Yale University
EliScholar – A Digital Platform for Scholarly Publishing at Yale

Public Health Theses   School of Public Health

January 2015

Pirna Sequence Variants Associated With Prostate Cancer In African Americans And Caucasians

Michael Cameron Lerro
Yale University, michael.lerro@yale.edu

Follow this and additional works at: http://elischolar.library.yale.edu/ysphtdl

piRNA SEQUENCE VARIANTS ASSOCIATED WITH PROSTATE CANCER IN AFRICAN AMERICANS AND CAUCASIANS

Michael C. Lerro

M.P.H. Thesis, Class of 2015

Department of Chronic Disease Epidemiology

Yale University School of Public Health

First Reader: Yong Zhu, PhD, MSc

Second Reader: Andrew DeWan, PhD, MA

ABSTRACT

Background: Prostate cancer is the most common cancer among men, and the incidence is 1.65 times higher in African Americans than Caucasians. The newly discovered PIWI/piRNA pathway, which regulates transposon and gene expression via small non-coding RNAs, has been implicated in certain aspects of cancer etiology. We hypothesize a novel involvement of the pathway via aberrant gene regulation due to sequence variants within piRNAs and test this hypothesis in relation to prostate cancer in both an African American and Caucasian sample. **Methods:** To interrogate SNPs embedded in piRNA sequences, we utilized genome-wide genotype data to impute 1,000 Genomes SNPs falling within piRNAs. We then tested for associations at these variants in bothe populations, while controlling for appropriate covariates and principal components. The regions encompassing significant SNPs were subsequently fine mapped. **Results:** In the African American sample one variant falling within a piRNA, rs61101785, was significantly associated with prostate cancer (FDR-p < 0.10). Fine mapping showed this variant to be the peak of an association signal. The variant is absent in the Caucasian sample. **Conclusions:** We have tested a novel hypothesis for the involvement of PIWI/piRNAs in cancer risk in a cancer estimated to make up 25% of new male cancer cases in 2015. Our results show the potential for a variant within a piRNA to affect cancer risk. Additionally, the risk variant is absent in Caucasians, potentially explaining some of the racial differences in prostate cancer risk.

INTRODUCTION

Prostate cancer is the most common cancer in men, with men facing a 15% lifetime risk

of developing the cancer and it predicted to make up 26% of new cancer cases in males in 2015,

for a total of 220,800 new cases[1]. Additionally, it is predicted to account for 9% of cancer related

deaths in men during 2015[1]. During the period of 2007-2011, the incidence of prostate cancer

was about 1.65 times higher in African Americans than in Caucasians[1]. Additionally, recent

research demonstrated that race modifies the risk of prostate cancer due to obesity, with obesity

being a stronger risk factor in African Americans than Caucasians[2]. Among other factors, it has

been hypothesized that the difference in prostate cancer risk between these two races has a

genetic component[3]. By performing this study in both an African American and Caucasian

sample, we may be able to shed light on genetic contributions to these racial differences.

The discovery of PIWI proteins and PIWI-interacting RNA (piRNAs), a class of small

non-coding RNAs, and the subsequent understanding of their biological role has spawned

interest in the potential role of these small RNAs in disease. The piwi gene was first identified in

drosophila through a genetic screen for genes affecting asymmetric division of germline stem

cells[4,5], and was then found to code for a highly conserved protein present in the stem and

somatic cells of the drosophila germline that is implicated in germline establishment and

maintenance[4-6]. PIWI proteins are members of the Argonaute family of proteins, which contain a

PAZ domain that binds single-stranded RNA, a MID domain, and a PIWI domain that resembles

the endonuclease RNase-H[7]. Homologs of drosophila piwi were then identified in various other

organisms including mice and humans[8-10]. After the discovery of PIWIs, it was shown that the

already characterized rasiRNAs and additional small RNAs interact with PIWI proteins, thus

being named piRNA[11-15]. The piRNAs identified primarily mapped to intergenic regions and are

enriched in repetitive elements, with about 20% in vertebrates mapping to transposon sequences[11-13].

Work to determine the function of PIWI/piRNAs has shown that the two are involved in the repression of transposable elements through transcriptional and post-transcriptional mechanisms, likely to maintain genome integrity[15-24]. In terms of transcriptional regulation, it has been shown that mutations in drosophila piwi and aub, two PIWIs, lead to a failure to establish H3K9me2/3 marks, a repressive histone modification[25-27]. In drosophila, this mechanism involves the interaction of PIWI proteins with Heterochromatin Protein-1 (HP1), demonstrating the ability of PIWI/piRNA complexes to recruit epigenetic modifiers to gene loci[28,29]. The processes of inducing epigenetic changes involves PIWI-bound piRNAs guiding PIWI in complex with epigenetic regulators to complementary DNA sequences or nascent transcripts where their action can take place[30].

Being of the same gene family as the proteins known to interact with miRNA, it is not surprising that evidence of a gene regulatory role for PIWI/piRNAs has also arisen. There is evidence in mice of PIWI/piRNAs directing the methylation of promoters, a DNA modification associated with decreased gene expression, in a sequence specific manner[31]. In drosophila, cytoplasmic PIWIs participate in inhibiting maternal mRNA translation and maternal mRNA decay via CCR4 mediated deadenylation by complementarity with their 3' UTRs[32]. Further, piRNAs can be generated from the 3' UTR of certain mRNAs in Drosophila, Xenopus, and mice, providing another possible method of regulation[33].

