the cytotoxic T lymphocyte response more readily than PSA or Prostatic Acid Phosphotase (PAP) [42]. In addition, in silico analysis using the signal peptide prediction program SignalP [43] predicted a Serine22Alanine substitution to alter the cleavage site from aa 26/27 to aa 21/22. This would result in a KLK4 pro-protein with an additional 5 aa, which could potentially affect localisation or possibly even activation of the KLK4 proenzyme. Of relevance, a form of PSA has been reported that has an altered signal/pro-peptide and, although the pro-PSA sequence is truncated (not lengthened as is predicted for KLK4), the
signal peptide alteration does result in an isoform of PSA that is unable to be activated [44]. This [-2]pro-PSA isoform is now also the basis of a commercially available prostate cancer serum test [45].

Attempting to predict the possible functional effects of the four associated SNPs located downstream of *KLK4*, rs1701927, rs1701926, rs1090649 and rs806019, is not as clearly directed. It is possible that some or all of these SNPs might alter enhancer/silencer binding sites, affecting expression of *KLK4*. *In silico* transcription factor binding site analysis predicts that rs1701926, rs1090649 and rs806019 may alter transcription factor binding sites [36,37,38,39] relevant to prostate cancer. The closest validated gene to the *KLK4* 3’ end is the Kallikrein family pseudogene *KLKP1*, thus these SNPs may regulate the activity of this pseudogene or expression of *KLKP1* transcripts, which have been shown to be down-regulated in prostate cancer tissues [46]. In addition, one other SNP not genotyped in this study, rs1654556, is in high LD (r²≥0.80) with these SNPs [23] and is predicted to alter mRNA folding [47] and miRNA binding (Table S2).

To the best of our knowledge, only two other studies have examined the role of *KLK4* SNPs in cancer (aside from genome-wide investigations). Recently Klein *et al.* investigated the effect of common variation in the exons and putative promoter regions of all 15 *KLK* genes on prostate cancer risk and levels of PSA forms and KLK2 [48]. Five *KLK4* SNPs – rs198969, rs198968, rs1654552, rs1654551 and rs1654553 - were assessed for association with prostate cancer risk in the Cancer Prostate in Sweden (CAPS) 1 sample set of over 1,400
cases and 700 controls and none were found to be associated. A second study in a small Korean sample set of 117 breast cancer cases and 194 controls found \textit{KLK4} SNP rs806019 to be associated with a decreased risk of breast cancer (Odds Ratio 0.53; 95% Confidence Interval 0.33-0.85; \(P=0.007\)) [49], a finding of similar magnitude and direction to that observed in our study of prostate cancer. Part of our study design was to exclude \textit{KLK4} SNPs already assessed in GWAS, except for those reported to be associated with prostate cancer at the \(P<0.05\) level in CGEMS. Four SNPs in CGEMS gave evidence of association with prostate cancer - rs17714461, rs8101572, rs8100631 and rs10427094 [31]. All four were genotyped in this study, but none were associated in our sample set. Of note, rs17714461 was recently reported to interact with the GWAS-detected \textit{KLK2/3} SNP rs2735839 in CGEMS [50]. The authors mention that this result is notable considering \textit{KLK4} and \textit{KLK2} collaborate to stimulate cellular proliferation in prostate cancer [51]. rs2735839 genotype data was available for only a small proportion of our samples and hence we did not investigate this interaction further.

Our well-sized study indicates a possible contribution of SNPs in the \textit{KLK4} gene to decreased risk of prostate cancer. However, these results should be interpreted cautiously considering the number of tests performed, and validation in much larger sample sets such as those of the PRACTICAL consortium is necessary.
Acknowledgements

The authors thank the many patients and control subjects who participated so willingly in this study, and the numerous institutions and their staff who have supported recruitment. The authors are very grateful to staff at the Australian Red Cross Blood Services for their assistance with the collection of risk factor information and blood samples of healthy donor controls; members of the Cancer Council Queensland for ProsCan participant information, including Megan Ferguson and Andrea Kittila; the hospitals that participated in recruitment for the ProsCan study: Greenslopes Private, Royal Brisbane, Mater Adults, Princess Alexandra, Ipswich, QEI, Redlands and Redcliffe Hospitals, Townsville General Hospital, and Mackay Base Hospital; Drs. John Yaxley and David Nicol for recruitment of patients into the Retrospective Queensland Study; and the Urological Society of Australia and New Zealand. Thank you to members of the Australian Prostate Cancer Research Centre-Queensland at QUT and the QIMR Molecular Cancer Epidemiology Laboratory for their assistance with recruitment and biospecimen processing, and Dr John Lai, XiaoQing Chen and Dr Jonathan Beesley for technical advice.
References

