A large scale analysis of genetic variants within putative miRNA binding sites in prostate cancer

Running Title: Association of miRSNPs with prostate cancer


1. Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Translational Research Institute, Queensland University of Technology, Brisbane, Australia
2. Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
3. Statistics Unit, QIMR Berghofer Medical Research Institute, Brisbane, Australia
4. Molecular Cancer Epidemiology Laboratory, Genetics and Computational Biology Division, QIMR Berghofer Medical Research Institute, Brisbane, Australia
5. Department of Molecular Oncology, Moffitt Cancer Center, Tampa, FL, USA
6. Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA
7. The Institute of Cancer Research, 15 Cotswold Road, London, SM2 5NG, UK
8. Royal Marsden NHS Foundation Trust, Fulham and Sutton, London and Surrey, UK
9. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK

10. University of Warwick, Coventry, UK

11. Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria, Australia

12. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia

13. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

14. Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA

15. Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

16. Institute of Biomedical Technology/BioMediTech, University of Tampere and FimLab Laboratories, Tampere, Finland

17. Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark

18. Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

19. Surgical Oncology (Uro-Oncology: S4), University of Cambridge, Box 279, Addenbrooke’s Hospital, Hills Road, Cambridge, UK and Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

20. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK

21. Cambridge Institute of Public Health, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 0SR
22. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
23. Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA
24. International Epidemiology Institute, 1455 Research Blvd., Suite 550, Rockville, MD 20850
25. Mayo Clinic, Rochester, Minnesota, USA
26. Department of Urology, University Hospital Ulm, Germany
27. Institute of Human Genetics University Hospital Ulm, Germany
28. Brigham and Women's Hospital/Dana-Farber Cancer Institute, 45 Francis Street- ASB II-3, Boston, MA 02115
29. Washington University, St Louis, Missouri
30. International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
31. Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine
32. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg Germany
33. German Cancer Consortium (DKTK), Heidelberg, Germany
34. Molecular Medicine Center and Department of Medical Chemistry and Biochemistry, Medical University - Sofia, 2 Zdrave St, 1431, Sofia, Bulgaria
35. Department of Genetics, Portuguese Oncology Institute, Porto, Portugal and Biomedical Sciences Institute (ICBAS), Porto University, Porto, Portugal
36. The PRACTICAL Consortium list of authors is provided in Acknowledgements
37. The Australian Prostate Cancer BioResource list of authors is provided in
Acknowledgement

*equal contribution

**Corresponding author:** Jyotsna Batra

**Email:** jyotsna.batra@qut.edu.au

**Fax:** (07) 344 37779

**Address:** Translational Research Institute Pty Ltd
37 Kent Street, Woolloongabba,
Queensland, Australia, 4102

**Conflict of interest**

No author has any competing financial interests or conflicts of interest in respect of this work and its publication

**Word count (excluding references):** 4202

**Total number of figures and tables:** The manuscript consists of 2 tables and 4 figures
Abstract

Prostate cancer is the second most common malignancy among men worldwide. Genome-wide association studies (GWAS) have identified 100 risk variants for prostate cancer, which can explain ~33% of the familial risk of the disease. We hypothesized that a comprehensive analysis of genetic variations found within the 3’ UTR of genes predicted to affect miRNA binding (miRSNPs) can identify additional prostate cancer risk variants. We investigated the association between 2,169 miRSNPs and prostate cancer risk in a large-scale analysis of 22,301 cases and 22,320 controls of European ancestry from 23 participating studies.

Twenty-two miRSNPs were associated (p<2.3x10^{-5}) with risk of prostate cancer, 10 of which were within the 7 genes previously not mapped by GWASs. Further, using miRNA mimics and reporter gene assays, we showed that miR-3162-5p has specific affinity for the KLK3 rs1058205 miRSNP T-allele whilst miR-370 has greater affinity for the VAMP8 rs1010 miRSNP A-allele, validating their functional role.

Significance: Findings from this large association study suggest that a focus on miRSNPs, including functional evaluation, can identify candidate risk loci below currently accepted statistical levels of genome-wide significance. Studies of miRNAs and their interactions with SNPs could provide further insights into the mechanisms of prostate cancer risk.
Introduction

Prostate cancer is the most common non-skin malignancy among men worldwide. In the US, an estimated 233,000 new cases and 29,480 deaths are expected in 2014 (1). Established risk factors for prostate cancer include advancing age, ethnicity, and a family history of the disease (2). Men with a family history of prostate cancer have a 2-fold increased risk of developing the disease and usually with an earlier age of onset (3). A significant role for genetic factors has been confirmed by genome-wide association studies (GWAS) and large scale replication studies, which have already identified 100 single nucleotide polymorphisms (SNP) associated with prostate cancer risk (4, 5). However, the identified SNPs account for only a small proportion of the (33%) excess familial risk suggesting additional SNPs remain to be identified (4).

MicroRNAs (miRNAs) are short ~19 – 24 nucleotide non-coding RNA molecules that post-transcriptionally regulate gene expression by cleaving or degrading mRNA and/or inhibiting its translation (6-8). Most miRNA binding has been observed within the 3’UTR of their target genes, although there are examples of binding within mRNA coding regions (9). As of March 2014, the miRBase database lists >2570 mature miRNAs for humans. miRNAs are expressed in a tissue and cell-specific manner with differential expression profiles in response to disease conditions, with many of these miRNA expression modulations contributing to disease progression (10-15). An impressive effort has been devoted to investigating miRNA dysregulation profiles in prostate cancer. Hence, miRNAs have emerged as not only potential biomarkers for prostate cancer but also as potential therapeutic targets (15-17).

miRNAs negatively regulate their target mRNAs primarily through Watson-Crick base-pairing interactions (18, 19). The most critical region for mRNA binding and repression are miRNA nucleotides 2-8, referred to as the miRNA seed site. Experiments have shown that genetic variations within the seed site or in the target mRNA at sites complementary to
miRNA seed sites, referred to as miRSNPs, may reduce effectiveness or abolish miRNA-mediated repression, having functional consequences for cancer risk (20, 21). For example, Liu et al recently reported that miRSNPs in ITGAv are associated with a decreased risk of prostate cancer (22). In another study assessing 61 putative miRSNPs in a Chinese population, three SNPs were associated with prostate cancer progression whilst four SNPs were associated with prostate cancer-specific mortality (23). However, all these studies have been conducted using small sample sizes and might not be reflective of true positive association.