As these functions of PIWI proteins and piRNA have been elucidated, evidence of their association with cancers has come to light. PIWI expression has been demonstrated in a variety of human cancers, including colorectal, hepatic, brain, pancreatic, testicular, prostate, breast,

gastrointestinal, ovarian, and endometrial cancers[34-40]. Additionally, expression of piRNAs has

been observed in cancer cell lines and tissue samples[41-43]. Of these, specific piRNAs have been

observed to be under- or over-expressed in tumor tissue as compared to adjacent normal tissue,

and amelioration of this aberrant expression showed the effect of decreasing cell

proliferation[42,43].

Herein, we propose a potential novel involvement of the PIWI/piRNA pathway in

tumorigenesis. Specifically, we hypothesize that sequence variants within piRNAs may play a

role in cancer risk by aberrant regulation of tumor suppressor or oncogene expression. As

piRNAs serve as a sequence specific guide for PIWI proteins, their action at certain loci may be

abolished or aberrantly target novel loci. This idea is supported by the fact that single nucleotide

changes in piRNAs can lead to a substantial loss of efficiency at intended target sites[30]. Here, we

test this hypothesis in relation to prostate cancer in an African American population derived from

the Multi-ethnic Cohort (MEC) and Caucasian population from the Cancer Genetic Markers of

Susceptibility (CGEMS) Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial

(PLCO) by investigating associations with single nucleotide polymorphisms (SNPs) embedded

in piRNA sequences. We have limited our analyses to piRNAs derived from 100 or fewer

genetic loci, as there is evidence that low copy number piRNAs are more likely to regulate

protein-coding gene expression[44]

METHODS

Data

Data for this study were obtained from the Database of Genotypes and Phenotypes

(dbGaP) and includes genotype and phenotype data for African American subjects from the

GENEVA Prostate Cancer study (phs000306.v4.p1) genotyped on the Illumina Human1M-Duov3_B platform and Caucasian subjects from the CGEMS PLCO Prostate Cancer study (phs000207.v1.p1) genotyped on the Illumina HumanHap300v1.1 and HumanHap250Sv1.0 platforms. The subjects in the GENEVA study were derived from the MEC[45], as well as six additional studies as documented by *Freedman et al.*[46,47]. Subjects in the PLCO study were drawn from the PLCO Cohort by an incidence density sampling method[48].

Data Cleaning

All data cleaning/management was performed using PLINKv1.07[49]. For both study populations, consent groups were all genotyped on the same platform and were merged to make a complete data set. However, prior to merging consent groups the data were cleaned to remove individuals with a call rate <90%, SNPs with a call rate <95%, and SNPs not following HWE (p<0.0001). SNPs of low MAF were not removed, as they were not to be used for association analyses and would help inform subsequent imputation. SNPs on the Y chromosome, from pseudo-autosomal regions, and mitochondrial SNPs were removed. The data was then lifted over to genome build 37, and variant coding was flipped as necessary to be on the (+) genomic strand in order to be compatible with the imputation reference panel. One sample was removed from each related or duplicate pair in the data, as determined by $\hat{\pi} \geq 0.2$ from IBS analysis in plink[49]. Individuals not self-declared for the ancestry of interest were also excluded. Ancestry checks were then performed by combining the genome-wide data with a HapMap reference panel followed by principal component analysis using EIGENSTRAT[50]. All subjects in the Caucasian population clustered well with HapMap Caucasians, so no removals were made. However, due to African Americans being an admixed population and no clear clustering in relation to HapMap

samples, PCA was performed using EIGENSTRAT[50] and subjects exceeding six standard deviations on any of the top ten principal components were removed with one iteration.

piRNA SNP Genotype Imputation

We utilized piRNA Bank[51] to determine the position, sequence, and copy number of all curated human piRNA sequences. This includes 32,149 unique piRNAs that map to 667,944 genomic loci. We then used the 1,000 Genomes Phase 3 reference data available for IMPUTE2[52] and piRNA coordinates to determine all SNPs mapping to genomic coordinates covered by piRNAs encoded at 100 or fewer loci, as imputation is limited to variants in the reference panel. Next, imputation was carried out using IMPUTE2 in 5 MB segments with the programs default settings[52]. The program outputs a probability of having each of the three possible genotypes for each individual. SNPTEST then uses these probabilities to determine allele dosages for use in a logistic regression model as described in the subsequent section.

Association Analyses

Association analyses were carried out in SNPTESTv2.5 using unconditional logistic regression with an additive allelic model that inputs posterior genotype probabilities as dosages and accounts for uncertainty due to imputation[53]. Prior to the analyses, monomorphic SNPs as well as those with MAF<1% or with an info score <0.9 from IMPUTE2 were excluded. For the analyses in African Americans, models were adjusted for 3 principal components, an ordinal variable representing age categories, and study. The analyses in Caucasians controlled for 3 principal components, an ordinal age category variable, and family history of prostate cancer. For both analyses age was grouped by ten-year increments, and the number of principal components to control for was determined by calculating a genomic inflation factor (GIF) and inspecting QQ plots both generated using genome-wide data. Principal component analysis was

carried out in EIGENSTRAT[50] with LD-pruned data generated from plink[49] using a pairwise R^2 threshold of 0.5.