1. Lawrence MG, Lai J, Clements JA (2010) Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. Endocr Rev 31: 407-446.
2. Mavridis K, Scorilas A (2010) Prognostic value and biological role of the kallikrein-related peptidases in human malignancies. Future Oncol 6: 269-285.
3. Nelson PS, Gan L, Ferguson C, Moss P, Gelinars R, et al. (1999) Molecular cloning and characterization of prostatase, an androgen-regulated serine protease with prostate-restricted expression. Proc Natl Acad Sci U S A 96: 3114-3119.
4. Yousef GM, Obiezu CV, Luo LY, Black MH, Diamandis EP (1999) Prostate/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated. Cancer Res 59: 4252-4256.
5. Obiezu CV, Shan SJ, Soosaipillai A, Luo LY, Grass L, et al. (2005) Human kallikrein 4: quantitative study in tissues and evidence for its secretion into biological fluids. Clin Chem 51: 1432-1442.
6. Avgeris M, Stravodimos K, Scorilas A (2011) Kallikrein-related peptidase 4 gene (KLK4) in prostate tumors: quantitative expression analysis and evaluation of its clinical significance. Prostate 71: 1780-1789.
7. Day CH, Fanger GR, Retter MW, Hylander BL, Penetrante RB, et al. (2002) Characterization of KLK4 expression and detection of KLK4-specific antibody in prostate cancer patient sera. Oncogene 21: 7114-7120.
8. Obiezu CV, Soosaipillai A, Jung K, Stephan C, Scorilas A, et al. (2002) Detection of human kallikrein 4 in healthy and cancerous prostatic tissues by immunofluorometry and immunohistochemistry. Clin Chem 48: 1232-1240.
9. Xi Z, Klokk TI, Korkmaz K, Kurys P, Elbi C, et al. (2004) Kallikrein 4 is a predominantly nuclear protein and is overexpressed in prostate cancer. Cancer Res 64: 2365-2370.
10. Klokk TI, Kilander A, Xi Z, Waehre H, Risberg B, et al. (2007) Kallikrein 4 is a proliferative factor that is overexpressed in prostate cancer. Cancer Res 67: 5221-5230.
11. Dong Y, Bui LT, Odorico DM, Tan OL, Myers SA, et al. (2005) Compartmentalized expression of kallikrein 4 (KLK4/hK4) isoforms in prostate cancer: nuclear, cytoplasmic and secreted forms. Endocr Relat Cancer 12: 875-889.
12. Kurlender L, Borgono C, Michael IP, Obiezu C, Elliott MB, et al. (2005) A survey of alternative transcripts of human tissue kallikrein genes. Biochim Biophys Acta 1755: 1-14.
13. Veveris-Lowe TL, Lawrence MG, Collard RL, Bui L, Herington AC, et al. (2005) Kallikrein 4 (hK4) and prostate-specific antigen (PSA) are associated with the loss of E-cadherin and an epithelial-mesenchymal transition (EMT)-like effect in prostate cancer cells. Endocr Relat Cancer 12: 631-643.
14. Gao J, Collard RL, Bui L, Herington AC, Nicol DL, et al. (2007) Kallikrein 4 is a potential mediator of cellular interactions between cancer cells and osteoblasts in metastatic prostate cancer. Prostate 67: 348-360.
15. Seiz L, Kotzsch M, Grebenchtchikov NI, Geurts-Moespot AJ, Fuessel S, et al. (2010) Polyclonal antibodies against kallikrein-related peptidase 4 (KLK4): immunohistochemical assessment of KLK4 expression in healthy tissues and prostate cancer. Biol Chem 391: 391-401.
16. Gronberg H (2003) Prostate cancer epidemiology. Lancet 361: 859-864.
17. Hindorff LA, Junkins HA, Hall PN, Mehta JP, Manolio TA A Catalog of Published Genome-Wide Association Studies.
18. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, et al. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40: 316-321.
19. Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, et al. (2008) Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. Cancer Epidemiol Biomarkers Prev 17: 2052-2061.
20. Eeles RA, Giles GG, Neal DE, Muir K, Easton DF (2008) Reply to "Variation in KLK genes, prostate-specific antigen and risk of prostate cancer". Nat Genet 40: 1035-1036.
21. Kote-Jarai Z, Amin Al Olama A, Leongamornlert D, Tymrakiewicz M, Saunders E, et al. (2011) Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. Hum Genet 129: 687-694.
22. (2003) The International HapMap Project. Nature 426: 789-796.
23. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.
24. Lose F, Batra J, O'Mara T, Fahey P, Marquart L, et al. (2011) Common variation in Kallikrein genes KLK5, KLK6, KLK12, and KLK13 and risk of prostate cancer and tumor aggressiveness. Urol Oncol.
25. Baade PD, Aitken JF, Ferguson M, Gardiner RA, Chambers SK (2010) Diagnostic and treatment pathways for men with prostate cancer in Queensland: investigating spatial and demographic inequalities. BMC Cancer 10: 452.
26. Lai J, Kedda MA, Hinek Z, Smith RL, Yaxley J, et al. (2007) PSA/KLK3 AREI promoter polymorphism alters androgen receptor binding and is associated with prostate cancer susceptibility. Carcinogenesis 28: 1032-1039.
27. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308-311.
28. Riva A, Kohane IS (2004) A SNP-centric database for the investigation of the human genome. BMC Bioinformatics 5: 33.
29. Goard CA, Bromberg IL, Elliott MB, Diamandis EP (2007) A consolidated catalogue and graphical annotation of dbSNP polymorphisms in the human tissue kallikrein (KLK) locus. Mol Oncol 1: 303-312.
30. Amigo J, Salas A, Phillips C, Carracedo A (2008) SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. BMC Bioinformatics 9: 428.
31. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, et al. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39: 645-649.
32. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ Jr, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24: 2938-2939.
33. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haplovie: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265.
34. Lose F, Nagle CM, O'Mara T, Batra J, Bolton KL, et al. (2010) Vascular endothelial growth factor gene polymorphisms and ovarian cancer survival. Gynecol Oncol 119: 479-483.
35. Lai J, Myers SA, Lawrence MG, Odorico DM, Clements JA (2009) Direct progesterone receptor and indirect androgen receptor interactions with the kallikrein-related peptidase 4 gene promoter in breast and prostate cancer. Mol Cancer Res 7: 129-141.
36. Grabe N (2000) AliBaba2.1.
37. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, et al. (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 26: 362-367.
38. Akiyama Y (2002) TFSEARCH: Searching Transcription Factor Binding Sites.
39. Cartharius K, Frech K, Grote K, Kloccke B, Haltmeier M, et al. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21: 2933-2942.
40. Gonzalez MI, Robins DM (2001) Oct-1 preferentially interacts with androgen receptor in a DNA-dependent manner that facilitates recruitment of SRC-1. J Biol Chem 276: 6420-6428.
41. Obinata D, Takayama K, UranO T, Murata T, Kumagai J, et al. (2012) Oct1 regulates cell growth of LNCaP cells and is a prognostic factor for prostate cancer. Int J Cancer 130: 1021-1028.
42. Wilkinson R, Woods K, D'Rozario R, Prue R, Vari F, et al. (2012) Human kallikrein 4 signal peptide induces cytotoxic T cell responses in healthy donors and prostate cancer patients. Cancer Immunol Immunother 61: 169-179.
43. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783-795.
44. Mikolajczyk SD, Marker KM, Millar LS, Kumar A, Saedi MS, et al. (2001) A truncated precursor form of prostate-specific antigen is a more specific serum marker of prostate cancer. Cancer Res 61: 6958-6963.
45. Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, et al. (2011) A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. J Urol 185: 1650-1655.
46. Kaushal A, Myers SA, Dong Y, Lai J, Tan OL, et al. (2008) A novel transcript from the KLKP1 gene is androgen regulated, down-regulated during prostate cancer progression and encodes the first non-serine protease identified from the human kallikrein gene locus. Prostate 68: 381-399.
47. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.
48. Klein RJ, Hallden C, Cronin AM, Ploner A, Wiklund F, et al. (2010) Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer. Cancer Prev Res (Phila) 3: 611-619.
49. Lee JY, Park AK, Lee KM, Park SK, Han S, et al. (2009) Candidate gene approach evaluates association between innate immunity genes and breast cancer risk in Korean women. Carcinogenesis 30: 1528-1531.
50. Ciampa J, Yeager M, Amundadottir L, Jacobs K, Kraft P, et al. (2011) Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study. Cancer Res 71: 3287-3295.
51. Mize GJ, Wang W, Takayama TK (2008) Prostate-specific kallikreins-2 and -4 enhance the proliferation of DU-145 prostate cancer cells through protease-activated receptors-1 and -2. Mol Cancer Res 6: 1043-1051.
**Supporting Information Legends**

