

Identification of a novel prostate cancer susceptibility variant in the *KLK3* gene transcript

Z. Kote-Jarai · A. Amin Al Olama · D. Leongamornlert · M. Tymrakiewicz · E. Saunders · M. Guy · G. G. Giles · G. Severi · M. Southey · J. L. Hopper · K. C. Sit · J. M. Harris · J. Batra · A. B. Spurdle · J. A. Clements · F. Hamdy · D. Neal · J. Donovan · K. Muir · P. D. P. Pharoah · S. J. Chanock · N. Brown · S. Benlloch · E. Castro · N. Mahmud · L. O'Brien · A. Hall · E. Sawyer · R. Wilkinson · D. F. Easton · R. A. Eeles

Received: 17 January 2011 / Accepted: 14 March 2011 / Published online: 5 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Genome-wide association studies (GWAS) have identified more than 30 prostate cancer (PrCa) susceptibility loci. One of these (rs2735839) is located close to a plausible candidate susceptibility gene, *KLK3*, which encodes prostate-specific antigen (PSA). PSA is widely used as a biomarker for PrCa detection and disease monitoring. To refine the association between PrCa and variants

in this region, we used genotyping data from a two-stage GWAS using samples from the UK and Australia, and the Cancer Genetic Markers of Susceptibility (CGEMS) study. Genotypes were imputed for 197 and 312 single nucleotide polymorphisms (SNPs) from HapMap2 and the 1000 Genome Project, respectively. The most significant association with PrCa was with a previously unidentified SNP, rs17632542 (combined $P = 3.9 \times 10^{-22}$). This association was confirmed by direct genotyping in three stages of the UK/Australian GWAS, involving 10,405 cases and 10,681 controls (combined $P = 1.9 \times 10^{-34}$). rs17632542 is also shown to be associated with PSA levels and it is a non-synonymous coding SNP (Ile179Thr) in *KLK3*. Using

Z. Kote-Jarai and A. Amin Al Olama are the joint first authors.
D. F. Easton, R. A. Eeles are the joint last authors.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-0981-1) contains supplementary material, which is available to authorized users.

Z. Kote-Jarai (✉) · D. Leongamornlert · M. Tymrakiewicz · E. Saunders · M. Guy · N. Brown · E. Castro · N. Mahmud · L. O'Brien · A. Hall · E. Sawyer · R. Wilkinson · R. A. Eeles
The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK
e-mail: zsofia.kote-jarai@icr.ac.uk

A. Amin Al Olama · S. Benlloch · D. F. Easton
Department of Public Health and Primary Care, Centre for Cancer Epidemiology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge CB1 8RN, UK

G. G. Giles · G. Severi · M. Southey
Cancer Epidemiology Centre, The Cancer Council Victoria, 100 Drummond Street, Carlton, VIC 3053, Australia

G. G. Giles · G. Severi · M. Southey · J. L. Hopper
Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, 723 Swanston Street, Carlton, VIC 3053, Australia

K. C. Sit · J. M. Harris · J. Batra · J. A. Clements
Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove Urban Village, Kelvin Grove, QLD 4059, Australia

A. B. Spurdle
Genetics and Population Health Division,
Queensland Institute of Medical Research,
300 Herston Rd, Herston, QLD 4006, Australia

F. Hamdy
Nuffield Department of Surgery, University of Oxford,
Oxford OX3 9DU, UK

D. Neal
Surgical Oncology, Departments of Oncology and Surgery,
University of Cambridge, Box 279, Addenbrooke's Hospital,
Hills Road, Cambridge CB2 2QQ, UK

J. Donovan
Department of Social Medicine, University of Bristol,
Canyng Hall, Whiteladies Road, Bristol, UK

K. Muir
Warwick Medical School, Warwick University,
Coventry CV4 7AL, UK

molecular dynamic simulation, we showed evidence that this variant has the potential to introduce alterations in the protein or affect RNA splicing. We propose that rs17632542 may directly influence PrCa risk.