Fine Mapping

Fine mapping was performed on regions containing variants associated with prostate cancer. For this, all variants from the Thousand Genomes reference panel in the original 5MB imputation window of the associated SNP were imputed using IMPUTE2 in the same manner as previously. Association testing was then carried out using SNPTEST for all variants in a 500KB window centered on the SNP of interest while controlling for all of the same variables as in the piRNA variant association analyses. Imputed variants were limited based on an info quality metric of 0.6. P-values were then used to generate Manhattan plots for inspection of the distribution of the association signal.

RESULTS

GENEVA Study

After data cleaning in preparation for piRNA variant imputation, the African American population consisted of 2,275 cases and 2,425 controls for a total of 4,700 individuals with genotype data at 1,121,335 SNPs. During cleaning, 48 individuals were removed following IBS analysis (Figure S1a) and 22 were removed due to being PCA outliers after it was determined this would be necessary by inspection of ancestry plots (Figure S2a). Following this, piRNA SNP genotypes were imputed for each subject at all possible piRNA variants using IMPUTE2[52]. Prior to association testing, variants that were monomorphic, had a MAF<1%, or an IMPUTE2 quality info score < 0.9 were removed. The association analyses were controlled for the study a subject was drawn from, age categorized in ten year increments, and the top three eigenvectors

from PCA. The choice to control for three principal components was made based on the observation of a GIF of 1.00 from genome-wide association analyses and examination of QQ plots generated from these analyses (Figure S3a).

Association analyses, controlled as described above, were carried out for 1847 variants, the results of which are displayed in Figure 1. The variant rs61101785, located in piR-021163, was associated with an increased risk of prostate cancer [FDR-p=0.070], with an odds ratio of 1.63 [95% CI: 1.29-2.05] (Table 1). The MAF of the variant was 4.1% in cases and 2.6% in controls, and is located at Chr4: 3,074,158. The piRNA it falls within maps only to this locus[51]. The locus lies within the first intron of the Huntingtin antisense 1 (HTT-AS1) transcript (UCSC Genome Browser). Fine mapping of the region encompassing rs61101785 revealed that the association signal peaks at that variant falling within piR-021163 (Figure 2).

PLCO Study

After data cleaning, there were 1,142 cases and 1,098 controls for a total of 2,240 Caucasian subjects from the PLCO study genotyped at 541,721 variants. During cleaning, 7 samples were removed due to not being genotyped on both platforms and 53 were removed following IBS analysis (Figure S1b). All remaining subjects clustered well with HapMap Caucasians on the top two principal components (Figure S2b), so no PCA outlier removal was performed. Next, all SNPs falling within piRNAs encoded at 100 or fewer loci were imputed for use in association analyses. The association tests were adjusted for family history of prostate cancer, age categorized in ten year increments, and the top three principal components from PCA based on a GIF of 1.00 and QQ plot inspection (Figure S3b). As in the African American population, variants that were monomorphic, had a MAF<1%, or an IMPUTE2 info score < 0.9 were removed. Associations were tested at 1,364 SNPs, the results of which are summarized in

Figure 3. The top three hits are all located within the same piRNA cluster on Chromosome 14[51], which lies in an intergenic region. Interestingly, the hits within this single piRNA cluster all correspond to single copy piRNAs. The fine mapping carried out on the imputation region encompassing rs8010969 and rs11625907 revealed that they are likely tagging a causal SNP (Figure 4).

DISCUSSION

Here, we have performed the first comprehensive analysis investigating the association between genetic variants within piRNAs and prostate cancer in both an African American and Caucasian sample. The study focused on an African American sample genotyped as part of the GENEVA study and Caucasian population drawn from the PLCO study, both available via dbGaP. Investigation of associations between imputed piRNA variants and prostate cancer revealed a highly interesting association in the African American study sample. A variant falling within the singly encoded piR-021163, rs61101785, was associated with an increased risk of prostate cancer in African Americans (FDR-p=0.0702). Fine mapping of the region encompassing rs61101785 demonstrated that the association signal peaks at the variant. This supports the idea of a functional role for this variant given the signal is real. The location of the variant (Chr4: 3,074,158) and piRNA (Chr4: 3,074,147-3,074,178) falls within the first intron of the HTT-AS1 transcript. HTT-AS1 is non-coding and antisense to the Huntingtin (HTT) gene[54], a gene causally linked to Huntington's disease when containing a PolyQ expansion[55,56], with the two being transcribed head-to-head. The HTT-AS1 transcript is known to regulate the expression of the HTT gene in a partially Dicer dependent manner[54]. The normal HTT gene has been implicated in cell survival[57], an important aspect of cancer development and progression.

