The combined effect of a polygenic risk score and rare genetic variants on prostate cancer risk

Burcu F. Darst[a], Xin Sheng[a], Rosalind A. Eeles[b,c], Zsofia Kote-Jarai[b], David V. Conti[a], Christopher A. Haiman[a]

[a] Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA

[b] The Institute of Cancer Research, London, UK

[c] The Royal Marsden NHS Foundation Trust, London, UK

Corresponding Author

Burcu F. Darst

1450 Biggy Street

Los Angeles, CA 90033 USA

1-232-442-0078

bdarst@usc.edu

Keywords (3-10): Biobank; Common variants; Exome sequencing; Genetics; Genomics; HOXB13; Polygenic risk score; Prostate cancer; Rare variants

Word count of text (including abstract): 1,673

Word count of abstract: 200

Abstract (200 word limit)

Although prostate cancer is known to have a strong genetic basis and is influenced by both common and rare variants, the ability to investigate the combined effect of such genetic risk factors has been limited to date. We conducted an investigation of 81,094 men from the UK Biobank, including 3,568 prostate cancer cases, to examine the combined effect of rare pathogenic/likely pathogenic/deleterious (P/LP/D) germline variants and common prostate cancer risk variants, measured using a polygenic risk score (PRS), on prostate cancer risk. Absolute risk of prostate cancer for *HOXB13*, *BRCA2*, *ATM*, and *CHEK2* P/LP/D carriers ranged from 9% to 56%, whereas absolute risk in non-carriers ranged from 2% to 31%, by age 85 for men in the lowest and highest PRS deciles, respectively. The high-penetrant *HOXB13* G84E prostate cancer risk variant was most common in cases in the lowest PRS quintile (4.4%) and least common in cases in the highest PRS quintile (0.5%; P=0.005), whereas there was no statistically significant difference in frequencies by PRS in controls. While rare and common variants strongly and distinctly influence prostate cancer onset, considering rare and common variants in conjunction will lead to more precise estimates of a man's lifetime risk of prostate cancer.

Patient summary: We found that the risk of prostate cancer conveyed by rare variants could vary depending on an individual's genetic profile of common risk variants. This implies that in order to comprehensively assess genetic risk of prostate cancer, it is important to consider both rare and common variants.

Prostate cancer (PCa) is a leading cause of death, with high heritability and risk among family members suggesting a strong genetic basis of this disease[1, 2]. Rare germline genetic variants have been shown to increase PCa risk[3, 4], as have common variants in aggregate as measured by polygenic risk scores (PRS), with men in the highest PRS decile having approximately 4-fold increased odds of PCa than men in the average 40-60% PRS category[5]. Until recently, the ability to investigate the combined influence of rare and common variants has been limited. Recent studies have shown that common variants modify the influence of rare variants on breast cancer, colorectal cancer, and coronary artery disease risk[6, 7], and the influence of rare *BRCA2* and *HOXB13* variants on PCa risk has been shown to vary by common variants[8-10]. Here, we investigated the combined effect of rare and common germline variants on PCa risk using whole-exome sequencing and genome-wide genotype data in a large sample of 81,094 European ancestry men from the UK Biobank (October 2020 release of 200K whole-exome sequences), including 3,568 PCa cases and 77,526 controls.

We first performed exome-wide gene-based analyses to determine whether novel PCa risk genes could be identified from rare pathogenic/likely pathogenic/deleterious (P/LP/D) variants. Across 14,905 tested genes, only *HOXB13* (OR=4.63, 95% CI=3.26-6.59, $P=1.4x10^{-17}$) and *CHEK2* (OR=2.06, 95% CI=1.51-2.80, $P=4.9x10^{-6}$) reached genome-wide significance (**Supplementary Figure 1**). Limiting to 151 DNA repair genes, which have been implicated in PCa risk[4, 11], only *CHEK2* (see above) and *BRCA2* (OR=2.15, 95% CI=1.40-3.28, $P=4.2x10^{-4}$) were significantly associated with PCa risk after multiple testing adjustment (**Supplementary Figure 2**). Testing individual exome-wide P/LP/D variants, two significant associations were