Background

Prostate cancer (PrCa) is the most common cancer in men living in the Western world, with a lifetime risk of approximately 1 in 8 for men in Europe and the USA (<http://www.info.cancerresearchuk.org/cancerstats/types/prostate/>). Its aetiology remains poorly understood. The increased risk of PrCa for men with a family history is consistent with variation in genetic susceptibility to the disease (Edwards and Eeles 2004). Genome-wide association studies (GWAS) are a proven way to identify common disease susceptibility alleles. Thus far, such studies have identified SNPs associated with prostate cancer in more than 30 independent regions (Liu et al. 2010). We previously conducted a GWAS in two stages and identified SNPs in seven regions associated with PrCa risk (Eeles et al. 2008). One of these SNPs, rs2735839 on 19q13.3, lies within an LD block of 34 kb, containing the kallikrein genes, *KLK3* and *KLK2* (combined stage 1 and 2 $P = 2.3 \times 10^{-17}$). This SNP was also shown to be associated with serum PSA level; however, an association with PrCa has been demonstrated using samples not ascertained by PSA screening, suggesting that the association is not solely due to selection of cases with elevated PSA and controls with a low PSA level (Ahn et al. 2008).

KLK3 encodes prostate-specific antigen (PSA), a member of the kallikrein family of serine proteases. Its main physiological function is to cleave semenogelins in the seminal coagulum, thereby regulating sperm motility. Like other kallikreins, it might also participate in the process of neoplastic growth and metastasis of PrCa (Lawrence et al. 2010; Paliouras et al. 2007). The expression of PSA is highly restricted to normal and malignant prostate epithelial cells in men. During tumorigenesis, however, serum PSA levels tend to be elevated, probably due to the loss of

normal architecture of the gland. For this reason, PSA is extensively used as a biomarker to screen for and monitor treatment of PrCa.

The association with rs2735839 implies that one or more variants in the *KLK3* region, correlated with rs2735839, are directly associated with PrCa risk. In an attempt to identify such variants, we performed a comprehensive analysis of SNPs across the region, using a combination of imputation and further genotyping.

Materials and methods

Samples

Analyses were based on the samples genotyped in first and second stages of an UK/Australian GWAS, collected as previously described (Eeles et al. 2008, 2009), together with a third stage involving a further 4,901 cases and 4,847 controls. Stage 3 samples were selected from UKGPCS as for stages 1 and 2; from Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), a case–control study based on the region covered by the Eastern UK Cancer Registry and Information Centre (ECRIC); and from the Australian epidemiological studies as in stage 2 (Supplementary Table 1). We also included data from the Cancer Genetic Markers of Susceptibility (CGEMS) study, a GWAS of 1,117 PrCa cases and 1,105 controls drawn from the PLCO study (<http://www.cgems.cancer.gov/>).

Imputation

To evaluate associations with PrCa for SNPs not included in genotyping arrays used in the GWAS, we defined a 34 kb LD region that included all SNPs in the region correlated with the original best hit, rs2735839, at $r^2 > 0.1$ according to the HapMap Phase II CEU dataset. We then identified all SNPs in the interval from the HapMap Phase II as well as SNPs identified through sequencing of European (CEU) samples in the 1000 Genomes study at a frequency of at least 2%. Genotypes for these SNPs in stages 1 and 2 of the UK/Australian GWAS, and the CGEMS study, were then imputed using MACH 1.0 (<http://www.sph.umich.edu/csg/abecasis/MACH/>) and 50 rounds of imputation. We included SNPs with quality score (r^2) > 0.3 in the analysis.

Genotyping

SNP rs17632542 was genotyped for stages 1 and 2 samples by the 5'-endonuclease assay (TaqmanTM) using the ABIPrism 7900HT sequence detection system according to the manufacturer's instructions. Primers and probes were

A. Amin Al Olama · P. D. P. Pharoah · S. Benlloch · D. F. Easton
Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge CB1 8RN, UK

S. J. Chanock
Division of Cancer Epidemiology and Genetics,
National Cancer Institute, 8717 Grovemont Circle,
Bethesda, MD 20892-4605, USA