Interestingly, piRNAs typically target transposons by deriving from antisense transcripts and imprinting of the Rasgfr1 locus in mice involves the targeting of an adjacent antisense transcript by a specific piRNA[31]. Although speculative, it is possible that this piRNA is derived from the antisense transcript and can then target the genomic locus. Another interesting aspect of this variant is that it was virtually monomorphic in the Caucasian sample, with only two cases being heterozygous at this position. This could partially account for the race differences observed in prostate cancer risk.

The associations observed in the piRNA cluster on chromosome 14 in the Caucasian sample are all likely reflecting the same functional variant if one is truly present. Future work may investigate if any of these piRNA variants differentially affect aspects of cancer development and progression as compared to the wildtype piRNAs.

A strength of this study is the Thousand Genomes reference panel used for imputation, as this data has highly comprehensive coverage of an immense number of variants. This allowed us to achieve coverage of many piRNA embedded SNPs. Howevever, we could not interrogate variants within piRNAs not included in the reference panel, which means some piRNAs were not interrogated. Another drawback is the fact that we cannot conclude that these variants are causal, as they may be tagging variants in linkage-disequilibrium with them. However, future functional analyses will help understand whether or not these variants are playing a role in cancer risk. Future work will focus on the effects of the wildtype and variant piRNA-021163 on aspects of tumorigenesis, including proliferation. Also, work may focus on the gene regulatory effects of the piRNA, which can be difficult to predict. Finally, we were limited to covariates provided in the dbGaP datasets, and could have potentially further controlled for confounding factors to bolster our results.

Overall, we provide the first evidence that piRNA sequence variants could potentially be associated with prostate cancer, with a strong finding coming from our African American sample that could in part explain racial differences in prostate cancer risk. Fine mapping of the region strengthened this idea, and future functional work will help to understand how this variant may be affecting risk. Therefore, it is possible that not only abberant expression of PIWIs or piRNAs can play a role in cancer but piRNA sequence changes may also be a factor.

ACKNOWLEDGEMENTS

TABLES & FIGURES
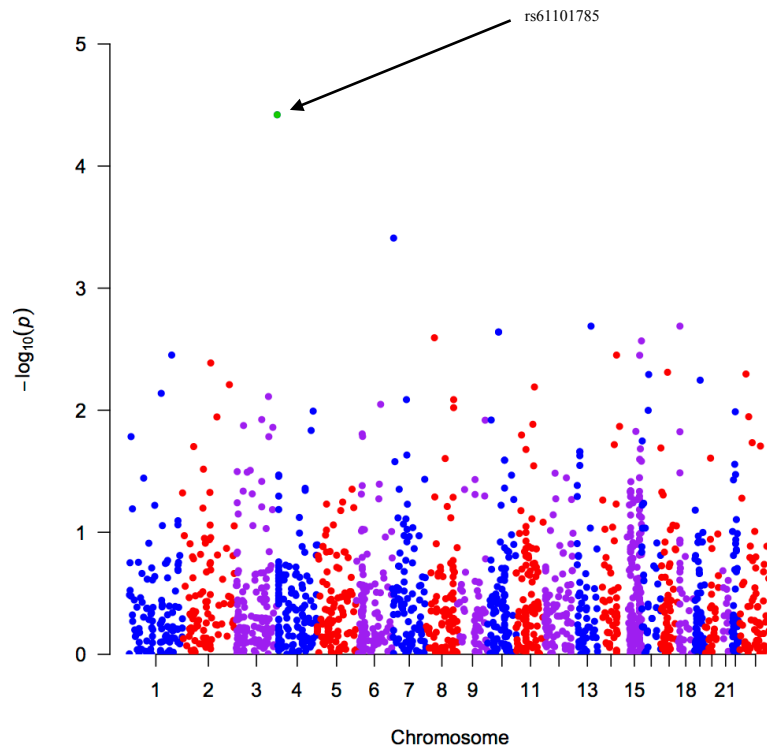
Figure 1: A Manhattan plot displaying the association results for imputed piRNA variants from the African American subjects of the GENEVA study. The significantly associated variant, rs61101785, is highlighted in green.

Variant	Location	piRNA	MAF Cases	MAF Controls	OR [95% CI]	P	FDR-Adjusted P
GENEVA (n=4,700)							
rs61101785	Chr4: 3,074,158	piR-021163	4.1%	2.6%	1.63 [1.29, 2.05]	3.80E-05	0.070
rs62439721	Chr7: 6,762,443	piR-003123	16.8%	19.3%	0.84 [0.76, 0.94]	3.89E-04	0.359
rs11074184	Chr15: 93,970,292	piR-008061	9.6%	11.6%	0.81 [0.71, 0.93]	3.55E-03	0.596
PLCO (n=2,240)							
rs8010969	Chr14: 88,626,243	piR-013783 piR-014246	19.6%	16.0%	1.28 [1.10, 1.49]	1.18E-03	0.180
rs11625907	Chr14: 88,625,605	piR-008286	19.7%	16.0%	1.28 [1.10, 1.49]	1.17E-03	0.199
rs8020378	Chr14: 88,624,946	piR-018495	19.7%	16.0%	1.28 [1.10, 1.50]	1.15E-03	0.225