Table S1. rs IDs found to be monomorphic in this study

Table S2. *In silico* function prediction of prostate cancer associated KLK4 gene variants
### Table 1. Association of *KLK4* SNPs and prostate cancer risk

| SNP          | Genotype | Controls | Cases | Adjusted |
|--------------|----------|----------|-------|----------|
|              | Chr position | n (%)   | n (%) | OR (95% CI) | P      |
| rs17714450   | GG       | 1000 (78.9) | 984 (77.5) | 1.00 |        |
| 56115669     | GA       | 255 (20.1)  | 277 (21.8)  | 1.11 (0.91-1.34) | 0.307 |
|              | AA       | 12 (0.9)    | 9 (0.7)     | 0.77 (0.32-1.84) | 0.556 |
|              | Per A allele |        |        | 1.07 (0.89-1.28) | 0.477 |
| rs8101572    | AA       | 395 (30.5)  | 389 (30.7)  | 1.00 |        |
| 56115440     | AC       | 619 (47.8)  | 634 (50.0)  | 1.03 (0.86-1.24) | 0.726 |
|              | CC       | 281 (21.7)  | 245 (19.3)  | 0.87 (0.70-1.09) | 0.236 |
| rs8100631    | GG       | 489 (37.8)  | 474 (37.4)  | 1.00 |        |
| 56115358     | GA       | 586 (45.3)  | 592 (46.7)  | 1.03 (0.87-1.22) | 0.731 |
|              | AA       | 220 (17.0)  | 201 (15.9)  | 0.93 (0.74-1.17) | 0.535 |
| rs268920C    | CC       | 1295 (100.0) | 1267 (99.9) | 1.00 |        |
| 56114746     | CG       | 0 (0.0)     | 1 (0.1)     | N/A    | N/A    |
|              | GG       | 0 (0.0)     | 0 (0.0)     | N/A    | N/A    |
|              | Per G allele |        |        | N/A    | N/A    |
| rs10427094   | CG       | 1130 (88.9) | 1113 (90.7) | 1.00 |        |
| 56114689     | CT       | 137 (10.8)  | 112 (9.1)   | 0.84 (0.64-1.09) | 0.183 |
|              | TT       | 4 (0.3)     | 2 (0.2)     | 0.53 (0.10-2.92) | 0.466 |
| rs268921     | CC       | 477 (36.9)  | 467 (36.8)  | 1.00 |        |
| 56114503     | CG       | 592 (45.7)  | 594 (46.8)  | 1.01 (0.85-1.20) | 0.891 |
|              | GG       | 225 (17.4)  | 207 (16.3)  | 0.93 (0.74-1.17) | 0.518 |
|              | Per G allele |        |        | 0.97 (0.87-1.09) | 0.613 |
| rs2569526    | AA       | 1182 (88.1) | 1185 (90.7) | 1.00 |        |
| 56106299     | AG       | 156 (11.6)  | 121 (9.3)   | 0.78 (0.60-1.00) | 0.048 |
|              | GG       | 3 (0.2)     | 1 (0.1)     | 0.35 (0.04-3.35) | 0.359 |
|              | Per G allele |        |        | 0.77 (0.60-0.98) | 0.033 |
| rs2569526    | AA       | 1259 (99.0) | 1225 (99.7) | 1.00 |        |
| 56106299     | AG       | 13 (1.0)    | 4 (0.3)     | 0.34 (0.11-1.06) | 0.062 |
| rs1654551          | TT  | 1060 (81.9) | 1082 (85.3) | 1.00 |
|------------------|-----|-------------|-------------|------|
|                  | TG  | 227 (17.5)  | 181 (14.5)  | 0.79 (0.64-0.97) | 0.028 |
|                  | GG  | 7 (0.5)     | 5 (0.4)     | 0.67 (0.21-1.23) | 0.499 |
| Per G allele     |     |             |             | **0.79 (0.65-0.97)** | **0.023** |
| rs1654552        | GG  | 403 (31.1)  | 376 (29.7)  | 1.00 |
|                  | GT  | 603 (46.6)  | 637 (50.3)  | 1.15 (0.96-1.37) | 0.140 |
|                  | TT  | 288 (22.3)  | 254 (20.0)  | 0.95 (0.76-1.19) | 0.658 |
| Per T allele     |     |             |             | 0.99 (0.89-1.10) | 0.847 |
| rs2242670        | CC  | 418 (31.2)  | 359 (27.4)  | 1.00 |
|                  | CT  | 634 (47.4)  | 674 (51.5)  | 1.24 (1.04-1.48) | 0.017 |
|                  | TT  | 286 (21.4)  | 277 (21.1)  | 1.12 (0.90-1.39) | 0.309 |
| Per T allele     |     |             |             | 1.07 (0.96-1.19) | 0.210 |
| rs198966         | CC  | 1294 (99.9) | 1265 (99.8) | 1.00 |
|                  | CT  | 1 (0.1)     | 2 (0.2)     | 2.47 (0.22-27.42) | 0.460 |
|                  | TT  | 0 (0.0)     | 1 (0.1)     | N/A | N/A |
| Per T allele     |     |             |             | 3.39 (0.45-25.58) | 0.236 |
| rs34626614       | GG  | 1294 (99.9) | 1267 (99.9) | 1.00 |
|                  | GA  | 1 (0.1)     | 1 (0.1)     | 1.18 (0.07-18.86) | 0.909 |
|                  | AA  | 0 (0.0)     | 0 (0.0)     | N/A | N/A |
| Per A allele     |     |             |             | 1.18 (0.07-18.86) | 0.909 |
| rs2569527        | AA  | 1274 (98.4) | 1255 (99.0) | 1.00 |
|                  | AC  | 21 (1.6)    | 13 (1.0)    | 0.62 (0.31-1.25) | 0.186 |
|                  | CC  | 0 (0.0)     | 0 (0.0)     | N/A | N/A |
| Per C allele     |     |             |             | 0.62 (0.31-1.25) | 0.186 |
| rs189903         | CC  | 1258 (100.0) | 1184 (99.8) | 1.00 |
|                  | CT  | 0 (0.0)     | 2 (0.2)     | N/A | N/A |
|                  | TT  | 0 (0.0)     | 0 (0.0)     | N/A | N/A |
| Per T allele     |     |             |             | N/A | N/A |
| rs2979451        | AA  | 686 (52.0)  | 625 (48.2)  | 1.00 |
|                  | AG  | 507 (38.4)  | 562 (43.3)  | 1.21 (1.03-1.42) | 0.021 |