R. A. Eeles
The Royal Marsden NHS Foundation Trust, Downs Road,
Sutton and London, Surrey SM2 5PT, UK

Table 1 Summary results for 4 SNPs in the KLK3 region at 19q13.3

Marker ^a Alleles ^b Position ^c	Stage	Per allele ^d OR (95%CI)	Het OR ^{d,e} (95%CI)	Hom OR ^{d,f} (95%CI)	P value ^g	
					Stage	Combined
rs17632542	Stage 1	0.35 (0.30–0.42)	0.35 (0.29–0.43)	0.13 (0.06–0.31)	2.6×10^{-29}	
T/C	Stage 2	0.78 (0.68–0.90)	0.79 (0.68–0.91)	0.53 (0.25–1.13)	0.0004	S1 and 2 1.6×10^{-24}
56053569	Stage3	0.65 (0.57–0.74)	0.65 (0.56–0.74)	0.47 (0.23–0.93)	1.2×10^{-10}	S 1, 2 and 3 3.6×10^{-33} S1, 2, 3 and CGEMS 1.9×10^{-34}
	Imputation stage 1				3.1×10^{-25}	
	Imputation stage 2				5.6×10^{-4}	S2 + CGEMS 2.5×10^{-5}
	Imputation CGEMS				0.01	S1, 2 and CGEMS 3.9×10^{-22}
rs2735839	Stage 1	0.56 (0.50–0.64)	0.58 (0.50–0.67)	0.29 (0.19–0.43)	8.2×10^{-20}	
G/A	Stage 2	0.84 (0.77–0.93)	0.83 (0.74–0.92)	0.80 (0.56–1.14)	0.001	S1 and 2 2.3×10^{-17}
56056435	Stage 3	0.80 (0.73–0.88)	0.78 (0.70–0.86)	0.74 (0.54–1.02)	1.1×10^{-6}	S 1, 2 and 3 5.4×10^{-22} S 1, 2, 3 and CGEMS 1.1×10^{-22}
rs266849	Stage 1	0.62 (0.55–0.69)	0.60 (0.52–0.69)	0.42 (0.30–0.58)	1.7×10^{-16}	
A/G	Stage 2	0.95 (0.87–1.03)	0.88 (0.79–0.98)	1.15 (0.89–1.49)	0.23	S1 and 2 2.1×10^{-9}
56040902	Stage3	0.81 (0.74–0.87)	0.81 (0.74–0.89)	0.64 (0.50–0.82)	1.2×10^{-7}	S 1, 2 and 3 1.3×10^{-15} S1, 2, 3 and CGEMS 1.4×10^{-14}
rs1058205	Stage 1	0.59 (0.52–0.66)	0.60 (0.52–0.69)	0.33 (0.24–0.45)	4.7×10^{-20}	
T/C	Stage 2	0.85 (0.78–0.93)	0.82 (0.74–0.91)	0.85 (0.64–1.11)	0.0004	S1 and 2 1.6×10^{-17}
56055210	Stage3	0.81 (0.75–0.88)	0.81 (0.73–0.89)	0.68 (0.53–0.89)	8.2×10^{-7}	S 1, 2 and 3 3.0×10^{-22} S1, 2, 3 and CGEMS 2.8×10^{-23}

rs17632542 is the most strongly associated SNP in this study. rs273539 was the most strongly associated SNP in the original GWAS. rs266849 and rs1058205 also reached $P < 10^{-7}$ in the original GWAS

^a dbSNP rs number, ^b major/minor allele, based on the frequencies in stage 1 controls, ^c build 36 position, ^d OR odds ratio, ^e OR in heterozygotes, relative to common homozygotes, ^f OR in homozygotes, relative to common homozygotes, ^g Cochran–Armitage test for trend

supplied by Applied Biosystems (<http://www.appliedbiosystems.com/>) as Assays-By-Design™. The Stage 3 genotyping was done using an Illumina Golden Gate Assay (<http://www.illumina.com>).

Statistical methods

We assessed associations between each SNP and disease using a 1 *df*.

Cochran–Armitage trend test, stratified by GWAS stage (UK/Australia stage 1, 2 or 3, and CGEMS). For imputed SNPs, we used the corresponding test based on the estimated allele dose. Odds ratios and confidence limits were estimated using unconditional logistic regression, stratified by stage. Tests of homogeneity of the odds ratios across strata were assessed using likelihood ratio tests. To determine independently associated SNPs from imputed data, we identified all SNPs significant at $P < 10^{-7}$ from

Table 2 Stepwise logistic regression analysis of 7 SNPs in this region associated with PrCa risk at $P < 10^{-7}$ (stages 1, 2, 3 and CGEMS)

SNPs in the model	OR (95%CI) Model with all SNPs	<i>P</i> value Model with all SNPs	Final model OR (95% CI) ^a	Final model <i>P</i> value
rs2735839	0.53 (0.06–4.62)	0.57		
rs2569735	1.93 (0.22–16.90)	0.55		
rs1058205	0.87 (0.77–0.99)	0.03		
rs2659122	1.07 (1.00–1.15)	0.07		
rs17632542	0.69 (0.62–0.76)	8.1×10^{-13}	0.66 (0.61–0.72)	2.3×10^{-23}
rs266849	0.95 (0.85–1.06)	0.34	0.91 (0.86–0.96)	0.001
rs2659051	0.96 (0.85–1.08)	0.52		