identified: known rs138213197 (G84E) in *HOXB13* (control carrier frequency=0.31%, case carrier frequency=1.29%, P=6.9x10^{-18}) and novel rs769540160 in *MYO3A* (control carrier frequency=0.004%, case carrier frequency=0.11%, P=1.1x10^{-7}) (**Supplementary Figure 3**). Although *MYO3A* was not genome-wide significant in gene-based tests, results were suggestive of carriers having 1.67-fold increased odds of PCa (95% CI=1.06-2.65, P=0.027). The carrier frequency of rs769540160 in 32,330 cancer-free European ancestry individuals in gnomAD was 0.009%[12], while it was 4-fold more common in a whole-exome sequencing study of 5,545 European ancestry men with aggressive and non-aggressive PCa, with a carrier frequency of 0.04% (carried by two men who both died due to PCa and had Gleason scores ≥8)[4]. Given the extreme rarity of this variant, additional large-scale PCa sequencing studies are necessary to further validate this novel association.

Combined rare and common variant analyses focused on carrier status of P/LP/D variants in known PCa risk genes (*HOXB13* and DNA repair genes *BRCA2*, *BRCA1*, *PALB2*, *ATM*, *CHEK2*, *NBN*, and *MSH2*)[4, 11, 13] and our recently developed multi-ancestry PRS[5]. *HOXB13*, *BRCA2*, *ATM*, and *CHEK2* had sufficient numbers of P/LP/D carriers for analyses and were consequently the focus of our investigation. Analyses jointly evaluating the PRS and carrier status excluded *HOXB13* G84E and/or *CHEK2* 1100delC from the PRS when carrier status included either of these variants. Of the total 1,576 carriers of P/LP/D alleles in these four genes, 19 men carried two P/LP/D alleles (including two cases) and the remaining 1,557 men carried one P/LP/D allele (including 143 cases). As expected, these four genes showed strong associations with PCa risk, as did the multi-ancestry PRS, which had stronger effects than a previously

developed European ancestry PRS[14] (**Supplementary Tables 1-2**). In aggregate, P/LP/D carriers had 2.52-fold increased odds of PCa (2.10-3.04, $P=1.40x10^{-22}$) and 4.73-fold increased odds of dying due to PCa (95% CI=2.82-7.94, $P=4.1x10^{-9}$). Although we had insufficient clinical data to further evaluate aggressive or lethal disease (220 men died due to PCa, of which 16 carried P/LP/D alleles in these genes), we previously reported that P/LP/D variants in *ATM* and *BRCA2* were more common in men with aggressive (and lethal) disease compared to men with non-aggressive PCa[4]. Aggregate effects of P/LP/D variants in these genes did not significantly differ in men with and without a first-degree family history of prostate cancer or in men ≤60 or >60 years of age (**Supplementary Tables 3-4**). PRS effects also did not significantly differ by family history; however, the PRS had significantly larger effects in younger compared to older men (**Supplementary Tables 5-6**), consistent with previous findings[5].

Relative to non-carriers in the average 40-60% PRS category, odds ratios ranged from 0.28 (95% CI=0.22-0.36) to 4.34 (95% CI=3.87-4.87) for non-carriers and 1.06 (95% CI=0.43-2.64) to 10.21 (95% CI=6.53-15.96) for carriers in the lowest and highest PRS decile, respectively (**Figure 1a**). Absolute risk of PCa by age 85 ranged from 2% to 31% for non-carriers and 9% to 56% for carriers in the lowest and highest PRS decile, respectively (**Figure 1b**). Absolute risk for carriers in the 90-100% PRS category (56%) was similar to the 55% absolute risk for men in the 99-100% PRS category (independent of carrier status) and two-fold higher than the 26% absolute risk for carriers (independent of PRS). Absolute risk for carriers in the 0-10% PRS category (9%) was similar to the 11% absolute risk for non-carriers (independent of PRS; **Figure 1c**). Effects and absolute risks were slightly weaker when excluding *HOXB13* G84E

from carrier status (**Supplementary Figure 4**). Evaluating the four genes separately revealed similar findings (**Supplementary Figures 5-8**), with *HOXB13* G84E carriers having notably increased PCa risk compared to non-carriers across PRS quintiles (used instead of deciles given smaller numbers of carriers within individual genes). Across PRS quintiles, odds ratios for *HOXB13* G84E carriers ranged from 2.96 (95% CI=1.19-7.34) to 10.10 (95% CI=5.03-20.28) (relative to *HOXB13* G84E non-carriers in the average 40-60% PRS category), while absolute risks for *HOXB13* G84E carriers ranged from 23% to 56% by age 85 (**Supplementary Figure 5**). We observed a statistically significant interaction between the continuous PRS and carrier status for *HOXB13* (P=0.041), but not for the other genes separately or in aggregate (P≥0.14; **Supplementary Table 1**).