To further explore the genetic association of miRSNPs and to derive more reliable risk estimates of previously identified prostate cancer risk miRSNPs, we investigated the association between 2,169 miRSNPs and prostate cancer risk and aggressiveness in 23 studies participating in the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) Consortium. This effort included 22,301 cases and 22,320 controls of European ancestry. We then validated the functional role of two prostate cancer risk miRSNPs, Kallikrein 3 (KLK3) rs1058205 (T>C) and Vesicle-associated membrane protein 8 (VAMP-8) rs1010 (A>G), as they were most strongly associated with disease aggressiveness. To our knowledge, this is the first large-scale investigation of the association between miRNA-related gene polymorphisms and prostate cancer risk.

Results

Patient characteristics

The characteristics of the study participants are presented in Table 1. The mean age at diagnosis for cases (64.8 years), was older than the age at interview for controls (60.6 years). Cases (22.1%) were more likely to have a family history of prostate cancer compared to
controls (13.9%). As expected, the majority of cases were diagnosed with tumours with a low (≤ 7) Gleason score (85.5%) that were localized (72.8%) and non-aggressive (82.1%).

Among the cases with data available on vital status, 14.6% died at a median follow up of 5 years with almost half (52.1%) of the deaths attributed to prostate cancer.

**Association of miRSNPs with prostate cancer**

Figure 1 and Supplementary Table 1 show the results of the association analyses for 2,169 putative miRSNPs with prostate cancer risk. Twenty five miRSNPs had a minor allele frequency <0.01 in control samples. A total of 22 SNPs (Table 2, Supplementary Figure 1) representing 16 genes were associated with risk of prostate cancer after correction for multiple testing (\(p<2.3\times10^{-5}\)). The most significant association was observed for rs1058205 located within the *KLK3* 3'UTR with an OR = 0.86 (0.83-0.9), \(p = 1.7\times10^{-14}\). This SNP was previously identified in a fine-mapping study (24). Similarly, *MDM4* rs4245739 was recently reported in the primary iCOGS analysis (4). Ten SNPs - rs2450975, rs3103353, and rs3127593 (*SLC22A2*), rs1567669 (*NKX3-1*), rs1010 (*VAMP8*), rs1810126 (*SLC22A3*), rs2647257 (*TET2*), rs14082 and rs1043853 (*PDLIM5*), and rs17664 (*ITGA6*) were found in the genes/regions previously implicated by prostate cancer GWAS studies. Ten SNPs rs879161 (*PHC3*), rs7615039 (*PHC3*), rs12492606 (*PHC3*), rs311497 (*GMEB2*), rs1530865 (*PDK1*), rs2357637 (*PDK1*), rs12573077 (*ARL3*), rs7402 (*MCAT*), rs47340 (*TTLL12*), and rs4233979 (*TMEM17*) in seven genes (*PHC3*, *GMEB2*, *PDK1*, *ARL3*, *MCAT*, *TTLL12*, *TMEM17*) (Table 2) are at least 20 kb away from the previously reported index GWAS SNPs within the locus. Although *GMEB2* and *ARL3* have been previously reported, these seven genes have not been mapped by previous GWAS as per the NHGRI catalogue on Nov 2014(25).
In secondary analysis, seven SNPs showed significant differences in per-allele odds ratios between aggressive and non-aggressive disease (Supplementary Table 2). The most significant difference was observed for the KLK3 rs1058205 SNP, however this SNP was more strongly associated with non-aggressive disease, which is in line with previous reports on other Kallikrein SNPs (24). Interestingly, only two SNPs, rs1010 in VAMP8 and rs311497 in GMEB2, showed stronger association with aggressive disease. The rs1567669 SNP (NKX3-1) was associated with PSA levels in the patient cohort (Supplementary Table 3). Six SNPs (rs1043853, rs1058205, rs14082, rs2450975, rs3103353, rs3127593) including KLK3 rs1058205 were marginally associated with PSA levels in controls (Supplementary Table 3). Six SNPs showed a trend with respect to age at diagnosis including rs1058205, rs1043853, rs12492606, rs14082, rs7615039 and rs879161 (Supplementary Table 4).

Gene expression and eQTL analysis

Using Oncomine™ (Compendia Bioscience, Ann Arbor, MI) analysis tool, we compared the expression levels of the 16 genes harbouring 22 significant miRSNPs. Using the Grasso dataset (26) of 59 tumour and 28 non tumour samples, expression of 7 genes was found to be deregulated in prostate cancer vs matched benign tissue analysis (Figure 2). We found KLK3 and VAMP8 to be the second and seventh highest deregulated genes within the 16 genes analysed.

We undertook a cis-eQTL analysis of the 22 prostate cancer associated miRSNPs using TCGA data. Three SNPs (rs1058205 (KLK3), rs1530865 and rs2357637 (PDK1)) were not covered by the TCGA genotyping platform (Affymetrix 6.0) and did not have an appropriate proxy SNP identified by SNAP(27), and thus could not be analysed. The remaining 19 SNPs (six index and 13 proxy SNPs identified by SNAP) were assessed for correlation with the respective gene expression harbouring these SNPs. SNP rs2450975 (indexed by rs316000,
$r^2 = 0.95$) in the $SLC22A2$ gene was found to be associated with mRNA transcript expression ($p = 1.76 \times 10^{-5}$), while the $SLC22A2$ SNP, rs10945656 (indexing rs3103353 and rs3127593, $r^2 = 1.0$) and $VAMP8$ rs1010 SNP showed a trend ($p=0.09$) towards genotype-transcript expression (Supplementary Table 5). None of the other SNPs were associated with transcript levels of the gene harbouring the SNP.

**Functional validation of the $KLK3$ rs1058205 and $VAMP8$ rs1010 miRSNPs**

Using a range of computational prediction algorithms, we identified three miRNAs predicted to have differences in binding affinity between the $KLK3$ rs1058205 SNP-alleles. SNPinfo (28) and mirsnpscore (29) predicted miR-219-1-3p to target the T-allele, MicroSNiPer (30) and mirsnpscore predicted miR-3162-5p also to target the T-allele and MicroSNiPer and mirsnpscore predicted miR-4278 to target the C-allele. Two miRNAs were predicted to have differences in binding affinity between the $VAMP8$ rs1010 SNP-alleles. SNPinfo predicted both miR-103 and miR-370-5p to target the A-allele (sense strand = T allele).