SNPs retained in the model if significant at $P < 10^{-3}$

We have used the combined data from stages 1, 2, 3 and CGEMS

^a Forward and backward stepwise regression converge to the same model

combined stages 1, 2 and CGEMS and used forward and backward stepwise logistic regression, based on estimated allele dose. SNPs were included in the model if they were significant at $P < 0.001$ after adjustment for other SNPs in the model. For the genotyped SNPs, we identified all SNPs significant at $P < 10^{-7}$ in the stages 1, 2, 3 and CGEMS data combined and used forward and backward stepwise logistic regression.

The associations of SNP genotypes with PSA level were assessed using linear regression, after log transformation of PSA level to reduce skewness. Analyses were performed in R and STATA.

Molecular dynamics modelling the structure of PSA

Potential conformations of the original and mutated crystal structure models were investigated using energy minimisation and force field (Goodford 1985). To gain further insights into the effects of substitution at Ile179 we simulated solvated Ile179 and Ile179Thr. Structure files were generated using Visual Molecular Dynamics (VMD) (Humphrey et al. 1996). The system was equilibrated using Nanoscale Molecular Dynamics (NAMD) (Phillips et al. 2005) and simulated using ACEMD (Giupponi et al. 2008) with the Chemistry at HARvard Molecular Mechanics (CHARMM) (Brooks et al. 2009) force field. Ten independent simulations were carried out for both variants which were then analysed to determine the models' average energies.

Results and discussion

We defined a 34-kb region on 19q13.3 based on HapMap phase II CEU data that included all SNPs correlated with rs2735839, the most strongly associated SNP in the

GWAS, at $r^2 > 0.1$. In this region, we identified 197 SNPs in HapMap2 and 312 SNPs in the 1000 Genomes Project. Genotypes for these SNPs were imputed for subjects in stages 1 and 2 of the UK/Australian GWAS and CGEMS. Based on the combined results, we identified 35 SNPs associated with PrCa at $P < 10^{-7}$. The most statistically significant associated SNPs were rs17632542 and rs62113212 ($P = 3.0 \times 10^{-24}$ and $P = 2.2 \times 10^{-24}$, respectively, Supplementary Table 2). These two SNPs were strongly correlated with one another ($r^2 = 0.99$ in HapMap CEU and $r^2 = 0.75$ in 1000 Genomes, Supplementary Table 3); neither of these SNPs was genotyped in the original GWAS.

To further evaluate these associations, we genotyped rs17632542 directly for the stages 1 and 2 sample set (5,504 cases and 5,834 controls) using a Taqman assay. We also genotyped rs17632542 in a further set of 4,901 PrCa cases and 4,847 controls (unselected for PSA) as part of an Illumina Golden Gate Assay (stage 3). We also genotyped the original associated SNP, rs2735839, and two other SNPs (rs266849 and rs1058205) that reached $P < 10^{-7}$ in the original GWAS stages 1 and 2. The estimated per-allele ORs for rs17632542 were similar to those estimated from the imputed data, and the statistical significance was further increased, stage 1 per allele OR 0.35 (95% CI 0.30–0.42) $P = 2.6 \times 10^{-29}$; stage 2 per allele OR 0.78 (95% CI 0.68–0.90) $P = 4 \times 10^{-4}$; stage 3 per allele OR 0.65 (95% CI 0.57–0.74), $P = 4.2 \times 10^{-10}$; combined $P = 1.9 \times 10^{-34}$ over all stages. The estimated OR in homozygotes for the C allele was lower than in heterozygotes in all stages, in a manner consistent with a log-additive allele dose model (Table 1). As expected, the estimated effect size was larger in stage 1, reflecting the selection of cases for family history of disease and the controls for low PSA (see “Materials and methods”), but the estimated ORs in stages 2 and 3 were similar.