**Table 1**: Summary of the top three hits from the GENEVA African American and PLCO Caucasian samples. FDR-adjusted p-values < 0.10 and < 0.20 were considered significant and suggestive, respectively.
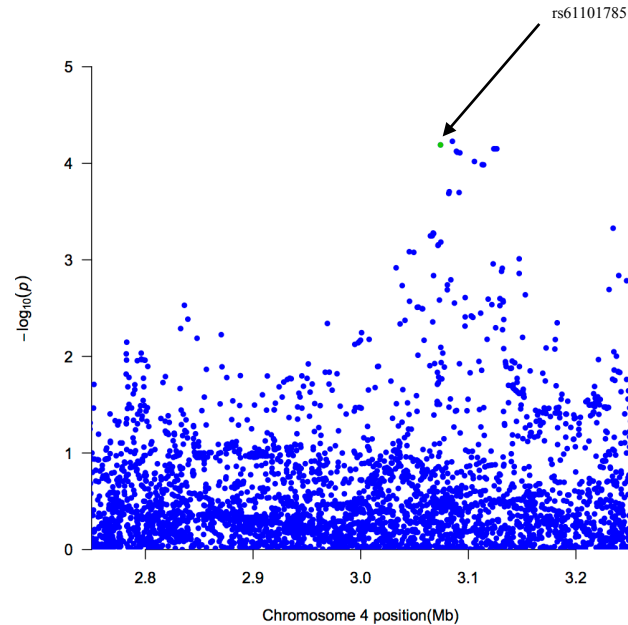
Figure 2: A Manhattan plot displaying the results of fine mapping of the region encompassing rs61101785, which is highlighted in green.
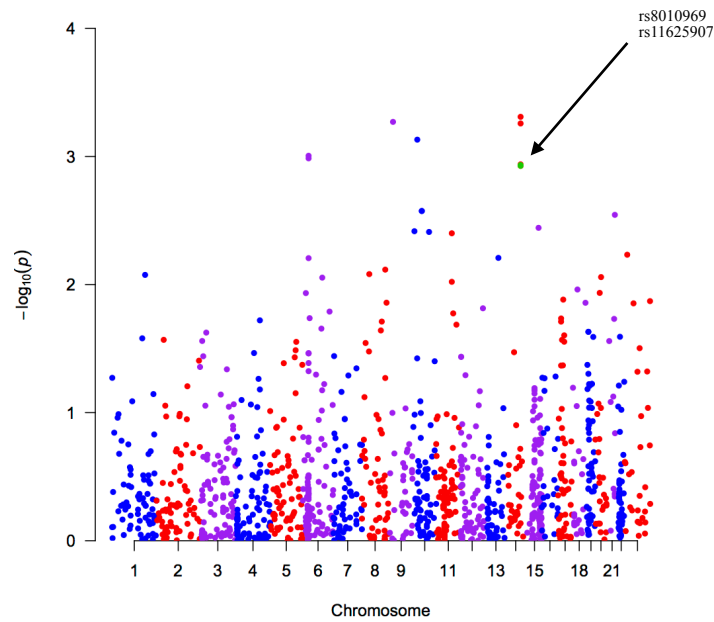
Figure 3: A Manhattan plot displaying the association results for imputed piRNA variants from the Caucasian subjects of the PLCO study. The variants rs8010969 and rs11625907 are highlighted in green.
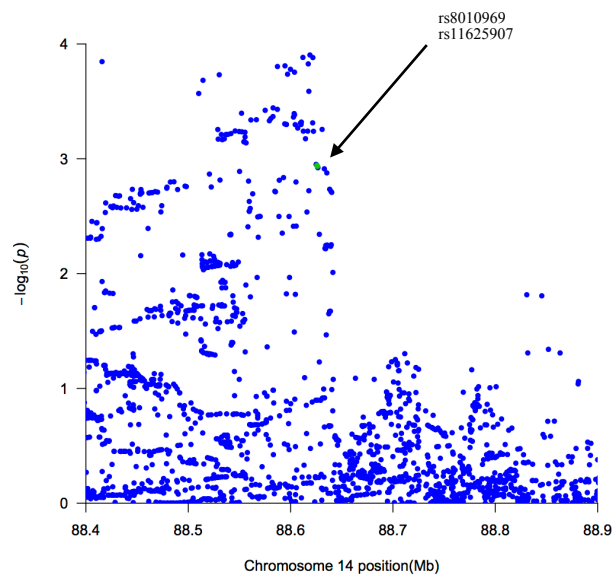

Figure 4: A Manhattan plot displaying the results of fine mapping of the region encompassing the variants rs8010969 and rs11625907, which are highlighted in green.