| SNP      | 5' Allele | 3' Allele | Per G Allele | 0.96 (0.73-1.27) | 0.772 | 1.06 (0.95-1.20) | 0.298 |
|----------|-----------|-----------|--------------|------------------|-------|------------------|-------|
| rs1701929<sup>b</sup> | TT        | CC        | 703 (52.8)  | 721 (55.0)     | 1.00  | 0.94 (0.80-1.10) | 0.411 |
|          | TC        | AG        | 540 (40.6)  | 520 (39.6)     | 0.99 (0.77-1.27) | 0.912 |
|          | CC        | GG        | 88 (6.6)    | 71 (5.4)       | 0.77 (0.56-1.08) | 0.130 |
|          | Per C Allele | Per A Allele | 0.91 (0.80-1.03) | 0.135 |
| rs7255024 | CC        | AA        | 1070 (88.2) | 1083 (88.3)    | 1.00  | 1.09 (1.04-1.49) | 0.019 |
|          | CA        | GG        | 140 (11.5)  | 140 (11.4)     | 1.24 (1.02-1.42) | 0.025 |
|          | AA        | Per G Allele | 3 (0.2)     | 4 (0.3)        | 1.36 (0.30-6.12) | 0.687 |
|          | Per C Allele | Per A Allele | 1.00 (0.79-1.27) | 0.986 |
| rs1654553<sup>b</sup> | TT        | CC        | 390 (29.2)  | 344 (26.3)     | 1.00  | 1.03 (0.83-1.27) | 0.821 |
|          | TC        | AG        | 626 (46.9)  | 679 (51.9)     | 1.24 (1.04-1.49) | 0.019 |
|          | CC        | GG        | 320 (24.0)  | 286 (21.8)     | 1.03 (0.83-1.27) | 0.821 |
|          | Per C Allele | Per A Allele | 1.02 (0.92-1.14) | 0.687 |
| rs2235091<sup>b</sup> | TT        | CC        | 583 (43.6)  | 520 (39.6)     | 1.00  | 1.21 (1.02-1.42) | 0.025 |
|          | TC        | AG        | 563 (42.1)  | 609 (46.4)     | 1.21 (1.02-1.42) | 0.025 |
|          | CC        | GG        | 192 (14.3)  | 183 (13.9)     | 1.06 (0.84-1.35) | 0.606 |
|          | Per C Allele | Per A Allele | 1.07 (0.96-1.20) | 0.208 |
| rs35945487 | CC        | TT        | 1262 (97.5) | 1238 (97.6)    | 1.00  | 0.93 (0.56-1.55) | 0.790 |
|          | CT        | 0 (0.0)   | 32 (2.5)    | 30 (2.4)       | 0.93 (0.56-1.55) | 0.790 |
|          | Per T Allele | Per A Allele | N/A         | N/A            |
| rs73042387 | GG        | CC        | 884 (73.7)  | 901 (74.6)     | 1.00  | 0.96 (0.80-1.16) | 0.686 |
|          | GA        | AG        | 291 (24.3)  | 287 (23.8)     | 0.96 (0.80-1.16) | 0.686 |
|          | AA        | GG        | 25 (2.1)    | 20 (1.7)       | 0.80 (0.44-1.45) | 0.460 |
|          | Per A Allele | Per A Allele | 0.94 (0.80-1.11) | 0.484 |
| rs1139132 | CC        | AA        | 540 (44.9)  | 488 (40.2)     | 1.00  | 1.24 (1.04-1.47) | 0.015 |
|          | CA        | AC        | 499 (41.4)  | 563 (46.4)     | 1.24 (1.04-1.47) | 0.015 |
|          | AA        | CC        | 165 (13.7)  | 163 (13.4)     | 1.08 (0.84-1.38) | 0.555 |
|          | Per A Allele | Per A Allele | 1.09 (0.97-1.22) | 0.152 |
| rs1701927 | AA        | CC        | 882 (68.1)  | 906 (71.5)     | 1.00  | 0.87 (0.73-1.03) | 0.110 |
|          | AC        | CG        | 374 (28.9)  | 336 (26.5)     | 0.87 (0.73-1.03) | 0.110 |
|          | CC        | TT        | 39 (3.0)    | 26 (2.1)       | 0.65 (0.39-1.07) | 0.091 |
|          | Per C Allele | Per A Allele | 0.85 (0.73-0.98) | 0.030 |
| rs1701926 | TT        | CC        | 863 (68.3)  | 871 (71.1)     | 1.00  | 0.91 (0.76-1.08) | 0.273 |
|          | TG        | CG        | 363 (28.7)  | 332 (27.1)     | 0.91 (0.76-1.08) | 0.273 |
|          | GG        | TT        | 37 (2.9)    | 22 (1.8)       | 0.58 (0.34-1.00) | 0.050 |
|          | Per G Allele | Per T Allele | 0.86 (0.74-1.00) | 0.058 |
| rs2569530 | GG        | CC        | 334 (30.1)  | 274 (25.6)     | 1.00  | 1.33 (1.09-1.62) | 0.006 |
|          | GT        | TG        | 509 (45.9)  | 553 (51.6)     | 1.33 (1.09-1.62) | 0.006 |
|          | TT        | TT        | 265 (23.9)  | 244 (22.8)     | 1.12 (0.89-1.42) | 0.332 |
|          | Per T Allele | Per A Allele | 1.07 (0.95-1.20) | 0.258 |
| rs1090649 | CC        | AA        | 860 (68.3)  | 871 (71.0)     | 1.00  | 0.92 (0.77-1.10) | 0.344 |
|          | CG        | AG        | 358 (28.4)  | 333 (27.2)     | 0.92 (0.77-1.10) | 0.344 |
|          | GG        | GG        | 41 (3.3)    | 22 (1.8)       | 0.53 (0.31-0.89) | 0.017 |
|          | Per G Allele | Per A Allele | 0.86 (0.73-1.00) | 0.044 |
| rs11881354 | AA        | CG        | 557 (44.1)  | 504 (40.6)     | 1.00  | 1.19 (1.00-1.41) | 0.045 |
|          | AG        | GG        | 531 (42.0)  | 574 (46.3)     | 1.19 (1.00-1.41) | 0.045 |
|          | GG        | Per G Allele | 1.01 (0.79-1.29) | 0.945 |
| rs806019  | CC        | CG        | 880 (68.0)  | 903 (71.2)     | 1.00  | 0.87 (0.73-1.04) | 0.129 |
| SNP     | Allele | Frequency | Per G allele | Per T allele | Per C allele |
|---------|--------|-----------|--------------|--------------|--------------|
| rs806020 | GG     | 365 (28.9)| 326 (26.4)   | 1.00         |              |
|         | GA     | 591 (46.9)| 632 (51.2)   | 1.21 (1.00-1.46) | 0.049        |