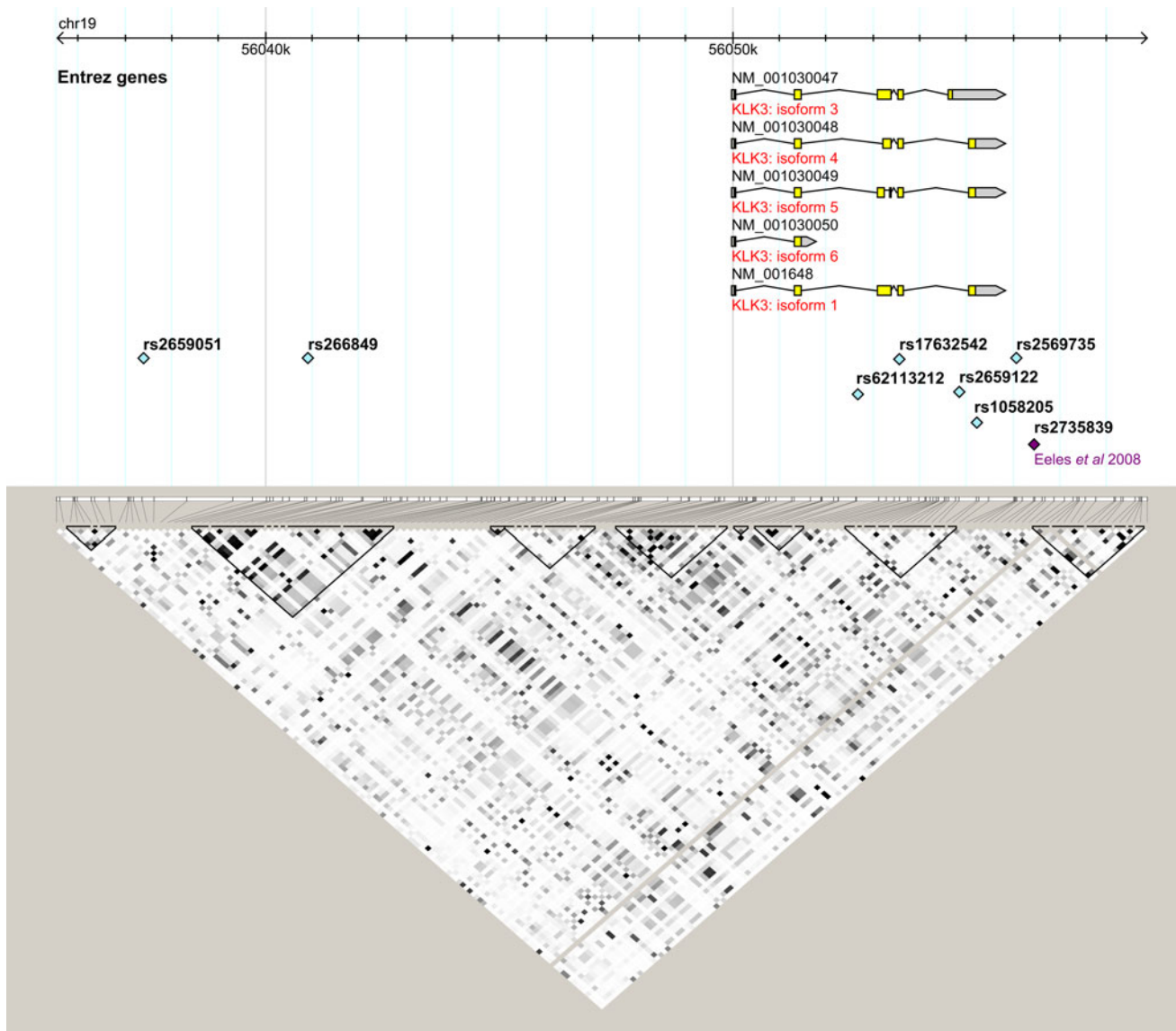


Fig. 1 Regional LD plot and multiple transcripts of the *KLK3* gene at 19q13.3. The newly identified SNPs associated with PrCa risk are labelled using blue diamonds, two SNPs are intragenic in *KLK3*, rs17632542 is a coding variant in exon 4, rs 62113212 is in intron 3

In a multiple logistic regression analysis, rs17632542 remained significant after adjustment for rs2735839 ($P = 8.6 \times 10^{-14}$) based on all stages combined, whereas rs2735839 showed weaker association after adjustment for rs17632542 ($P = 0.02$).