Interestingly, among cases, *HOXB13* G84E was most common in the lowest PRS quintile (4.4%) and least common in the highest PRS quintile (0.5%), whereas control carrier frequencies across PRS quintiles were consistently 0.3% (**Figure 2a**). Accordingly, the average PRS was higher in *HOXB13* G84E non-carriers than carriers among cases (P=0.005) and did not significantly differ by carrier status among controls (P=0.3; **Figure 2c**). The frequency of *CHEK2* P/LP/D carriers was also most common in cases in the lowest PRS quintile (2.3%) and least common in the highest PRS quintile (1.3%; **Supplementary Figure 9**); however, PRS did not significantly differ by carrier status (P=0.3; **Supplementary Figure 10**). The *HOXB13* G84E finding was validated using GWAS data in an independent sample of 5,197 cases and 115,796 controls in the UK Biobank, with cases having a carrier frequency of 3.2% in the lowest PRS quintile and 1.2% in the highest PRS quintile (**Supplementary Figure 11**). Similar results were

observed in the full UK Biobank sample (8,765 cases and 193,322 controls; **Supplementary Figure 12**). This finding suggests that *HOXB13* G84E may account for more PCa in men with low versus high PRS. While carrying rare or common risk variants could serve as independent pathways to PCa onset, our results suggest that carrying rare variants in these genes and having high PRS compound PCa risk.

Findings from this investigation suggest that PCa risk may vary depending on an individual's genetic profile of common risk variants, measured by PRS, and carrier status for rare P/LP/D variants in *HOXB13* and *BRCA2*, with novel evidence for variants in *CHEK2* and *ATM*. In particular, men in the top PRS decile had higher absolute risk of PCa than carriers (31% vs 25%); however, considering the PRS and carrier status jointly, absolute risk for non-carriers in the top PRS decile was 31%, while it increased to 56% for carriers in the top PRS decile. This is supported by previous findings of rare and common variants collectively improving discriminative ability of PCa risk models[15] and could have important clinical implications, such as informing decisions regarding PCa screening, with P/LP/D carriers and/or men with a high PRS potentially benefiting from earlier and more frequent screening. Further studies are underway and needed to evaluate the impact of such clinical implementations. Consistent with studies of other diseases[16], our findings also suggest that rare and common variants could independently lead to PCa onset, with low PRS cases being more likely to carry *HOXB13* G84E, for example. Whole-genome sequencing efforts could have improved power to identify additional moderate- to high-penetrant rare PCa risk variants by prioritizing low PRS cases, as extreme sampling has been shown to improve power to detect rare variants[17]. It will be important to extend this to clinical investigations to

161 determine whether PRS in conjunction with carrier status for rare P/LP/D variants could
162 better discern aggressive PCa, which we were unable to investigate in this study.
163 Further, similar investigations in non-European ancestry men will be critical, particularly
164 in men of African ancestry given the established genetic contribution to high PCa
165 incidence rates in this population[5].

Acknowledgements

This work was supported by the National Cancer Institute at the National Institutes of Health grant (K99 CA246063, BFD and R01 CA196931, CAH) and an award from the Achievement Rewards for College Scientists Foundation Los Angeles Founder Chapter (BFD). This research has been conducted using the UK Biobank Resource under application number 42195.

Author Contributions

Study conception/design: BFD, DVC, CAH; Data analysis: BFD; Bioinformatics support: XS; Interpretation of data: BFD, RAE, ZK-J, DVC, CAH; Drafted the manuscript: BFD; Contributed to manuscript revisions: BFD, XS, RAE, ZK-J, DVC, CAH; All authors approved of the final manuscript.

Competing Interests Statement

The authors have no conflicts of interest to disclose.

References

[1] Hjelmborg JB, Scheike T, Holst K, Skytthe A, Penney KL, Graff RE, et al. The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev. 2014;23:2303-10.

[2] Hemminki K. Familial risk and familial survival in prostate cancer. World J Urol. 2012;30:143-8.

[3] Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, Quinque D, et al. The contribution of rare variation to prostate cancer heritability. Nat Genet. 2016;48:30-5.

