medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Mendelian randomization revealing the protective effect of sodium-glucose cotransporter 2 inhibition on prostate cancer with verified evidence from electronic healthcare and biological data

Jie Zheng, PhD[1,2,3]*[§], Jieli Lu, MD, PhD[1,2]*, Jiying Qi, PhD[1,2], Qian Yang, PhD[3], Huiling Zhao, MPhil[3], Haoyu Liu[1,2], Youqiong Ye [4,5], Min Xu, MD, PhD[1,2], Yu Xu, MD, PhD[1,2], Tiange Wang, MD, PhD[1,2], Mian Li, MD, PhD[1,2], Zhiyun Zhao, MD, PhD[1,2], Ruizhi Zheng, PhD[1,2], Shuangyuan Wang, PhD[1,2], Hong Lin, PhD[1,2], Chunyan Hu, MPH[1,2], Celine Sze Ling Chui, PhD[6,7,8], Shiu Lun Au Yeung, PhD[7], Shan Luo, PhD[7], Olympia Dimopoulou, MPH[3,9], Padraig Dixon, PhD[3,10], Sean Harrison, PhD[3,10], Yi Liu, PhD[3], Jamie Robinson, PhD[3], James Yarmolinsky, PhD[3], Philip Haycock, PhD[3], Jinqiu Yuan, PhD[12,13,14,15], Sarah Lewis, PhD[3], Tom R. Gaunt, PhD[3,16], George Davey Smith, MD, PhD[3,16], Ning Guang, MD, PhD[1,2], Richard M. Martin, MD, PhD[3,7,16§], Bin Cui, PhD[1,2], Weiqing Wang, MD, PhD[1,2§], Yufang Bi, MD, PhD[1,2§]

[1] Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2] Shanghai National Clinical Research Center for metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai Key Laboratory for Endocrine Tumor, Shanghai Digital Medicine Innovation Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3] MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom

[4] Center for Immune-Related Diseases at Shanghai Institute of Immunology, Department of Gastroenterology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[5] Shanghai Institute of Immunology, State Key Laboratory of Oncogenes and Related Genes, Department of Immunology and Microbiology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

[6] School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administration Region, China

[7] School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administration Region, China

[8] Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, Hong Kong Science and Technology Park, Hong Kong Special Administration Region, China

[9] Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom BS8 2PS

[10] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, OX2 6GG

[11] Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, United Kingdom

[12] Clinical Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, 518107, China

[13] Center for Digestive Disease, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, 518107, China

[14] Guangzhou Women and Children Medical Center, Guangzhou, Guangdong, 510623, China

[15] Division of Epidemiology, the JC School of Public Health & Primary Care, the Chinese University of Hong Kong. Hong Kong

[16] NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, United Kingdom

*** Equal first author**

**§ Senior and corresponding author**



**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Yufang Bi, Professor of medicine, Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, E-mail: byf10784@rjh.com.cn

Weiqing Wang, Professor of medicine, Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; Email: wqingw61@163.com

Richard M. Martin, Professor of Clinical Epidemiology, MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol. Population Health Sciences, Bristol Medical School, University of Bristol. NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, United Kingdom. Tel: +44 117 455 3987; Email: Richard.Martin@bristol.ac.uk

Jie Zheng, Professor in health informatics, Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; E-mail: zj12477@rjh.com.cn

Word count: abstract 333; main text 4,347.
Number of figures and tables: 4 figures and 2 tables.
Appendix materials: 3 figures,14 tables and 3 notes.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .


## Abstract

**Background** Observational studies indicated a decreased risk of prostate cancer by SGLT2 inhibitors, but high-quality evidence is lacking to make a clear conclusion. We evaluated the effect of SGLT2 inhibition on prostate cancer risk by triangulating evidence from three methods.

**Methods** Genetic variants associated with $HbA_{1c}$ levels ($P<5\times10^{-8}$) in the genomic region of the target gene, *SLC5A2*, were used to proxy SGLT2 inhibition. In discovery, Mendelian randomization (MR) was applied to estimate effects of genetically proxied SGLT2 inhibition on risk of prostate cancer and its subtypes (79,148 cases and 61,106 controls). In a validation using electronic healthcare data, the association of incidence of prostate cancer between 24,155 new users of SGLT2 inhibitors and 24,155 new users of the active comparator, dipeptidyl peptidase 4 inhibitors, was estimated using electronic health-care data. In a biological validation, the differential gene expression of *SLC5A2* between normal prostate tissue and tumour tissue were estimated in 691 prostate cancer patients. To validate the influence of glucose, the association between $HbA_{1c}$ levels and incident prostate cancer during 10-years of follow-up were estimated.