To determine whether there are additional SNPs in the region, other than rs17632542, that are independently associated with PrCa, we used two different analytical approaches. First, we used stepwise logistic regression analyses of 7 SNPs in this region that were genotyped in all stages and were significant at $P < 10^{-7}$. In this analysis, one additional SNP, rs266849, remained in the model at $P < 0.001$ after adjusting for rs17632542 ($P = 0.001$, Table 2) Secondly, we used the imputed data (Stages 1, 2

and CGEMS) and put all 35 SNPs significant at $P < 10^{-7}$ in a logistic regression analysis (Supplementary Table 2). From this analysis, one SNP was significant at $P < 0.001$ after adjusting for rs17632542 (either rs2659052 or rs2569753, as both were correlated with one another, $r^2 = 0.64$). rs266849 is about 12 kb centromeric to rs17632542 between the *KLK15* and *KLK3* genes. rs266849 and rs2659052 were weakly correlated with each other and with rs17632542 (Fig. 1; Supplementary Table 3), suggesting that there are two additional associated variants. In both analyses, however, rs17632542 remained the most significantly associated SNP after adjusting for the other significant SNPs (Table 2; Supplementary Table 2).

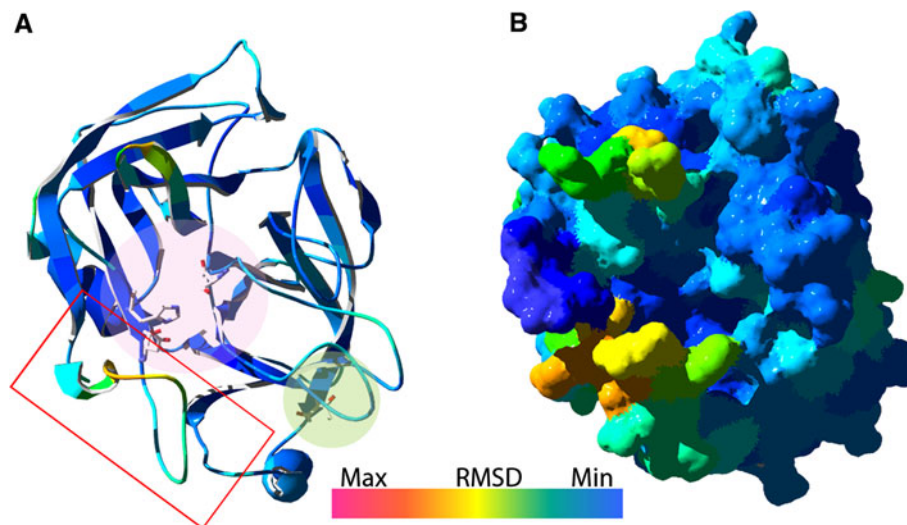


Fig. 2 Comparative molecular dynamic simulation of threonine and isoleucine PSA variants. Structures shown are averaged from 10 independent simulations. **a** Ribbon diagram of I179T variant of PSA coloured according to side chain displacement from the isoleucine variant. Position of the catalytic triad is highlighted by a *pink oval* and T179 substitution by a *green oval*. The kallikrein loop's position is

indicated by a *red rectangle*. **b** Molecular surface view of I179T variant PSA showing solvent accessible surface coloured according to RMS displacement from wild-type (I179) PSA. Maximal displacement is coloured *red*, whilst minimal displacement is *blue* as indicated by the heat map

We previously reported that rs2735839 was associated with serum PSA level (Eeles et al. 2008). Based on the serum PSA levels of the stages 2 and 3 controls, the C allele of rs17632542 was also associated with lower PSA levels, with lowest levels in CC homozygotes ($P = 1.5 \times 10^{-11}$, Supplementary Table 4; Supplementary Fig. 1). In a multiple regression analysis, rs2735839 was marginally significantly associated with PSA levels, after adjusting for rs17632542 ($P = 0.06$). In a recent publication (Gudmundsson et al. 2010) it was also reported that at the *KLK* region rs17632542 was significantly associated with PSA level and prostate cancer risk.

We found evidence for a significant association between rs17632542 and Gleason score, with a stronger association for lower grade disease (based on Gleason score <7 vs. 7+; P -trend = 7.04×10^{-7} and based on Gleason score <8 vs. 8+ P -trend 3.2×10^{-7} , Supplementary Table 5). A similar association has previously been reported for rs2735839 (Kader et al. 2009). We found no association between rs17632542 and family history or age of onset of the disease (data not shown).