[4] Darst BF, Dadaev T, Saunders E, Sheng X, Wan P, Pooler L, et al. Germline sequencing DNA repair genes in 5,545 men with aggressive and non-aggressive prostate cancer. J Natl Cancer Inst. 2020.

[5] Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat Genet. 2021;53:65-75.

[6] Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020;11:3635.

[7] Gallagher S, Hughes E, Wagner S, Tshiaba P, Rosenthal E, Roa BB, et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. JAMA Netw Open. 2020;3:e208501.

[8] Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, et al. A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. Eur Urol. 2014;65:169-76.

[9] Kote-Jarai Z, Mikropoulos C, Leongamornlert DA, Dadaev T, Tymrakiewicz M, Saunders EJ, et al. Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes. Ann Oncol. 2015;26:756-61.

[10] Lecarpentier J, Silvestri V, Kuchenbaecker KB, Barrowdale D, Dennis J, McGuffog L, et al. Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. J Clin Oncol. 2017;35:2240-50.

[11] Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med. 2016;375:443-53.

[12] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434-43.

[13] Matejcic M, Patel Y, Lilyquist J, Hu C, Lee KY, Gnanaolivu RD, et al. Pathogenic Variants in Cancer Predisposition Genes and Prostate Cancer Risk in Men of African Ancestry. JCO Precis Oncol. 2020;4:32-43.

[14] Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50:928-36.

[15] Shi Z, Platz EA, Wei J, Na R, Fantus RJ, Wang CH, et al. Performance of Three Inherited Risk Measures for Predicting Prostate Cancer Incidence and Mortality: A Population-based Prospective Analysis. Eur Urol. 2020.

[16] Lu T, Zhou S, Wu H, Forgetta V, Greenwood CMT, Richards JB. Individuals with common diseases but with a low polygenic risk score could be prioritized for rare variant screening. Genet Med. 2020.

[17] Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genet Epidemiol. 2011;35:790-9.