Reporter vector assays were then used to test the validity of these *in silico* predicted miRNA binding potential to their target gene/s with specific genotype. For $KLK3$ rs1058205 SNP, miRNA miR-3162-5p induced a ~29% ($p = 0.048$) decrease in luciferase levels for the T-allele compared to the C-allele suggesting that miR-3162-5p has specific affinity for the T allele (Figure 3A). No significant changes were observed for miR-219-1-3p or miR-4278 with either of the alleles for SNP rs1058205 (Supplementary Figure 2A-B). For $VAMP8$ rs1010 (A>G), though miR-370-5p induced a change in luciferase activity for both alleles, the decrease in luciferase levels for the A-allele was ~2 fold ($p = 0.0067$) stronger than for the G-allele (Figure 3B). Although miR-103 was found to regulate VAMP8 expression, it showed comparable results for both alleles (Supplementary Figure 2C).
Expression of miR-3162-5p and miR-370-5p in prostate cancer

Though previously reported in melanoma, breast cancer and cervical cancer (with expression up-regulated in cervical cancer) (31-33), prostatic expression for miR-3162-5p has not been determined. However, the miR-3162-5p gene is located within intron seven of the Oxysterol binding protein gene, which is known to be expressed in the normal and cancerous prostate (31-34). Using qPCR, we specifically confirmed miR-3162-5p expression in a range of cancerous and non-cancerous prostatic cell lines (Supplementary Figure 3A) as well as in patient tissue samples (Supplementary Figure 3B). miR-3162-5p was detected in all the model cell lines and patient samples with varying expression levels. Prostatic expression for miR-370-5p has been reported previously to be up-regulated in cancer (17, 35).

Regulation of KLK3 mRNA and protein levels by miR-3162-5p

Given the importance of KLK3/PSA as a serum biomarker for prostate cancer, we further characterised the miR-3162-5p and KLK3 rs1058205 SNP interaction, to determine if miR-3162-5p was able to affect endogenous KLK3 mRNA and protein levels in cell lines using LNCaP cells homozygous TT for the rs1058205 SNP. With reference to the negative control miRNA mimic, over-expression of miR-3162-5p resulted in a 25% decrease in KLK3 mRNA ($p = 0.016$) as determined using qPCR analysis (Figure 4A).

We then assessed if miR-3162-5p was able to affect endogenous KLK3 protein levels. With reference to the negative control miRNA mimic, over-expression of miR-3162-5p resulted in a 32% decrease in cellular KLK3 protein ($p = 0.007$) as determined using Western blot analysis (Figure 4B-C).

Discussion
In this study we investigated the association between 2,169 putative miRSNPs and prostate cancer risk in a large sample including 22,301 cases and 22,320 controls of European ancestry. We identified 22 SNPs within the 3’ UTR of the 16 genes to be associated with risk of prostate cancer, seven of which although within the previously identified GWAS locus are not previously mapped by the GWAS studies. The most significant association is for the KLK3 rs1058205 SNP previously identified to be associated with prostate cancer risk in a recent study (24).

Seven of these SNPs including KLK3 (rs1058205) and VAMP8 (rs1010), a gene for which little is known about its prostatic function, showed significant differences between aggressive and non-aggressive disease. As expected, these results were not as robust due to the small sample size in the aggressive disease sub cohort. Thus, it was not surprising that, as has been the case for many previously GWAS identified SNPs (36), we could identify only two SNPs to be more significantly associated with aggressive disease.

Using Oncomine, we compared the expression levels of these 16 genes in a dataset consisting of 59 tumor and 28 non-tumor samples, revealing the expression of seven of these genes to be deregulated in prostate cancer. KLK3 was the second most deregulated gene. We then chose to validate the functional role of the KLK3 rs1058205 SNP, it being the most significant miRSNP identified. We demonstrated that miR-3162-5p has specific affinity for the KLK3 rs1058205 T-allele. Our results support the emerging "miRNA network" that contributes to prostate cancer by regulating kallikrein and non-kallikrein genes (37).

As prostatic expression for miR-3162-5p had not been determined previously, we confirmed its expression in a range of cancerous and non-cancerous prostatic cell lines and in patient tissue samples. It is interesting to note that prostate cancer PC3 cell lines with the highest miR-3162-5p expression do not produce any endogenous KLK3. In previous studies this differential expression has been attributed to absence of an androgen receptor in the PC3 cell
lines since KLK3 expression is androgen dependent in other prostate cancer cell lines. However, its regulation by miR-3162-5p could be an alternative regulatory mechanism and rationale for no KLK3 expression in these cells.

Kallikrein-3 (KLK3) is also referred to as Prostate-specific antigen (PSA), as it is expressed at orders of magnitude higher in the prostate compared to other tissues (38). As KLK3 serum levels are often elevated in prostate cancer, largely due to leakage associated with a loss of tissue architecture, KLK3 is thus utilized as the major serum biomarker for this disease (38, 39). Given the potential importance of KLK3 for prostate cancer diagnosis we further characterized the miRNA-\(KLK3\) rs1058205 interaction demonstrating that miR-3162-5p is able to cause a decrease in \(KLK3\) mRNA and KLK3 protein expression in LNCaP cells homozygous for the T-allele. Interestingly, the \(KLK3\) rs1058205 SNP was associated with PSA levels in the control population (Supplementary Table 3), which may reflect some effects of strong regulatory factors (such as miRNA) exerting genotype-specific effects for this locus. The rs1058205 SNP, in addition to other PSA associated SNPs, may therefore have implications for PSA-based diagnoses, hence requiring adjustments to PSA ranges for specific genotype. Furthermore, as KLK3 belongs to a family of 15 homologous genes, it would be important to consider potential additional effects of miR-3162-5p miRNA on other kallikrein and non-kallikrein targets in future studies.

The \(KLK3\) rs1058205 T allele was previously shown to be associated with increased prostate cancer risk (24). Here we demonstrate that decreased KLK3 expression induced by miR-3162-5p targeting of the T-allele represents a mechanism by which the rs1058205 T-allele may be associated with increased prostate cancer risk. Interestingly, it has been shown that more aggressive prostate tumors have lower tissue levels of KLK3 (40). One mode via which KLK3 may act in a protective capacity in prostate cancer is through inhibition of angiogenesis (41, 42). Although the full mechanism is unclear, the anti-angiogenic effect of
KLK3 has been attributed to its proteolytic function on various angiogenic and anti-angiogenic proteins. (43, 44). However, due to its additional proteolytic function and its subsequent potential to target components of cell-cell adhesion and the extra cellular matrix, high levels of KLK3 have also been proposed as a risk for prostate cancer. It is therefore possible that the effects of KLK3 on tumor development are stage-specific, with low KLK3 contributing to increased localised tumor growth (as observed in genetic risk analysis) whilst high KLK3 poses a risk at later metastatic stages of tumor development.