|         | AA     | 305 (24.2)| 277 (22.4)   | 1.03 (0.82-1.28) | 0.802        |
|         | Per A |              | 1.02 (0.91-1.14) | 0.704        |
| rs806021 | AA     | 374 (29.6)| 338 (27.3)   | 1.00         |              |
|         | AG     | 592 (46.9)| 625 (50.5)   | 1.17 (0.97-1.41) | 0.093        |
|         | GG     | 297 (23.5)| 274 (22.2)   | 1.03 (0.82-1.28) | 0.802        |
|         | Per G |              | 1.02 (0.92-1.14) | 0.696        |
| rs806022 | GG     | 378 (29.2)| 332 (26.2)   | 1.00         |              |
|         | GT     | 610 (47.1)| 650 (51.3)   | 1.22 (1.01-1.46) | 0.037        |
|         | TT     | 306 (23.6)| 285 (22.5)   | 1.07 (0.86-1.33) | 0.559        |
|         | Per T |              | 1.04 (0.93-1.16) | 0.473        |
| rs806023 | AA     | 379 (29.4)| 342 (27.1)   | 1.00         |              |
|         | AG     | 608 (47.1)| 640 (50.7)   | 1.17 (0.97-1.40) | 0.097        |
|         | GG     | 303 (23.5)| 281 (22.2)   | 1.03 (0.83-1.29) | 0.766        |
|         | Per G |              | 1.02 (0.92-1.14) | 0.669        |
| rs2569535 | GG    | 1278 (98.9)| 1256 (99.5) | 1.00         |              |
|         | GC    | 14 (1.1)  | 6 (0.5)      | 0.48 (0.18-1.24) | 0.130        |
|         | CC    | 0 (0.0)   | 0 (0.0)      | N/A          | N/A          |
|         | Per C |              | 0.48 (0.18-1.24) | 0.130        |
| rs1701941 | TT    | 357 (27.8)| 316 (25.0)   | 1.00         |              |
|         | TA    | 606 (47.9)| 650 (51.4)   | 1.19 (0.99-1.44) | 0.065        |
|         | AA    | 312 (24.3)| 298 (23.6)   | 1.08 (0.87-1.35) | 0.478        |
|         | Per A |              | 1.04 (0.94-1.17) | 0.441        |
| rs1654513 | AA    | 550 (42.5)| 555 (43.8)   | 1.00         |              |
|         | AG    | 593 (45.8)| 578 (45.6)   | 0.97 (0.82-1.14) | 0.701        |
|         | GG    | 152 (11.7)| 135 (10.6)   | 0.88 (0.68-1.14) | 0.341        |
|         | Per G |              | 0.95 (0.84-1.07) | 0.374        |
| rs8104538 | TT    | 621 (48.0)| 606 (47.8)   | 1.00         |              |
|         | TC    | 542 (41.9)| 535 (42.2)   | 1.00 (0.85-1.18) | 0.978        |
|         | CC    | 132 (10.2)| 126 (9.9)    | 0.98 (0.75-1.28) | 0.883        |
|         | Per C |              | 0.99 (0.88-1.12) | 0.926        |
| rs10409668 | CC   | 1295 (100.0)| 1267 (99.9) | 1.00         |              |
|         | CT    | 0 (0.0)   | 1 (0.1)      | N/A          | N/A          |
|         | TT    | 0 (0.0)   | 0 (0.0)      | N/A          | N/A          |
|         | Per T |              | N/A          | N/A          |
| rs1560719 | TT    | 521 (40.4)| 475 (38.3)   | 1.00         |              |
|         | TC    | 569 (44.1)| 581 (46.8)   | 1.12 (0.95-1.33) | 0.178        |
|         | CC    | 201 (15.6)| 185 (14.9)   | 1.01 (0.80-1.27) | 0.956        |
|         | Per C |              | 1.03 (0.92-1.15) | 0.614        |
| rs2739493 | TT    | 489 (37.8)| 435 (34.4)   | 1.00         |              |
|         | TG    | 583 (45.1)| 609 (48.1)   | 1.17 (0.98-1.39) | 0.076        |
|         | GG    | 221 (17.1)| 221 (17.5)   | 1.11 (0.88-1.39) | 0.368        |
|         | Per G |              | 1.07 (0.96-1.20) | 0.210        |
| rs1701934 | GG    | 1283 (99.1)| 1256 (99.1) | 1.00         |              |
|         | GA    | 12 (0.9)  | 11 (0.9)     | 0.97 (0.43-2.21) | 0.944        |
|         | AA    | 0 (0.0)   | 1 (0.1)      | N/A          | N/A          |
|         | Per A |              | 1.15 (0.54-2.46) | 0.716        |
| rs1560723 | TT    | 1282 (99.1)| 1255 (99.1) | 1.00         |              |
|         | TC    | 12 (0.9)  | 11 (0.9)     | 0.97 (0.43-2.21) | 0.944        |
|         | CC    | 0 (0.0)   | 1 (0.1)      | N/A          | N/A          |
| SNP          | Allele | n (%) | n (%) | OR (95% CI) | P value |
|--------------|--------|-------|-------|-------------|---------|
| rs1701936    | CC     | 1292 (99.8) | 1263 (99.7) | 1.15 (0.54-2.46) | 0.716   |
|              | CT     | 3 (0.2)      | 3 (0.2)      | 1.17 (0.23-5.83) | 0.849   |
|              | TT     | 0 (0.0)      | 1 (0.1)      | N/A          | N/A     |
| Per C allele |        |           |           |             |         |
| rs1654514    | AA     | 1282 (99.0) | 1256 (99.1) | 1.74 (0.47-6.42) | 0.405   |
|              | AG     | 13 (1.0)     | 11 (0.9)     | 0.89 (0.40-2.00) | 0.780   |
|              | GG     | 0 (0.0)      | 1 (0.1)      | N/A          | N/A     |
| Per G allele |        |           |           |             |         |
| rs1701937    | CC     | 1292 (99.8) | 1264 (99.7) | 1.00         |         |
|              | CA     | 3 (0.2)      | 3 (0.2)      | 1.17 (0.23-5.82) | 0.850   |
|              | AA     | 0 (0.0)      | 1 (0.1)      | N/A          | N/A     |
| Per A allele |        |           |           |             |         |