SUPPLEMENTS
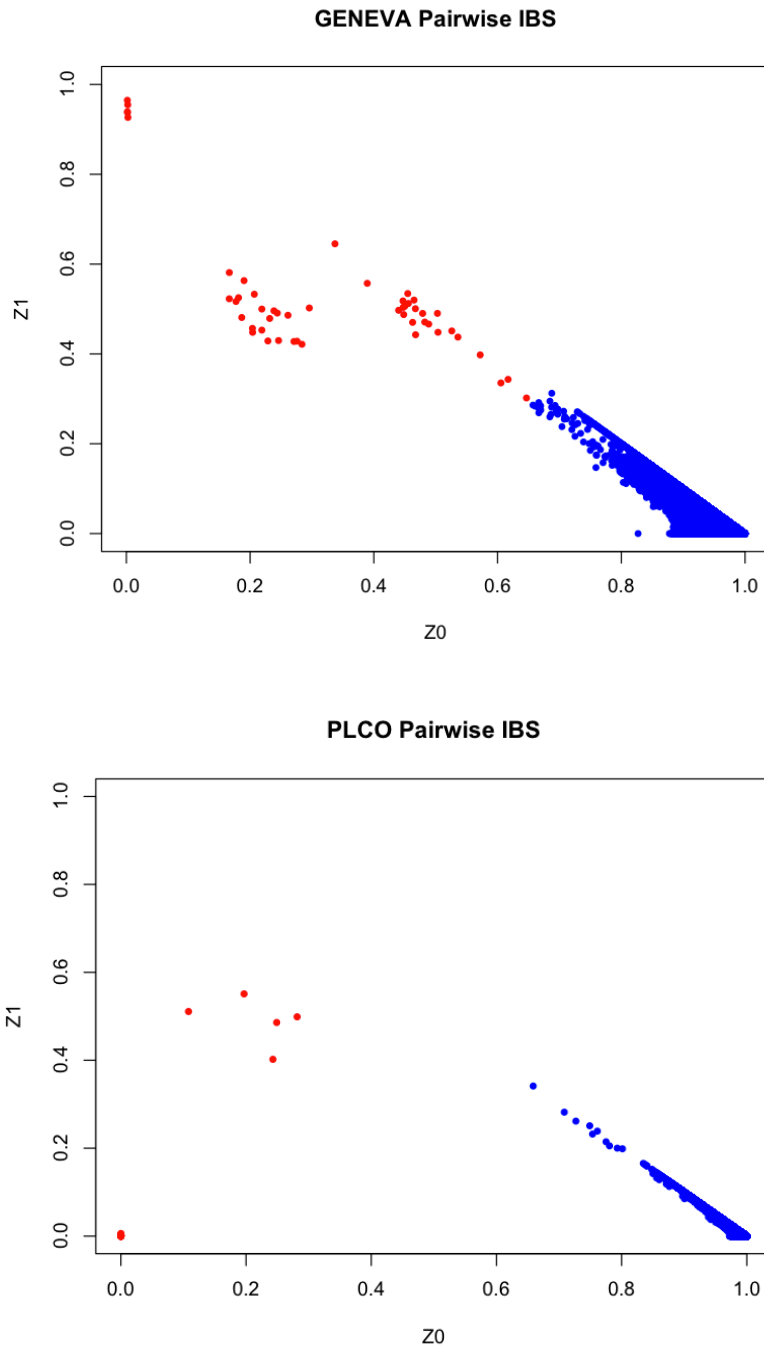
**Figures S1a and S1b**: IBS analysis. Plots of the probability of sharing one allele at a given locus vs. the probability of sharing zero alleles at a given locus. One sample was removed from each pair in red.
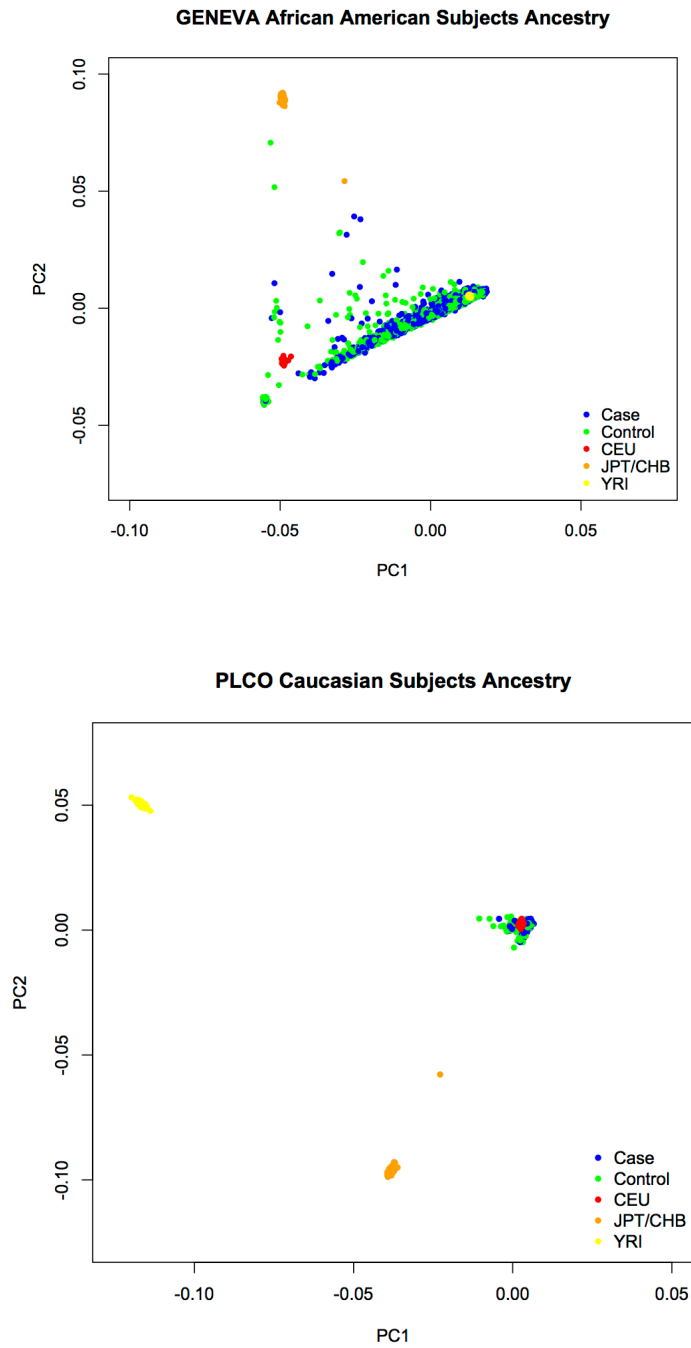
Figures S2a and S2b: Plots of the first two principle