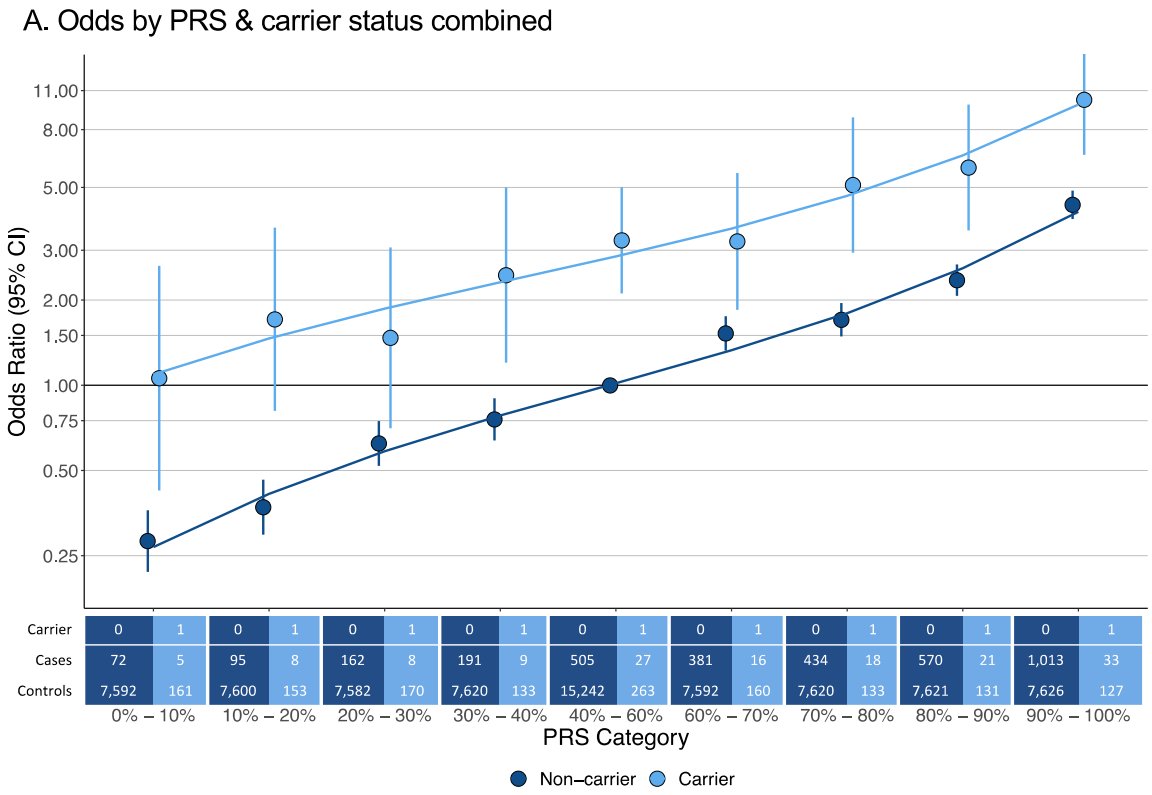
Figure Legends

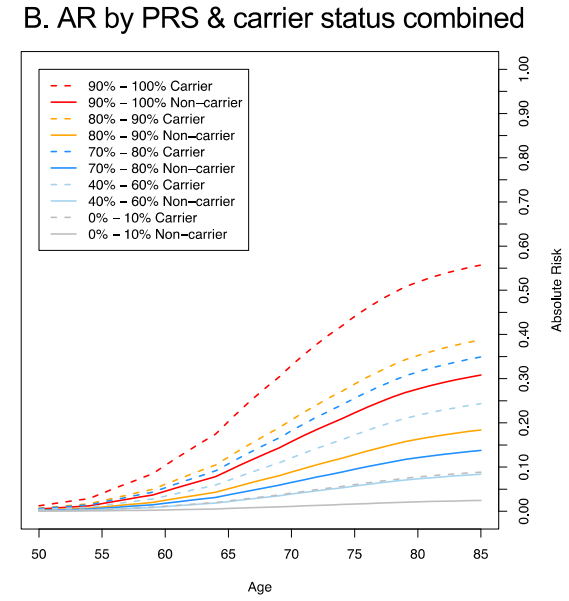
Figure 1. Aggregate effect of P/LP/D variants in *ATM*, *BRCA2*, *CHEK2*, and *HOXB13* and a polygenic risk score (PRS) on prostate cancer risk. A) Odds of prostate cancer by PRS category and carrier status. Odds ratios are calculated with respect to the referent non-carrier 40-60% PRS category. Percentage of total cases are annotated for each effect estimate, and sample sizes of carriers and non-carriers by case status and PRS category are indicated below the figure. In the 40-60% PRS category, 0.76% and 14.15% of total cases are carriers and non-carriers, respectively. OR are plotted on a log-scale. B) Absolute risk (AR) of prostate cancer by age and the combination of carrier status and PRS category. The 40-60% PRS non-carrier line estimates baseline AR by age (8.4% lifetime AR). C) Absolute risk of prostate cancer by age and carrier status (independent of PRS) and PRS category (independent of carrier status). The 40-60% PRS line (8.1% lifetime AR) and non-carrier line (11.3% lifetime AR) estimates baseline AR by age.

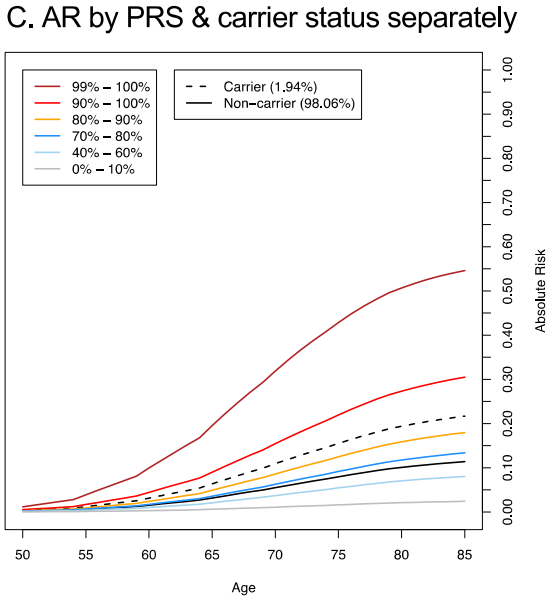
Figure 2. Polygenic risk score (PRS) distribution of *HOXB13* G84E carriers. A) *HOXB13* G84E carrier frequency by PRS category and prostate cancer status. C) PRS distribution by *HOXB13* G84E carrier status and prostate cancer status. PRS differences between carriers and non-carriers are calculated using a two-sided t-test.