The $VAMP8$ rs1010 SNP was also selected for functional validation due to its significant association with aggressive prostate cancer, where miR-370-5p was found to have greater affinity for the $VAMP8$ rs1010 A-allele versus the minor G-allele. Interestingly, prostatic expression for miR-370-5p has been reported previously to be up-regulated in cancer (17, 35).

To our knowledge, this is the first study to report an association and mechanism of action between a $VAMP8$ miRSNP and prostate cancer risk. VAMP8 is an integral membrane protein that is involved in the fusion of synaptic vesicles with the pre-synaptic membrane. It also plays a complex role in the control of granule secretion, transport vesicle trafficking, phagocytosis and endocytosis (45, 46). Loss of VAMP8 has been shown to affect glucose metabolism, energy expenditure and insulin sensitivity in mice (47). Though a direct role of VAMP8 in cancer is unknown, its ability to influence glucose metabolism and energy expenditure makes it a potential candidate in carcinogenesis, in relation to the shift in cellular metabolism from oxidative phosphorylation to glycolysis (the Warburg effect) that occurs in cells undergoing malignant transformation (47, 48). Hence, the role of $VAMP8$ may be important for prostate cancer. The $VAMP8$ rs1010 SNP was previously associated with risk for early onset myocardial infarction (49) and is in high LD ($r^2=0.98$) with an intergenic SNP (rs10187424) identified in a previous GWAS for prostate cancer risk (OR=0.92, 95%
CI=0.89-0.94, \( P=2.1 \times 10^{-9} \) (36). No functional relevance has been assigned to rs10187424 SNP. In the current study, we demonstrated that miR-370-5p has greater affinity for the VAMP8 rs1010 A-allele, thus identifying the likely causal variant behind the GWAS marker SNP. Nevertheless, the possibility of another functional variant in LD with rs1010 or any other putative functional miRSNPs (including KLK3) identified in our study cannot be ruled out. Larger sample sizes are now required to provide additional power to assess true independence and/or the effect of these SNPs as modifiers of the unknown functional variants and/or top risk GWAS SNPs using conditional regression and/or haplotype analysis.

Furthermore, it should be noted that the size effects of these variants are very small though comparable to previous GWAS studies. Thus, once independent causal variants or haplotypes at each of the known GWAS loci are identified, it would be interesting to undertake risk score calculations to assess the additive effects of all GWAS identified SNPs including the miRSNPs identified in our study.

Although our analysis has identified several miRSNPs previously not reported by the GWAS analysis, functional validation of these variants is required. Our eQTL analysis did not yield any significant results for genotype - mRNA expression correlation except for a SNP in SLC22A2, which is not surprising given that the miRNA machinery might not affect the mRNA levels in situ but will only inhibit the translation of these genes. Future studies are warranted to correlate genotypes with protein expression using immunohistochemistry and/or western blot analysis. Additional functional studies may further clarify the role of these novel miRSNPs in prostate cancer aetiology.

In conclusion, our study has identified putative functional SNPs associated with prostate cancer risk in several genes that further show differential expression in tumor vs normal tissue from prostate cancer patients. The functional validation for the rs1058205 and rs1010
miRSNPs herein provides increasing evidence that miRSNPs may be associated with prostate cancer risk.

Materials and Methods

Study populations

The Collaborative Oncological Gene-environment Study (COGS) is a large collaborative effort among different consortia, including PRACTICAL, to evaluate genetic variants for associations with the risk of prostate, ovarian and breast cancers. Details of the study have been reported previously (4). Briefly, 32 studies participating in the PRACTICAL consortium contributed samples from 25,074 prostate cancer cases and 24,272 controls to COGS. The majority of studies were nested, population-based or hospital-based case-control studies. Individuals were excluded from the study based on strict quality control criteria including: overall genotype call rate <95%, genotypically non-European origin, samples that were XX or XXY and therefore not genotypically males or samples not concordant with previous genotyping within PRACTICAL. The present analysis included 44,621 samples (22,301 cases and 22,320 controls) of European ancestry. Demographic and clinical information on study participants including age at diagnosis, Gleason score, stage of disease, prostate-specific antigen (PSA) and cause of death were obtained through in-person interviews or medical or death records. Aggressive disease was defined as Gleason score ≥ 8, prostate-specific antigen (PSA) >100 ng/ml, disease stage of ‘distant’ (outside the pelvis) or prostate cancer associated death. Study was approved by each institutional review board (IRB) and informed consent was obtained from each participant. Patient studies were conducted in accordance with the Declaration of Helsinki.

miRSNP selection and genotyping
A total of 2,169 miRSNPs within the 3' UTRs of the cancer associated genes were selected for genotyping. A SNP was selected if differential miRNA binding potential for the alternative alleles was predicted by at least two of four algorithms: 1. Mirsnpscore (29); 2. Miranda and 3. Sanger (both available through SNPinfo) (28); and 4. MicroSNiPer (30). Genotyping was performed using a custom Illumina Infinium array that included 211,115 SNPs (the iCOGS chip) (4). Genotypes were called using Illumina’s proprietary GenCall algorithm. SNPs were excluded from further analysis if the call rate was <95%, deviated from Hardy-Weinberg Equilibrium (HWE) in controls at $P < 10^{-7}$, or if genotypes were discrepant in more than 2% of duplicate samples.

**cis-eQTL analysis**

For each index miRSNP, we retrieved all the correlated ($r^2 \geq 0.8$) variants in EUR populations from 1000 Genomes using SNAP (27). The pre-processed (Level 2) germline genotypes of the index or correlated SNPs were downloaded from the TCGA data portal and the expression levels of genes harbouring these SNPs were obtained via the cBio Portal for Cancer Genomics. Using standard QC analysis, 6 samples were removed either due to discordant sex information (X-chromosome homozygosity rate between 0.2 and 0.8) or due to heterozygosity rate >3 standard deviations from the mean. An additional 45 individuals were removed due to ethnic heterogeneity as calculated using principal component analysis. Data from 178 Caucasian individuals was used for the final genotype expression correlation analysis by Kruskal-Wallis test using IBM SPSS Statistics (version 22).