components from PCA analysis of study subjects merged with HapMap references. CEU = Caucasian, JPT/CHB = East Asian, YRI = Yoruba (West Africa).
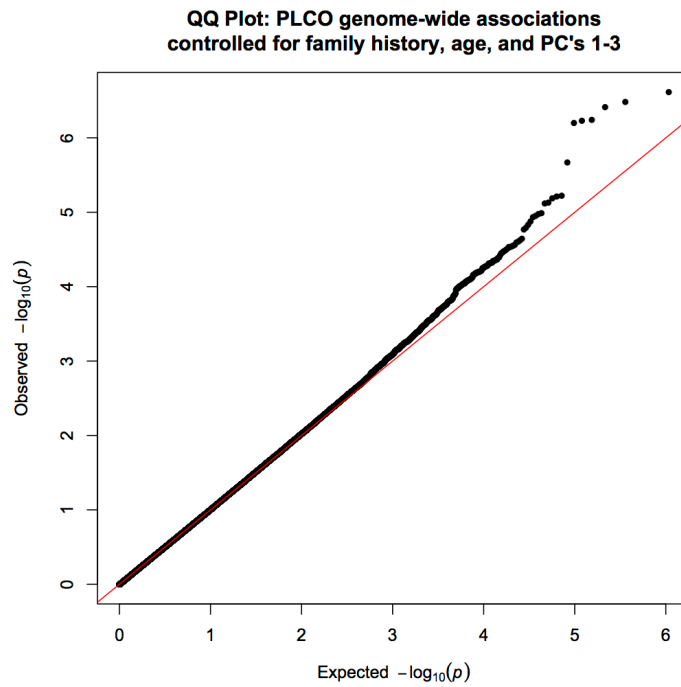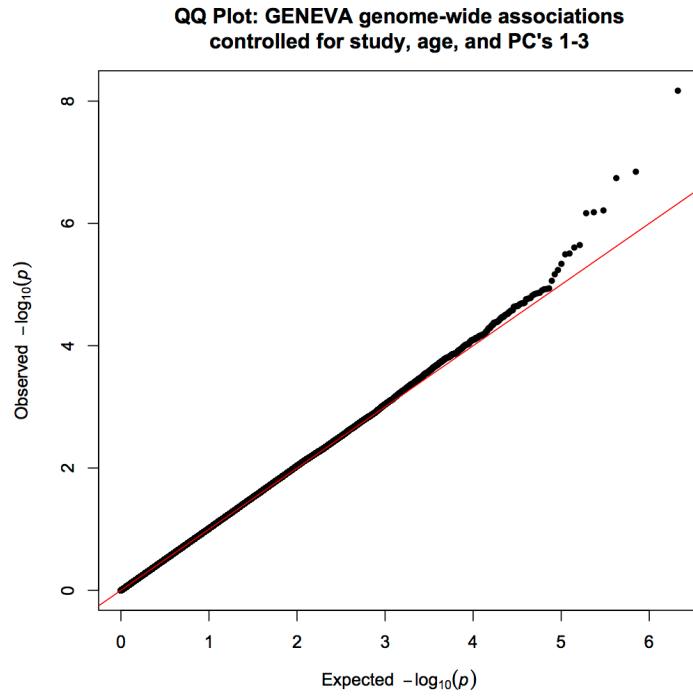
Figure S3a and S3b: QQ plots based on genome-wide associations adjusted for the covariates used in the association analyses for both studies.