SNP, single nucleotide polymorphism; chr, chromosome; n, number; OR, odds ratio; CI, confidence interval.

\(^a\)Co-ordinates from hg18

\(^b\)tagSNP

\(^c\)SNP too infrequent in these groups to calculate OR (95% CI)

**Bold**, SNPs displaying a $P_{\text{trend}}$ value <0.05
### Table 2. Association of KLK4 SNPs and prostate tumour aggressiveness

| SNP          | Genotype | Gleason <7 | Gleason ≥7 | Adjusted Chr positiona | n (%) | n (%) | OR (95% CI) | P       |
|--------------|----------|------------|------------|-------------------------|-------|-------|-------------|---------|
| rs17714461   | GG       | 241 (75.1) | 640 (76.9) |                         | 1.00  |       |             |         |
|              | GA       | 77 (24.0)  | 257 (22.4) |                         | 0.86  | 0.63-1.17 | 0.340       |
|              | AA       | 3 (0.9)    | 7 (0.6)    |                         | 0.60  | 0.13-2.73 | 0.510       |
|              |          |            |            |                         |       |       |             |         |
|              | Per A allele |       |            |                         | 0.85  | 0.64-1.13 | 0.269       |
| rs17714450   | GG       | 316 (98.8) | 813 (98.7) |                         | 1.00  |       |             |         |
|              | GA       | 4 (1.3)    | 11 (1.3)   |                         | 1.05  | 0.33-3.34 | 0.936       |
|              | AA       | 0 (0.0)    | 0 (0.0)    |                         | N/A   | N/A   |             |         |
|              | Per A allele |       |            |                         | 1.05  | 0.33-3.34 | 0.936       |
| rs8101572    | AA       | 106 (33.1) | 246 (29.8) |                         | 1.00  |       |             |         |
|              | AC       | 154 (48.1) | 417 (50.5) |                         | 1.14  | 0.85-1.53 | 0.397       |
|              | CC       | 60 (18.8)  | 162 (19.6) |                         | 1.12  | 0.77-1.63 | 0.563       |
|              | Per C allele |       |            |                         | 1.07  | 0.89-1.29 | 0.494       |
| rs8100631    | GG       | 125 (39.1) | 301 (36.5) |                         | 1.00  |       |             |         |
|              | GA       | 145 (45.3) | 393 (47.7) |                         | 1.10  | 0.83-1.46 | 0.506       |
|              | AA       | 50 (15.6)  | 130 (15.8) |                         | 1.05  | 0.71-1.55 | 0.817       |
|              | Per A allele |       |            |                         | 1.04  | 0.86-1.25 | 0.685       |
| rs268920     | CC       | 320 (100.0)| 824 (99.9) |                         | 1.00  |       |             |         |
|              | CG       | 0 (0.0)    | 1 (0.1)    |                         | N/A   | N/A   |             |         |
|              | GG       | 0 (0.0)    | 0 (0.0)    |                         | N/A   | N/A   |             |         |
|              | Per G allele |       |            |                         | N/A   | N/A   |             |         |
| rs10427094   | CC       | 292 (91.8) | 717 (90.4) |                         | 1.00  |       |             |         |
|              | CT       | 26 (8.2)   | 75 (9.5)   |                         | 1.16  | 0.72-1.85 | 0.540       |
|              | TT       | 0 (0.0)    | 1 (0.1)    |                         | N/A   | N/A   |             |         |
|              | Per T allele |       |            |                         | 1.19  | 0.75-1.89 | 0.466       |
| rs268921     | CC       | 123 (38.4) | 297 (36.0) |                         | 1.00  |       |             |         |
|              | CG       | 144 (45.0) | 395 (47.9) |                         | 1.11  | 0.83-1.48 | 0.476       |
|              | GG       | 53 (16.6)  | 133 (16.1) |                         | 1.01  | 0.69-1.48 | 0.974       |
|              | Per G allele |       |            |                         | 1.02  | 0.85-1.23 | 0.806       |
| rs268923     | AA       | 113 (36.8) | 297 (37.4) |                         | 1.00  |       |             |         |
|              | AT       | 140 (45.6) | 390 (49.1) |                         | 1.04  | 0.78-1.40 | 0.778       |
|              | TT       | 54 (17.6)  | 108 (13.6) |                         | 0.75  | 0.51-1.12 | 0.163       |
|              | Per T allele |       |            |                         | 0.90  | 0.74-1.09 | 0.293       |
| rs10419776   | CC       | 190 (79.2) | 414 (73.8) |                         | 1.00  |       |             |         |
|              | CG       | 49 (20.4)  | 138 (24.6) |                         | 1.23  | 0.85-1.79 | 0.273       |
|              | GG       | 1 (0.4)    | 9 (1.6)    |                         | 3.70  | 0.46-29.58 | 0.217       |
|              | Per G allele |       |            |                         | 1.31  | 0.93-1.85 | 0.128       |
| rs56112930   | CC       | 312 (99.7) | 804 (99.6) |                         | 1.00  |       |             |         |
|              | CT       | 1 (0.3)    | 3 (0.4)    |                         | 1.34  | 0.14-13.06 | 0.803       |
|              | TT       | 0 (0.0)    | 0 (0.0)    |                         | N/A   | N/A   |             |         |
|              | Per T allele |       |            |                         | 1.34  | 0.14-13.06 | 0.803       |
| rs7248321b   | AA       | 303 (90.4) | 762 (91.0) |                         | 1.00  |       |             |         |
| SNP         | Allele | Ref | Alt | Prevalence | Effect Size | 95% CI       | p-value |
|-------------|--------|-----|-----|------------|-------------|--------------|---------|
| rs2569526   | AA     | 318 (99.7) | 792 (99.7) | 1.00       |             |             |         |
| rs2978642   | TT     | 184 (57.9)  | 475 (57.6)   | 1.00       |             |             |         |
| rs198969b   | GG     | 84 (25.1)    | 231 (27.6)   | 1.00       |             |             |         |
| rs198968b   | CC     | 209 (66.8)   | 557 (69.1)   | 1.00       |             |             |         |
| rs198967    | CC     | 319 (99.7)   | 824 (99.9)   | 1.00       |             |             |         |
| rs1654551   | TT     | 269 (84.1)   | 709 (85.9)   | 1.00       |             |             |         |
| rs1654552   | GG     | 98 (30.6)    | 247 (30.0)   | 1.00       |             |             |         |
| rs2242670b  | CC     | 90 (26.8)    | 236 (28.1)   | 1.00       |             |             |         |
| rs198966    | CC     | 319 (99.7)   | 824 (99.9)   | 1.00       |             |             |         |
| rs34626614c | GG     | 320 (100.0)  | 825 (100.0)  | 1.00       |             |             |         |
| rs            | Allele 1 | Allele 2 | Allele 3 | Allele 4 | Allele 5 | Allele 6 |
|--------------|----------|----------|----------|----------|----------|----------|
| rs2569527    | AA       | 0 (0.0)  | 0 (0.0)  | N/A      | N/A      |          |
| rs189903     | CC       | 299 (100.0) | 762 (99.9) | 1.00     |          |          |
| rs73042387   | Per T    | N/A      | N/A      |          |          |          |
| rs2979451    | AA       | 175 (53.2) | 394 (47.1) | 1.00     |          |          |
| rs1701929    | TT       | 180 (53.6) | 460 (54.7) | 1.00     |          |          |
| rs7255024    | CC       | 272 (87.7) | 696 (88.3) | 1.00     |          |          |
| rs1654553    | AA       | 87 (26.0)  | 226 (26.9) | 1.00     |          |          |
| rs2235091    | TT       | 146 (43.5) | 329 (39.1) | 1.00     |          |          |
| rs35945487   | CC       | 312 (97.5) | 806 (97.7) | 1.00     |          |          |
| rs73042387   | GG       | 73 (21.8)  | 189 (22.5) | 1.03     | 0.71     | 0.93     |
| rs1139132    | CC       | 126 (41.7) | 326 (41.5) | 1.00     |          |          |
| rs1701927    | AA       | 223 (69.7) | 592 (71.8) | 1.00     |          |          |