253 Figures

254 Figure 1.

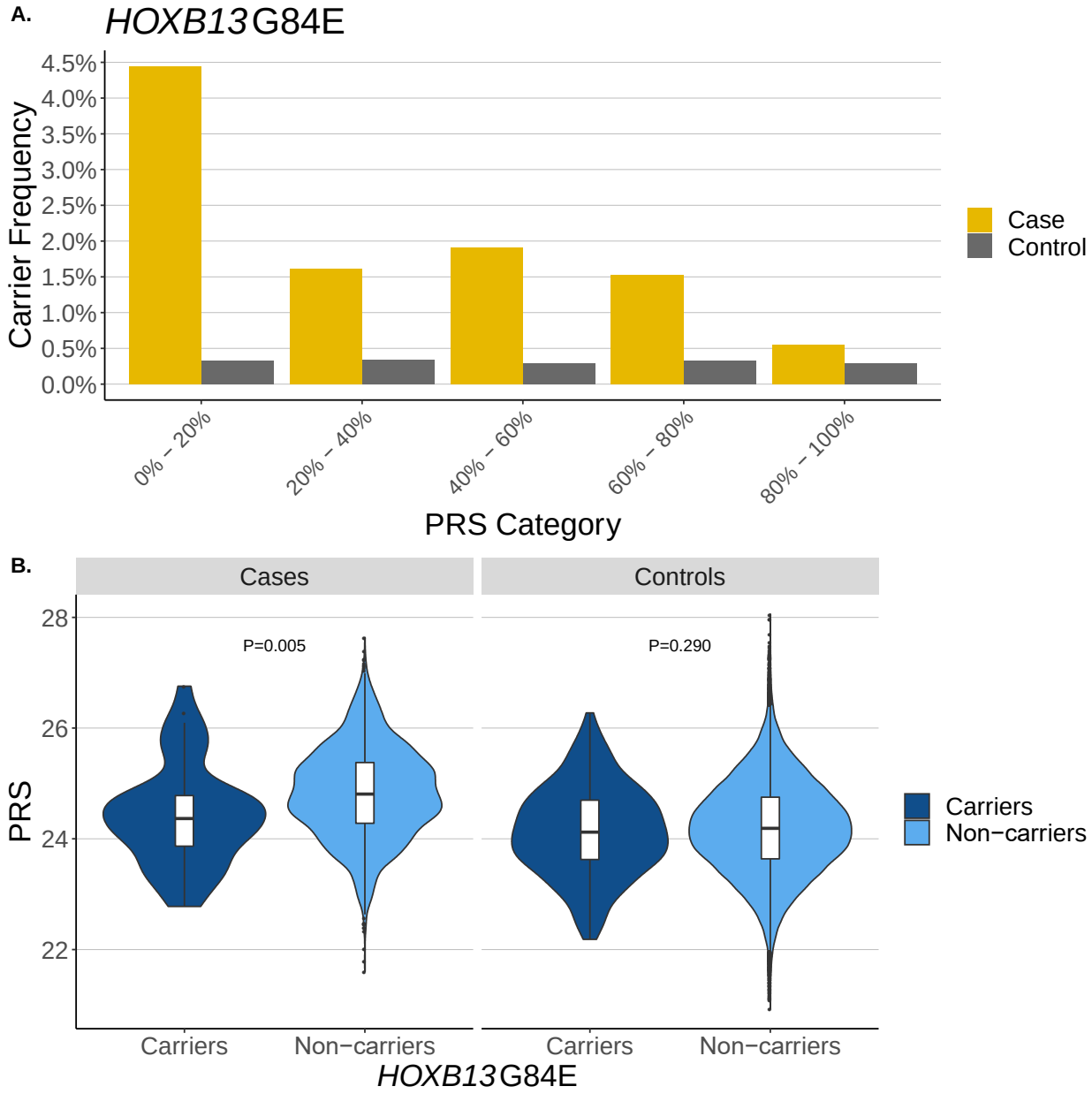






255

Figure 2.


A.
HOXB13G84E
Carrier Frequency
4.5%
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
0% – 20%
20% – 40%
40% – 60%
60% – 80%
80% – 100%
PRS Category
Case
Control
B.
Cases
Controls
P=0.005
P=0.290
PRS
28
26
24
22
Carriers
Non−carriers
Carriers
Non−carriers
HOXB13G84E
Carriers
Non−carriers