REFERENCES

1.   Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians* **65**, 5-29 (2015).

2.   Barrington, W.E., Schenk, J.M., Etzioni, R. & et al. DIfference in association of obesity with prostate cancer risk between us african american and non-hispanic white men in the selenium and vitamin e cancer prevention trial (select). *JAMA Oncology* (2015).

3.   Freedland, S.J. & Isaacs, W.B. Explaining racial differences in prostate cancer in the United States: Sociology or biology? *The Prostate* **62**, 243-252 (2005).

4.   Lin, H. & Spradling, A.C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development* **124**, 2463-2476 (1997).

5.   Cox, D.N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & Development* **12**, 3715-3727 (1998).

6.   Cox, D.N., Chao, A. & Lin, H. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503-514 (2000).

7.   Hutvagner, G. & Simard, M.J. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**, 22-32 (2008).

8.   Qiao, D., Zeeman, A.M., Deng, W., Looijenga, L.H. & Lin, H. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. *Oncogene* **21**, 3988-99 (2002).

9.   Sasaki, T., Shiohama, A., Minoshima, S. & Shimizu, N. Identification of eight members of the Argonaute family in the human genome☆. *Genomics* **82**, 323-330 (2003).

10.   Kuramochi-Miyagawa, S. *et al.* Two mouse piwi-related genes: miwi and mili. *Mechanisms of Development* **108**, 121-133 (2001).

11.   Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-207 (2006).

12.   Grivna, S.T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development* **20**, 1709-1714 (2006).

13.   Girard, A., Sachidanandam, R., Hannon, G.J. & Carmell, M.A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202 (2006).

14.   Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes & Development* **20**, 1732-1743 (2006).

15.   Yin, H. & Lin, H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* **450**, 304-8 (2007).

16.   Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G.J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744-7 (2007).

17.   Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-103 (2007).

18.   Chen, Y., Pane, A. & Schupbach, T. Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. *Curr Biol* **17**, 637-42 (2007).

19.   Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**, 908-17 (2008).

20.   Vagin, V.V. *et al.* The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. *RNA Biol* **1**, 54-8 (2004).

21.   Xu, M. *et al.* Mice deficient for a small cluster of Piwi-interacting RNAs implicate Piwi-interacting RNAs in transposon control. *Biol Reprod* **79**, 51-7 (2008).

22.   Aravin, A.A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-99 (2008).

23.   Pelisson, A., Sarot, E., Payen-Groschene, G. & Bucheton, A. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. *J Virol* **81**, 1951-60 (2007).

24.   Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. *Genetics* **166**, 1313-21 (2004).

25.   Klenov, M.S. *et al.* Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. *Nucleic Acids Res* **35**, 5430-8 (2007).

26.   Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell* **151**, 964-980 (2012).

27. Wang, S.H. & Elgin, S.C.R. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proceedings of the National Academy of Sciences* **108**, 21164-21169 (2011).

28. Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21**, 2300-11 (2007).

29. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**, 669-72 (2004).

30. Huang, Xiao A. *et al.* A Major Epigenetic Programming Mechanism Guided by piRNAs. *Developmental Cell* **24**, 502-516 (2013).

31. Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848-52 (2011).

32. Rouget, C. *et al.* Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature* **467**, 1128-32 (2010).

33. Robine, N. *et al.* A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* **19**, 2066-76 (2009).

34. Zeng, Y. *et al.* HIWI expression profile in cancer cells and its prognostic value for patients with colorectal cancer. *Chin Med J (Engl)* **124**, 2144-9 (2011).

35. Jiang, J., Zhang, H., Tang, Q., Hao, B. & Shi, R. Expression of HIWI in Human Hepatocellular Carcinoma. *Cell Biochemistry and Biophysics* **61**, 53-58 (2011).

36. Sun, G. *et al.* Clinical significance of Hiwi gene expression in gliomas. *Brain Research* **1373**, 183-188 (2011).

37. Grochola, L.F. *et al.* The stem cell-associated Hiwi gene in human adenocarcinoma of the pancreas: expression and risk of tumour-related death. *Br J Cancer* **99**, 1083-8 (2008).

38. Liu, J.J. *et al.* Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. *Int J Clin Exp Pathol* **3**, 328-37 (2010).

39. Lee, J.H. *et al.* Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway. *Hum Mol Genet* **15**, 201-11 (2006).

40. Liu, X. *et al.* Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. *International Journal of Cancer* **118**, 1922-1929 (2006).

41. Lu, Y. *et al.* Identification of piRNAs in Hela cells by massive parallel sequencing. *BMB Rep* **43**, 635-41 (2010).

42. Cheng, J. *et al.* piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells. *Cancer Letters* **315**, 12-17 (2012).

43. Cheng, J. *et al.* piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clinica Chimica Acta* **412**, 1621-1625 (2011).

44. Fu, A., Jacobs, D.I. & Zhu, Y. Epigenome-wide analysis of piRNAs in gene-specific DNA methylation. *RNA Biology* **11**, 1301-1312 (2014).

45. Kolonel, L.N. *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* **151**, 346-57 (2000).

46. Freedman, M.L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* **103**, 14068-73 (2006).

47. Haiman, C.A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* **39**, 638-44 (2007).

48. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**, 645-9 (2007).

49. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).

50. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).

51. Sai Lakshmi, S. & Agrawal, S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Research* **36**, D173-D177 (2008).

52. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).

53. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).

54. Chung, D.W., Rudnicki, D.D., Yu, L. & Margolis, R.L. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Human Molecular Genetics* **20**, 3467-3477 (2011).

55. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971-83 (1993).

56. Walker, F.O. Huntington's disease. *Lancet* **369**, 218-28 (2007).

57.   Rigamonti, D. *et al.* Wild-type huntingtin protects from apoptosis upstream of

        caspase-3. *J Neurosci* **20**, 3705-13 (2000).