Taken together, the above results suggest that observed associations with both PrCa and PSA level are largely driven by rs17632542 or rs62113212. These SNPs are intragenic to *KLK3*: rs17632542 is a non-synonymous SNP in exon 4 resulting in an Ile179Thr substitution, while rs62113212 is intronic and lies in a sequence region with no enhancer or splice effect function (Fig. 1). However, based on the TFSEARCH database, <http://www.cbrc.jp/research/db/TFSEARCH.html>, rs62113212 changes the

scores for the binding of the octamer factor 1 (Oct-1) transcription factor. Although we cannot rule out rs62113212 as the directly associated variant and it remains possible that another correlated variant was missed by the resequencing in the 1000 genomes dataset, rs17632542 appears to be the most plausible functional variant.

Based on the previous structural studies, 3D modeling and X-ray crystallography of PSA and other kallikreins, Ile179 is not part of the active catalytic site (Villoutreix et al. 1994). However, when we carried out multiple molecular dynamic simulations in explicit solvent, we found that the solvated threonine (Thr) variant consistently achieved lower energy levels compared with the isoleucine (Ile) variant ($-3,532 \pm 54$ kcal/mol compared with $-3,053 \pm 41$ kcal/mol), suggesting superior stability in solution. Furthermore, modeling suggested that the Thr variant causes displacement of the kallikrein loop (residues 111–118), although verification of this will require empirical structural determination (Fig. 2).

Because multiple alternative transcripts have been reported for *KLK3* in normal and tumor-derived tissue (Heuze-Vourc'h et al. 2003; Pampalakis et al. 2008; Whitbread et al. 2010), we also assessed the possible effect of the nucleotide substitution at the pre-mRNA level using the Human Splice Finder bioinformatic prediction tool (Desmet et al. 2009). New sites for exonic splice enhancer motifs (SRp40, SRp55) were created in the presence of the rs17632542 minor allele (C) by two algorithms. This putative enhancer site is close to the splicing site for PSA-RP2, a splice variant which results in retention of the intron

between exons 3 and 4. Notably, the same splice variant is found in four other *KLK* genes (Michael et al. 2005) (*KLK1*, *KLK2*, *KLK5* and *KLK15*) that surround *KLK3* at the centromeric end of the locus, suggesting this may be a biologically important splicing event, given the evolutionary conservation. Further molecular studies (EMSA, luciferase assay) would be required to determine if this SNP has an effect at the pre-mRNA level.

The contribution of common genetic variants in *KLK3* to PrCa risk has been a subject of considerable debate because SNPs in this region were also found to be associated with serum levels of PSA. The association seen for rs2735839 in stage 1 of our GWAS could reflect the use of low PSA controls. However, the association was also seen in stage 2 of the GWAS, in the CGEMS study, and in a validation study of our GWAS results in 29 groups in the PRACTICAL (Prostate cancer association group to investigate cancer associated alterations in the genome) consortium in 56,300 samples (unpublished data), the majority of which have not selected controls on the basis of PSA level. There is still some potential for confounding due to opportunistic PSA testing or testing as part of the diagnostic process. The newly identified SNP, rs17632542, exhibits a more significant association with PrCa than the previous tag SNP and studies of this SNP in men who have prostate biopsy as the primary PrCa screen, irrespective of PSA levels, could determine whether this SNP is causally related to PrCa risk independent of PSA. Given its stronger effect on PrCa risk, inclusion of rs17632542 in place of rs2735839 in risk prediction models will improve discrimination in such models.