**Allele frequencies and p-values**

- **Per A allele**
- **Per C allele**
- **Per T allele**
- **Per G allele**

**Allele frequencies**

- **CC**
- **CA**
- **AA**
- **TC**
- **TT**
- **AG**
- **GG**
- **TT**
- **AC**
- **CC**

**P-values**

- **(0.00)**
- **(0.01)**
- **(0.02)**
- **(0.03)**

**Allele counts**

- **rs2569527**
- **rs189903**
- **rs73042387**
- **rs2979451**
- **rs1701929**
- **rs7255024**
- **rs1654553**
- **rs2235091**
- **rs35945487**
- **rs73042387**
- **rs1139132**
- **rs1701927**

**Allele counts**

- **rs2569527**
- **rs189903**
- **rs73042387**
- **rs2979451**
- **rs1701929**
- **rs7255024**
- **rs1654553**
- **rs2235091**
- **rs35945487**
- **rs73042387**
- **rs1139132**
- **rs1701927**
| SNP             | Allele | Chi-sq | Odds Ratio (95% CI) | P      |
|-----------------|--------|--------|---------------------|--------|
| rs1701926       | Per C allele | 0.88 (0.68-1.13) | 0.325 |
| 56100570        | TT     | 568 (71.7) | 1.00               |
|                 | TG     | 211 (26.6) | 0.87 (0.65-1.17)   | 0.359  |
|                 | GG     | 13 (1.6)   | 0.60 (0.24-1.49)   | 0.272  |
|                 | Per G allele | 0.85 (0.65-1.09) | 0.198 |
| rs2569530       | GG     | 181 (26.2) | 1.00               |
| 56100420        | GT     | 352 (51.0) | 0.87 (0.62-1.23)   | 0.438  |
|                 | TT     | 157 (22.8) | 0.89 (0.59-1.33)   | 0.570  |
|                 | Per T allele | 0.94 (0.77-1.15) | 0.560 |
| rs1090649       | CC     | 567 (71.7) | 1.00               |
| 56098343        | CG     | 368 (51.0) | 1.05 (0.79-1.39)   | 0.726  |
|                 | GG     | 111 (13.8) | 1.15 (0.75-1.74)   | 0.521  |
|                 | Per G allele | 0.84 (0.65-1.09) | 0.183 |
| rs11881354      | AA     | 328 (40.6) | 1.00               |
| 56097941        | AG     | 368 (45.6) | 1.05 (0.79-1.39)   | 0.726  |
|                 | GG     | 111 (13.8) | 1.15 (0.75-1.74)   | 0.521  |
|                 | Per G allele | 0.88 (0.68-1.13) | 0.323 |
| rs806019        | CC     | 590 (71.5) | 1.00               |
| 56097794        | CG     | 220 (26.7) | 0.93 (0.69-1.24)   | 0.618  |
|                 | GG     | 15 (1.8)   | 0.60 (0.26-1.41)   | 0.241  |
|                 | Per G allele | 0.84 (0.65-1.09) | 0.183 |
| rs806020        | GG     | 216 (26.8) | 1.00               |
| 56097620        | GA     | 402 (51.0) | 0.94 (0.69-1.29)   | 0.692  |
|                 | AA     | 187 (23.2) | 1.11 (0.76-1.63)   | 0.579  |
|                 | Per A allele | 1.05 (0.87-1.27) | 0.615 |
| rs806021        | AA     | 223 (27.7) | 1.00               |
| 56097485        | AG     | 397 (49.3) | 0.91 (0.67-1.25)   | 0.564  |
|                 | GG     | 185 (23.0) | 1.10 (0.75-1.61)   | 0.613  |
|                 | Per G allele | 1.04 (0.86-1.26) | 0.669 |
| rs806022        | GG     | 220 (26.7) | 1.00               |
| 56096999        | GT     | 417 (50.5) | 0.95 (0.70-1.30)   | 0.763  |
|                 | TT     | 188 (22.8) | 1.03 (0.71-1.50)   | 0.884  |
|                 | Per T allele | 1.01 (0.84-1.22) | 0.903 |
| rs806023        | AA     | 227 (27.6) | 1.00               |
| 56096896        | AG     | 409 (49.8) | 0.93 (0.68-1.27)   | 0.639  |
|                 | GG     | 186 (22.6) | 1.04 (0.71-1.51)   | 0.854  |
|                 | Per G allele | 1.01 (0.84-1.22) | 0.891 |
| rs2569535       | GG     | 821 (99.9) | 1.00               |
| 56096639        | GC     | 1 (0.1)    | 0.16 (0.02-1.56)   | 0.114  |
|                 | CC     | 0 (0.0)    | N/A                | N/A    |
|                 | Per C allele | 0.16 (0.02-1.56) | 0.114 |
| rs1701941       | TT     | 209 (25.4) | 1.00               |
| 56096032        | TA     | 418 (50.9) | 0.97 (0.71-1.34)   | 0.872  |
|                 | AA     | 195 (23.7) | 1.00 (0.69-1.45)   | 0.995  |