**miRNA target reporter vector assays**

To assess validity of *in silico* predictions for miRNA-mRNA affinity, miRNA target luciferase reporter vector assays were performed. Reporter vectors were constructed for the
major and minor SNP allele variants for both \textit{KLK3} and \textit{VAMP-8} using the pmirGLO Dual-Luciferase miRNA Target Expression Vector (Promega) for \textit{KLK3} constructs, and the pMIR REPORT vector (Ambion) for \textit{VAMP-8} constructs. (Portions of \textit{KLK3} and \textit{VAMP-8} pertaining to predicted miRNA binding regions were synthesised by Integrated DNA Technologies). LNCaP cells were co-transfected with vector and mirVana miRNA Mimpics (Life Technologies) using FuGENE transfection reagent (Promega) then analysed 24 hours later using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer’s instructions. Luciferase levels were normalised against Renilla co-expressed from the same vector or against the β-galactosidase derived from co-expression with a second vector measured using Galacto-Light (Tropix) for VAMP8. For KLK3, on the day of transfection cells were cultured in 1\% charcoal stripped serum for the remainder of the experiment to induce lower levels of endogenous \textit{KLK3} to minimise the impact of miRNA mimic-endogenous \textit{KLK3} binding on reporter vector assay sensitivity. A negative control mirVana miRNA Mimic - Negative Control #1 (Life Technologies) was used for analysis alongside candidate miRNAs. A single experiment consisted of each miRNA / vector treatment cultured in triplicate. Three independent experiments were conducted in total.

RT-qPCR analysis to assess miRNA expression

miRNA was extracted as total RNA from cell lines (LNCaP, LAPC4, DUCAP, DU145, PC3, 22Rv1, RWPE1 and BPH1 obtained from ATCC unless otherwise stated (see Acknowledgements) using TRIzol reagent (Life Technologies). Cell line authentication (STR profiling) was performed by either the Queensland Institute of Medical Research (Brisbane, Queensland, Australia) or DDC Medical (Fairfield, Ohio, U.S.A.). Formalin fixed and paraffin-embedded (FFPE) blocks from prostate tumours and their adjacent non-cancer prostate were obtained from the Australian Prostate Cancer Bio-
Resource tumour bank. Tissue blocks containing the tumour cells were serially sectioned (20 μm sections) and transferred to glass slides. Slides were stained with methyl green and the tumour areas were marked by a pathologist. Marked areas were then manually micro-dissected under a microscope using a sterile injection needle (size 0.65 × 25 mm). RNA was extracted using the miRNeasy FFPE kit (Qiagen, Chadstone, Australia).

To assess expression, reverse transcription and qPCR was performed using the TaqMan MicroRNA Reverse Transcription Kit and TaqMan MicroRNA Assays (Life Technologies). The small nuclear RNA - RNU24 was used as an endogenous quantitative normalization control (16). Relative expression levels were calculated using the Comparative Ct method. As TaqMan MicroRNA Assays were not commercially available for miR-3162-5p, we designed in-house assays for miR-3162-5p (Supplementary Methods 1) following the TaqMan methodology (50).

RT-qPCR and Western blot analysis to assess KLK3 mRNA and protein expression

LNCaP cells were plated at 150 000 cells per well on a 6 well plate overnight, then cultured in 1% charcoal stripped serum (for the remaining duration of the experiment) and transiently transfected with 30 nM of mirVana miRNA Mimics using Lipofectamine RNAiMAX transfection reagent (Life Technologies) followed by a treatment with 10 nM dihydrotestosterone (DHT) to stimulate KLK3 expression via the androgen receptor pathway. Total RNA was isolated after 24 hours incubation post DHT stimulation using the RNAeasy Mini Kit (Qiagen) and assessed for quality and yield using a Nanodrop ND-1000 spectrophotometer. RNA was reversed transcribed using oligo dT primers and 500 ng of total RNA. qPCR for KLK3 was then performed using the SYBR Green PCR Master Mix (Life Technologies) for each sample in triplicate with Beta-actin used as an endogenous quantitative normalisation control. Relative expression levels were calculated using the
Comparative Ct method. Primers were synthesised by Integrated DNA Technologies. Primers sequences for *KLK3* were: forward primer 5’- agtgcgagaagcattcccaacc -3’, reverse primer 5’-ccagcaagatcagctttgtcct -3’. Primers sequences for Beta-actin were: forward primer 5’-gcgttacaccctttcttgacaaaacct -3’, reverse primer 5’- gctgtcaccttcaccgttcca -3’.

Total protein was isolated using SDS lysis buffer (1% SDS, 5% glycerol, 10 nM Tris, Roche Complete protease inhibitor); concentration was assessed via the BCA method and 10 ug of total protein was run using standard techniques on a 12% resolving poly-acrylamide gel.

Western blotting was performed using standard techniques with primary antibodies including Rb anti KLK3 (Dako - A0562) and Rb anti Beta-actin (Abcam - ab25894). Western blots were imaged on an Odyssey Imaging System (LI-COR Biosciences) using fluorescently labelled secondary antibodies (Alexa Fluor 680 & 790 - Invitrogen) with protein band intensities analysed via densitometry using Odyssey Imaging System software.

**Statistical analysis**

Demographic, clinical and mortality information was summarised by mean (SD) and number (%). Ethnic groups were defined based on a subset of 37,000 uncorrelated markers that passed quality control (including ~1,000 selected as ancestry informative markers). The COGS data was combined with the three Hapmap2 populations and multi-dimensional scaling was used to identify and exclude ethnic outliers (4). After exclusion of ethnic outliers, principal component analyses were carried out for Europeans. The first six principal components were used to control for population substructure as additional principal components did not reduce inflation further (4). Associations between individual SNPs and prostate cancer risk or aggressive disease were evaluated using logistic regression models to estimate per minor allele odds ratios (OR) and 95% confidence intervals (95% CI).

Associations between individual SNPs and prostate cancer risk were also evaluated in a
similar manner for different age categories. The associations between SNP genotypes and PSA level were assessed using linear regression, after log-transformation of PSA level to correct for skewness. Analyses were performed using SPSS and R. All models included study site and principal components as covariates.

Unless otherwise stated, for all other analyses three independent experiments were conducted with results presented as mean +/- standard deviation, and analysed using a Student’s t test with a p-value of <0.05 considered statistically significant for the functional studies.