As this variant is predicted to introduce changes in the stability of the protein and might have a role in pre-mRNA processing of alternative transcripts; it is a good candidate variant for further functional investigation of the underpinnings of the GWAS signal in this region with respect to PrCa risk. Similarly, additional studies are required to determine if this coding variant also influences PSA levels.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Ahn J, Berndt SI, Wacholder S, Kraft P, Kibel AS, Yeager M, Albanes D, Giovannucci E, Stampfer MJ, Virtamo J, Thun MJ, Feigelson HS, Cancel-Tassin G, Cussenot O, Thomas G, Hunter DJ, Fraumeni JF Jr, Hoover RN, Chanock SJ, Hayes RB (2008) Variation in *KLK* genes, prostate-specific antigen and risk of prostate cancer. *Nat Genet* 40:1032–1034
- Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodosek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614
- Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67
- Edwards SM, Eeles RA (2004) Unravelling the genetics of prostate cancer. *Am J Med Genet Part C Semin Med Genet* 129C:65–73
- Eeles RA, Kote-Jarai Z, Giles GG, Al Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, rdern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40:316–321
- Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, Schleutker J, Hamdy FC, Neal DE, Donovan JL, Stanford JL, Ostrander EA, Ingles SA, John EM, Thibodeau SN, Schaid D, Park JY, Spurdle A, Clements J, Dickinson JL, Maier C, Vogel W, D'Ark T, Rebbeck TR, Cooney KA, Cannon-Albright L, Chappuis PO, Hutter P, Zeegers M, Kaneva R, Zhang HW, Lu YJ, Foulkes WD, English DR, Leongamornlert DA, Tymrakiewicz M, Morrison J, rdern-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Saunders EJ, Page EC, Sawyer EJ, Edwards SM, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Southey MC, Lophatananon A, Liu JF, Kolonel LN, Le Marchand L, Wahlfors T, Tammela TL, Auvinen A, Lewis SJ, Cox A, FitzGerald LM, Koopmeiners JS, Karyadi DM, Kwon EM, Stern MC, Corral R, Joshi AD, Shahabi A, McDonnell SK, Sellers TA, Pow-Sang J, Chambers S, Aitken J, Gardiner RAF, Batra J, Kedda MA, Lose F, Polanowski A, Patterson B, Serth JÄ, Meyer A, Luedeke M, Stefflova K, Ray AM, Lange EM, Farnham J, Khan H, Slavov C, Mitkova A, Cao G, Easton DF (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 41:1116–1121
- Giupponi G, Harvey MJ, De FG (2008) The impact of accelerator processors for high-throughput molecular modeling and simulation. *Drug Discov Today* 13:1052–1058
- Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
- Gudmundsson J, Besenbacher S, Sulem P, Gudbjartsson DF, Olafsson I, Arinbjarnarson S, Agnarsson BA, Benediktsdottir KR, Isaksson HJ, Kostic JP, Gudjonsson SA, Stacey SN, Gylfason A, Sigurdsson A, Holm H, Bjornsdottir US, Eyjolfsson GI, Navarrete S, Fuertes F, Garcia-Prats MD, Polo E, Checherita IA, Jinga M, Badea P, Aben KK, Schalken JA, van O, I, Sweep FC, Helfand BT, Davis M, Donovan JL, Hamdy FC, Kristjansson K, Gulcher JR, Masson G, Kong A, Catalona WJ, Mayordomo JI, Geirsson G, Einarsson GV, Barkardottir RB, Jonsson E, Jinga V, Mates D, Kiemenev LA, Neal DE, Thorsteinsdottir U, Rafnar T, Stefansson K (2010) Genetic correction of PSA values using sequence variants associated with PSA levels. *Sci Transl Med* 2:62ra92

- Heuze-Vourc'h N, Leblond V, Courty Y (2003) Complex alternative splicing of the hKLK3 gene coding for the tumor marker PSA (prostate-specific-antigen). *Eur J Biochem* 270:706–714
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38
- Kader AK, Sun J, Isaacs SD, Wiley KE, Yan G, Kim ST, Fedor H, DeMarzo AM, Epstein JI, Walsh PC, Partin AW, Trock B, Zheng SL, Xu J, Isaacs W (2009) Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. *Prostate* 69:1195–1205
- 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
- Liu H, Wang B, Han C (2010) Meta-analysis of genome-wide and replication association studies on prostate cancer. *Prostate*
- Michael IP, Kurlender L, Memari N, Yousef GM, Du D, Grass L, Stephan C, Jung K, Diamandis EP (2005) Intron retention: a common splicing event within the human kallikrein gene family. *Clin Chem* 51:506–515
- Paliouras M, Borgono C, Diamandis EP (2007) Human tissue kallikreins: the cancer biomarker family. *Cancer Lett* 249:61–79
- Pampalakis G, Scorilas A, Sotiropoulou G (2008) Novel splice variants of prostate-specific antigen and applications in diagnosis of prostate cancer. *Clin Biochem* 41:591–597
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
- Villoutreix BO, Getzoff ED, Griffin JH (1994) A structural model for the prostate disease marker, human prostate-specific antigen. *Protein Sci* 3:2033–2044
- Whitbread AK, Veveris-Lowe TL, Dong Y, Tan OL, Gardiner R, Samaratunga HM, Nicol DL, Clements JA (2010) Expression of PSA-RP2, an alternatively spliced variant from the PSA gene, is increased in prostate cancer tissues but the protein is not secreted from prostate cancer cells. *Biol Chem* 391:461–466