|                 | Per A allele | 1.00 (0.83-1.20) | 0.998 |
| SNP          | Allele 1 | Allele 2 | Allele 3 | Allele 4 | P-value (OR 95% CI) |
|--------------|----------|----------|----------|----------|---------------------|
| rs1654513    | AA       | 132 (41.3) | 367 (44.5) | 1.00     | 0.92 (0.70-1.21)    |
|              | AG       | 148 (46.3) | 375 (45.5) | 0.09     | 0.75 (0.49-1.15)    |
|              | GG       | 40 (12.5)  | 83 (10.1)  | 0.00     | 0.88 (0.73-1.07)    |
| Per G allele |          |           |          |          | 0.189              |
| rs8104538    | TT       | 161 (50.3) | 393 (47.7) | 1.00     | 1.11 (0.84-1.47)    |
|              | TC       | 126 (39.4) | 345 (41.9) | 0.00     | 1.09 (0.70-1.70)    |
|              | CC       | 33 (10.3)  | 86 (10.4)  | 0.00     | 1.07 (0.88-1.30)    |
| Per C allele |          |           |          |          | 0.512              |
| rs10409668   | CC       | 320 (100.0) | 824 (99.9) | 1.00     | 1.07 (0.88-1.30)    |
|              | CT       | 0 (0.0)    | 1 (0.1)   | N/A      | N/A                |
|              | TT       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per T allele |          |           |          |          | 0.857              |
| rs1560719    | TT       | 129 (41.1) | 301 (37.4) | 1.00     | 1.07 (0.72-1.59)    |
|              | TC       | 137 (43.6) | 383 (47.6) | 0.00     | 1.04 (0.71-1.53)    |
|              | CC       | 48 (15.3)  | 121 (15.0) | 0.00     | 1.06 (0.85-1.29)    |
| Per C allele |          |           |          |          | 0.819              |
| rs2739493    | TT       | 112 (35.0) | 276 (33.6) | 1.00     | 1.07 (0.72-1.59)    |
|              | TG       | 153 (47.8) | 401 (48.8) | 0.00     | 1.04 (0.71-1.53)    |
|              | GG       | 55 (17.2)  | 145 (17.6) | 0.00     | 1.02 (0.85-1.23)    |
| Per G allele |          |           |          |          | 0.819              |
| rs1701934    | GG       | 316 (98.8) | 819 (99.3) | 1.00     | 1.07 (0.72-1.59)    |
|              | GA       | 4 (1.3)    | 6 (0.7)   | 0.66     | 0.66 (0.18-2.37)    |
|              | AA       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per A allele |          |           |          |          | 0.521              |
| rs1560723    | TT       | 316 (98.8) | 818 (99.3) | 1.00     | 1.07 (0.72-1.59)    |
|              | TC       | 4 (1.3)    | 6 (0.7)   | 0.66     | 0.66 (0.18-2.37)    |
|              | CC       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per C allele |          |           |          |          | 0.522              |
| rs1701936    | CC       | 318 (99.4) | 823 (99.9) | 1.00     | 1.07 (0.72-1.59)    |
|              | CT       | 2 (0.6)    | 1 (0.1)   | 0.28     | 0.28 (0.03-3.18)    |
|              | TT       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per T allele |          |           |          |          | 0.307              |
| rs1654514    | AA       | 316 (98.8) | 819 (99.3) | 1.00     | 1.07 (0.72-1.59)    |
|              | AG       | 4 (1.3)    | 6 (0.7)   | 0.66     | 0.66 (0.18-2.37)    |
|              | GG       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per G allele |          |           |          |          | 0.521              |
| rs1701937    | CC       | 318 (99.4) | 824 (99.9) | 1.00     | 1.07 (0.72-1.59)    |
|              | CA       | 2 (0.6)    | 1 (0.1)   | 0.28     | 0.28 (0.03-3.17)    |
|              | AA       | 0 (0.0)    | 0 (0.0)   | N/A      | N/A                |
| Per A allele |          |           |          |          | 0.306              |

SNP, single nucleotide polymorphism; chr, chromosome; n, number; OR, odds ratio; CI, confidence interval. aCo-ordinates from hg18

bTagSNP; cSNP too infrequent in these groups to calculate OR (95% CI)

**Bold**, SNPs displaying a P_trend value <0.05