Acknowledgements

J. Batra is supported by an NHMRC Career Development Fellowship. A.B. Spurdle is an NHMRC Senior Research Fellow; J.A. Clements is an NHMRC Principal Research Fellow. This work is supported by Cure Cancer Australia Foundation and Cancer Australia PdCCRS grant 1068321 and NHMRC grant 1050742 (PI: Batra), DOD grant W81XWH-12-1-0113(PI: Park) and NCI R01CA128813 (PI: Park). E.K. Amankwah was supported by a Cancer prevention fellowship from the NCI (R25T CA147832) during the conduct of this study. The APCB funding was obtained from the National Health and Medical Research Council (Enabling Grant 614296) and an infrastructure grant from the Prostate Cancer Foundation of Australia. The authors would also like to thank John Lai, Farhana Matin and Leire Moya for technical support and to Charles Sawyers for providing the prostate cancer cell line LAPC4 and Matthias Nees for providing the prostate cancer cell line DuCAP. Additionally, the results published here are based partly on data generated by The Cancer Genome Atlas (TCGA), established by the National Cancer Institute and the National Human Genome Research Institute, and we are grateful to the specimen donors and relevant research groups associated with this project.

Funding for the CRUK study and PRACTICAL consortium:
This work was supported by the Canadian Institutes of Health Research,
European Commission's Seventh Framework Programme grant agreement n° 223175
(HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118,
C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH)
Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON
initiative). We acknowledge support from the NIHR to the Biomedical Research Centre at
The Institute of Cancer Research and Royal Marsden NHS Foundation Trust.

COGS acknowledgement:
This study would not have been possible without the contributions of the following: Per Hall
(COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin
Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin
Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench,
Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis,
Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre
for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of
the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel
Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University
and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G.
Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham,
Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic
Genotyping Core Facility.

Funding for the iCOGS infrastructure came from: the European Community's Seventh
Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175)
(COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174,
C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of
Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

The PRACTICAL Consortium- In addition to those named in the author list

<table>
<thead>
<tr>
<th>Authors</th>
<th>Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret Cook</td>
<td>PRACTICAL coordination</td>
</tr>
<tr>
<td>John Pedersen, John L Hopper and Melissa C Southey</td>
<td>MCCS</td>
</tr>
<tr>
<td>Ami Karlsson, Carin Cavalli-Bjoerkman, Jan-Erik Johansson, Jan Adolfson, Markus Aly, Michael Broms, Paer Stattin</td>
<td>CAPS &amp; STHM1</td>
</tr>
<tr>
<td>Brian E. Henderson, Fredrick Schumacher</td>
<td>MEC</td>
</tr>
<tr>
<td>Andreas Roder, Peter Iversen, Peter Klarskov, Sune F. Nielsen, Maren Weischer</td>
<td>CPCS1 &amp; CPCS2</td>
</tr>
<tr>
<td>Tim J. Key, Hans Wallinder, Sven Gustafsson</td>
<td>EPIC</td>
</tr>
<tr>
<td>Jenny L. Donovan, Freddie Hamdy, Anne George, Athene Lane, Gemma Marsden, Michael Davis, Paul Brown</td>
<td>ProtecT &amp; ProMPT</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Nora Pashayan</td>
<td>SEARCH</td>
</tr>
<tr>
<td>Sarah Holt</td>
<td>FHCRC</td>
</tr>
<tr>
<td>Lisa B. Signorello, Wei Zheng</td>
<td>SCCS</td>
</tr>
<tr>
<td>Liang Wang, Lori Tillmans, Shaun Riska</td>
<td>MAYO</td>
</tr>
<tr>
<td>Antje Rinckleb, Kathleen Herkommer, Manuel Luedeke, Walther Vogel Poland Dominika Wokozorczyk, Jan Lubiski, Wojciech Kluzniak</td>
<td>ULM</td>
</tr>
<tr>
<td>Aida k. Dieffenbach, Christa Stegmaier, Volker Arndt</td>
<td>ESTHER</td>
</tr>
<tr>
<td>Babu Zachariah, Hyun Park, Julio Pow-Sang, Maria Rincon, Selina Radlein,</td>
<td>MOFFITT</td>
</tr>
<tr>
<td>Aleksandrina Vlahova, Atanaska Mitkova, Chavdar Slavov, Darina Kachakova, Elenko Popov, Svetlana Christova, Tihomir Dikov, Vanio Mitev</td>
<td>PCMUS</td>
</tr>
<tr>
<td>Felicity Lose, APCB*</td>
<td>QLD</td>
</tr>
<tr>
<td>Joana Santos, Joao Barros-Silva, Paula Paulo, Pedro Pinto, Rui Henrique, Sofia Maia</td>
<td>IPO-Porto</td>
</tr>
</tbody>
</table>

The Australian Prostate Cancer BioResource is contributed by following author other than those named in the author list

Prof Gail Risbridger$^{1,2}$, Dr Renea Taylor$^{1,2}$, Prof Wayne Tilley$^{3,4}$, A/Prof Lisa Butler$^{3,4}$, A/Prof Lisa Horvath$^{5,6,7}$, Dr Trina Yeadon$^{8,9,10}$, Ms Allison Eckert$^{8,9,10}$, Dr Glen Wood$^{11}$, Dr Peter Heathcote$^{11}$, Dr Greg Malone$^{11}$, Dr Kris Kerr$^{12}$, Dr Megan Turner$^{12}$, Dr Angus Collins$^{12}$


Author Contributions


References


46. Behrendorff N, Dolai S, Hong W, Gaisano HY, Thorn P. Vesicle-associated membrane protein 8 (VAMP8) is a SNARE (soluble N-ethylmaleimide-sensitive factor
attachment protein receptor) selectively required for sequential granule-to-granule fusion.

The Journal of biological chemistry. 2011;286:29627-34.


### Tables

#### Table 1 Participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Controls N=22,320</th>
<th>Cases N=22,301</th>
<th>p value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at diagnosis/interview mean± SD</strong></td>
<td>60.6±10.7</td>
<td>64.8± 8.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Family history of prostate cancer</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>10992(86.1)</td>
<td>10300(77.9)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1779(13.9)</td>
<td>2918(22.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Gleason score</strong></td>
<td>N/A</td>
<td>8863(52.6)</td>
<td>5548(32.9)</td>
</tr>
<tr>
<td>2-6</td>
<td>8863(52.6)</td>
<td>5548(32.9)</td>
<td>2437(14.5)</td>
</tr>
<tr>
<td>7</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td>8-10</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td><strong>SEER Stage</strong></td>
<td>N/A</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
</tr>
<tr>
<td>Local</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td>Distant</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td>Regional</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>13246(72.8)</td>
<td>883(4.9)</td>
<td>3555(19.6)</td>
</tr>
<tr>
<td><strong>PSA at diagnosis (ng/ml)</strong></td>
<td>N/A</td>
<td>12692(95.7)</td>
<td>565(4.3)</td>
</tr>
<tr>
<td>&lt;100</td>
<td>12692(95.7)</td>
<td>565(4.3)</td>
<td></td>
</tr>
<tr>
<td>≥100</td>
<td>12692(95.7)</td>
<td>565(4.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Aggressive disease&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td>N/A</td>
<td>17504(82.1)</td>
<td>3812(17.9)</td>
</tr>
<tr>
<td>No</td>
<td>17504(82.1)</td>
<td>3812(17.9)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>17504(82.1)</td>
<td>3812(17.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Vital status</strong></td>
<td>N/A</td>
<td>13794(85.4)</td>
<td>1233(7.6)</td>
</tr>
<tr>
<td>Alive</td>
<td>13794(85.4)</td>
<td>1233(7.6)</td>
<td>1134(7.0)</td>
</tr>
<tr>
<td>Prostate-specific death</td>
<td>0</td>
<td>1233(7.6)</td>
<td>1134(7.0)</td>
</tr>
<tr>
<td>Other death</td>
<td>1822(27.8)</td>
<td>1134(7.0)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> t-test for a continuous variable and chi-square test for a categorical variable

<sup>b</sup> Aggressive disease is defined as a Gleason score of 8-10, PSA at diagnosis≥100 ng/ml, distant stage or prostate cancer-specific death.
Table 2 Risk estimates and predicted miRNAs for the 22 putative miRSNPs associated with prostate cancer risk

<table>
<thead>
<tr>
<th>miRSNP</th>
<th>Gene</th>
<th>Transcript variant</th>
<th>Predicted miRNA(s)</th>
<th>chr</th>
<th>Position</th>
<th>Risk OR (95% CI)</th>
<th>Risk p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4245739</td>
<td>MDM4</td>
<td>NM_002393, NM_001204171, NM_001204172</td>
<td>miR-191-5p, miR-887, miR-3669</td>
<td>1</td>
<td>204518842</td>
<td>0.92 (0.89-0.95)</td>
<td>7.81E-08</td>
</tr>
<tr>
<td>rs4233979</td>
<td>TMEM17</td>
<td>NM_198276</td>
<td>miR-299-5p</td>
<td>2</td>
<td>62727902</td>
<td>1.08 (1.04-1.12)</td>
<td>1.61E-05</td>
</tr>
<tr>
<td>rs1010</td>
<td>VAMP8</td>
<td>NM_003761</td>
<td>miR-370, miR-103</td>
<td>2</td>
<td>85808982</td>
<td>0.91 (0.89-0.94)</td>
<td>3.79E-11</td>
</tr>
<tr>
<td>rs17664</td>
<td>ITGA6</td>
<td>NM_00079818, NM_0020210</td>
<td>miR-548c-3p, miR-548aj, miR-4691-3p</td>
<td>2</td>
<td>173369231</td>
<td>0.93 (0.91-0.96)</td>
<td>5.51E-07</td>
</tr>
<tr>
<td>rs1530865</td>
<td>PDK1</td>
<td>NM_002610</td>
<td>miR-877-5p, miR-3125, miR-3916</td>
<td>2</td>
<td>173461090</td>
<td>0.8 (0.75-0.86)</td>
<td>2.28E-09</td>
</tr>
<tr>
<td>rs2357637</td>
<td>PDK1</td>
<td>NM_002610</td>
<td>miR-3916, miR-3125, miR-877-5p</td>
<td>2</td>
<td>173463138</td>
<td>0.81 (0.76-0.88)</td>
<td>3.66E-08</td>
</tr>
<tr>
<td>rs7615039</td>
<td>PHC3</td>
<td>NM_024947</td>
<td>miR-208a, miR-208b</td>
<td>3</td>
<td>169806170</td>
<td>0.86 (0.83-0.9)</td>
<td>8.72E-12</td>
</tr>
<tr>
<td>rs12492606</td>
<td>PHC3</td>
<td>NM_024947</td>
<td>miR-939, miR-362-5p</td>
<td>3</td>
<td>169808354</td>
<td>0.86 (0.83-0.9)</td>
<td>1.23E-11</td>
</tr>
<tr>
<td>rs879161</td>
<td>PHC3</td>
<td>NM_024947</td>
<td>miR-27a-5p, miR-220c, miR-3158-3p</td>
<td>3</td>
<td>169812115</td>
<td>0.86 (0.83-0.9)</td>
<td>6.57E-12</td>
</tr>
<tr>
<td>rs14082</td>
<td>PDLIM5</td>
<td>NM_006457, NM_001011513, NM_001256425, NM_001256426, NM_001256428</td>
<td>miR-128, miR-494</td>
<td>4</td>
<td>95586224</td>
<td>1.09 (1.06-1.12)</td>
<td>1.27E-09</td>
</tr>
<tr>
<td>rs1043853</td>
<td>PDLIM5</td>
<td>NM_006457, NM_001011513, NM_001256425, NM_001256426, NM_001256427, NM_001256428</td>
<td>miR-567, miR-3120, miR-4310</td>
<td>4</td>
<td>95588274</td>
<td>1.08 (1.05-1.11)</td>
<td>9.65E-09</td>
</tr>
<tr>
<td>rs2647257</td>
<td>TET2</td>
<td>NM_001127208</td>
<td>miR-301a, miR-301b, miR-4330</td>
<td>4</td>
<td>106199505</td>
<td>1.1 (1.06-1.13)</td>
<td>2.05E-10</td>
</tr>
<tr>
<td>rs2450975</td>
<td>SLC22A2</td>
<td>NM_003058, NM_003058</td>
<td>miR-412, miR-4282</td>
<td>6</td>
<td>160637975</td>
<td>1.12 (1.09-1.16)</td>
<td>9.36E-13</td>
</tr>
<tr>
<td>SNP</td>
<td>Gene</td>
<td>miRNAs</td>
<td>Count</td>
<td>Log2 Odds</td>
<td>p-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------</td>
<td>-----------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3127593</td>
<td>SLC22A2</td>
<td>miR-200a, miR-302a, miR-488-3p</td>
<td>6</td>
<td>1.12</td>
<td>2.22E-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3103353</td>
<td>SLC22A2</td>
<td>miR-942, miR-4268</td>
<td>6</td>
<td>1.12</td>
<td>1.75E-08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1810126</td>
<td>SLC22A3</td>
<td>miR-1205, miR-124-3p, miR-216b</td>
<td>6</td>
<td>1.1</td>
<td>2.81E-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1567669</td>
<td>NKX3-1</td>
<td>miR-637, miR-1275, miR-625</td>
<td>8</td>
<td>1.11</td>
<td>3.15E-13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1257307</td>
<td>ARL3</td>
<td>miR-432, miR-1258, miR-1224-5p</td>
<td>10</td>
<td>0.93</td>
<td>6.43E-06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1058205</td>
<td>KLK3</td>
<td>miR-3162-5p, miR-219-1-3p, miR-4278</td>
<td>19</td>
<td>0.87</td>
<td>1.73E-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs311497</td>
<td>GMEB2</td>
<td>miR-26c, miR-492, miR-619, miR-4648</td>
<td>20</td>
<td>0.92</td>
<td>1.28E-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7402</td>
<td>MCAT</td>
<td>miR-616-3p</td>
<td>22</td>
<td>0.93</td>
<td>4.41E-06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs47340</td>
<td>TTLL12</td>
<td>let-7f, let-7g, let-7i, miR-103, miR-107, miR-764</td>
<td>22</td>
<td>1.07</td>
<td>5.79E-06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a SNP previously identified in prostate cancer GWAS and fine-mapping studies.

b Other SNPs within or around the genes for the listed SNPs have been identified in previous GWAS studies.

c miRNAs were predicted using four algorithms. (See Materials & Methods).

d adjusted for 6 principal components and study group.

e miRSNPs are present within the 3'UTRs of the most the of the common splice variants of the miRNA target genes.
Figure Titles and Legends

Figure 1 miRSNP association with prostate cancer risk- Manhattan Plot with –log \( p \)-values adjusted for study site and principal components. 2,169 miRSNPs were assessed for association with prostate cancer risk. 22 SNPs representing 16 genes were found to be associated with risk of prostate cancer after correction for multiple testing (\( p<2.3 \times 10^{-5} \)). (The 10 most significant SNPs are labelled). KLK3 rs1058205 and VAMP8 rs1010 (in bold) were selected for further functional validation studies. (OR = odds ratio).

Figure 2 Expression levels of 16 genes harbouring 22 significant miRSNPs in cancerous and normal tissue from prostate cancer patients- Oncomine analysis of the Grasso dataset (26) of 59 tumor and 28 non-tumor (normal) samples, shows the expression of 7 genes (PDLIM5 – VAMP8) to be deregulated (\( p<0.05 \)) in prostate cancer.

Figure 3 miR-3162-5p directly targets the KLK3 rs1058205 SNP T allele and miR-370-5p targets the VAMP8 rs1010 SNP A allele with greater affinity - Following over-expression with miR-3162-5p, reporter vector assays demonstrated a ~29% decrease in luciferase levels (\( p = 0.048 \)) for the KLK3 rs1058205 SNP T-allele compared to the C-allele (A). Over-expression of miR-370-5p resulted in a change in luciferase activity for both VAMP8 rs1010 SNP variants with the decrease for the A-allele ~2 fold (\( p = 0.0067 \)) stronger than for the G-allele (B). Mean +/- SD, n = 3. (* = \( p < 0.05 \)) (** = \( p < 0.01 \))

Figure 4 miR-3162-5p induces a reduction in KLK3 mRNA and KLK3 protein expression in LNCaP cells homozygous for the rs1058205 T SNP-allele (A) qPCR analyses in LNCaP cells revealed a 25% decrease in KLK3 mRNA (\( p = 0.016 \)) following over-expression of miR-3162-5p compared to the negative control miRNA mimic treatment.
(B) Western blot analyses in LNCaP cells revealed a 32% decrease in cellular KLK3 protein
($p = 0.007$) following over-expression of miR-3162-5p compared to the negative control. (C)
Representative Western blot. Mean +/- SD, $n = 3$. (* = $p < 0.05$) (** = $p < 0.01$)
Figure 1
Figure 2

<table>
<thead>
<tr>
<th>P-value</th>
<th>Fold Change</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.71E-12</td>
<td>3.04</td>
<td>PDLM5</td>
</tr>
<tr>
<td>4.06E-4</td>
<td>2.06</td>
<td>KLK3</td>
</tr>
<tr>
<td>0.001</td>
<td>1.21</td>
<td>MCAT</td>
</tr>
<tr>
<td>0.001</td>
<td>1.31</td>
<td>TTLL12</td>
</tr>
<tr>
<td>0.006</td>
<td>1.27</td>
<td>PDK1</td>
</tr>
<tr>
<td>0.006</td>
<td>1.23</td>
<td>MDM4</td>
</tr>
<tr>
<td>0.040</td>
<td>1.18</td>
<td>VAMP8</td>
</tr>
<tr>
<td>0.058</td>
<td>1.09</td>
<td>PHC3</td>
</tr>
<tr>
<td>0.109</td>
<td>1.37</td>
<td>NKX3-1</td>
</tr>
<tr>
<td>0.174</td>
<td>1.11</td>
<td>TET2</td>
</tr>
<tr>
<td>0.358</td>
<td>1.02</td>
<td>GMEB2</td>
</tr>
<tr>
<td>0.425</td>
<td>1.04</td>
<td>SLC22A3</td>
</tr>
<tr>
<td>0.518</td>
<td>-1.00</td>
<td>ITGA6</td>
</tr>
<tr>
<td>0.947</td>
<td>-1.29</td>
<td>SLC22A2</td>
</tr>
<tr>
<td>0.970</td>
<td>-1.09</td>
<td>ARL3</td>
</tr>
<tr>
<td>0.999</td>
<td>-1.42</td>
<td>TMEM17</td>
</tr>
</tbody>
</table>

Legend:
- **Blue** represents up-regulation.
- **Red** represents down-regulation.
- **Black** represents no change.
- **Not measured**

Comparisons are made between **Normal** and **Tumour** conditions.
Figure 3

A. KLK3 rs1058205

B. VAMP8 rs1010
Figure 4

A

B

C

Relative KLK3 mRNA

Control miR-3162-5p

Relative KLK3 protein

Control miR-3162-5p

33 kDa

Control miR-3162-5p

KLK3 β-actin
A large scale analysis of genetic variants within putative miRNA binding sites in prostate cancer

Shane Stegeman, Ernest Amankwah, Kerenaftali Klein, et al.

Cancer Discovery  Published OnlineFirst February 17, 2015.

Updated version  Access the most recent version of this article at: doi:10.1158/2159-8290.CD-14-1057

Supplementary Material  Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2015/02/17/2159-8290.CD-14-1057.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.