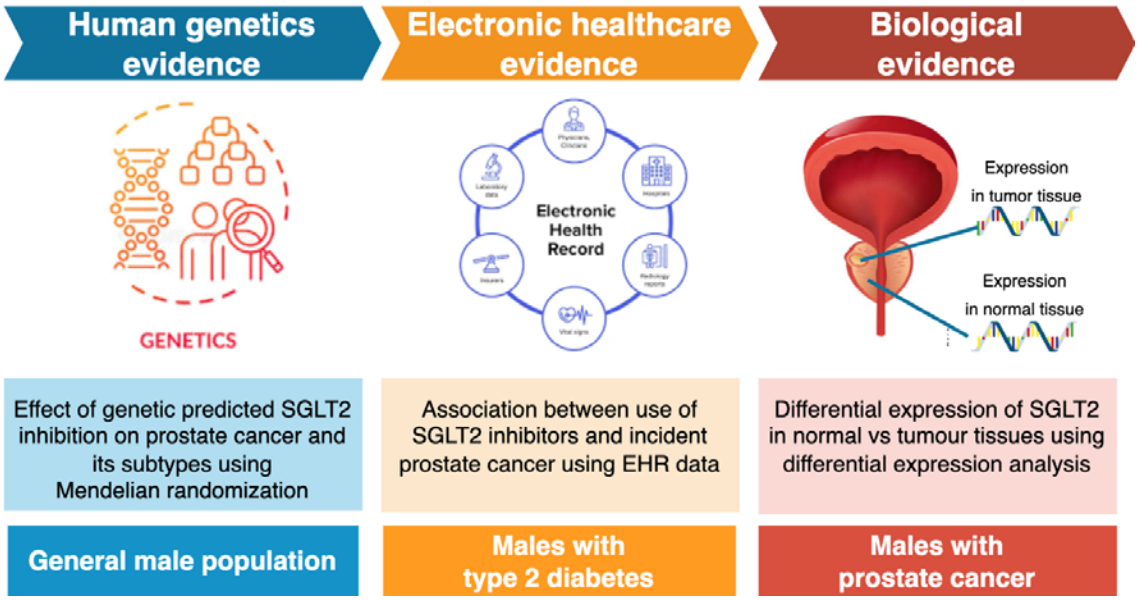
**Findings** For genetic evidence, genetically proxied SGLT2 inhibition reduced the risk of overall (odds ratio=0.56, 95%CI=0.38 to 0.82), advanced (OR=0.52, 95%CI=0.27 to 0.99) and early-onset (OR=0.27, 95%CI=0.11 to 0.72) prostate cancer. For electronic healthcare evidence, usage of SGLT2 inhibitor was associated with a 23% reduced risk of prostate cancer (hazard ratio=0.77, 95%CI=0.61 to 0.99) in males with diabetes. For biological evidence, expression levels of the *SLC5A2* gene in tumour prostate tissue was 2.02-fold higher than that in normal tissue (P=0.006). Genetically proxied $HbA_{1c}$ and observed $HbA_{1c}$ provided little evidence to support an association with total/incident prostate cancer, implies a non-glycemic effect of SGLT2 inhibition on prostate cancer.

**Interpretation** This study provides genetic, electronic healthcare, and biological evidence to support a beneficial effect of SGLT2 inhibition on reducing prostate cancer risk. Future trials are warranted to investigate whether SLGT2 inhibitors can been recommended for diabetic individuals with high risk of prostate cancer or considered as an anti-prostate cancer therapy.


**Funding** AMS, MRC and NSFC.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Graphical abstract



medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Research in context

### Evidence before this study

We searched PubMed, Embase and clinicaltrials.gov databases from inception up to July 11, 2023 using the search terms: “SGLT2 inhibitor”, “canagliflozin”, “dapagliflozin”, or “empagliflozin” and “prostate cancer” and “clinical trials”, without language restrictions. Some functionals studies provided evidence of SGLT2 inhibition on reduce the viability of prostate cancer cells but lack of human-based evidence. Only one clinical trial study is investigating the role of SGLT2 inhibitors on prostate cancer in individuals with diabetes. Other 46 trials of SGLT2 inhibitors set prostate cancer as secondary outcome, the prostate cancer cases were limited for these studies, power issues have prevented clear causal inference. Little has been done to establish the causal role of SGLT2 inhibition on total and incident prostate cancer.

### Added value of this study

In this study, Mendelian randomization (MR) was applied in 140,254 men (79,148 with prostate cancer), and suggested that genetically proxied SGLT2 inhibition showed an effect on 44%, 48% and 73% reduced risk of total-, advanced- and early-onset prostate cancer in the general male population. Validation analysis using electronic health-care record data (81,122 men with diabetes) suggested that usage of SGLT2 inhibitor was associated with a 23% reduced risk of prostate cancer in males with diabetes. The differential expression analysis in 639 men with prostate cancer showed that the expression of SGLT2 was 2.02 folds higher in prostate cancer tissues compared with that in surrounding normal prostate tissues. As a benchmark, MR and observational analyses showed little evidence to support an effect of HbA1c on prostate cancer, which suggests a potential non-$HbA_{1c}$ effect of SGLT2 inhibition on prostate cancer.

### Implication of all the available evidence

There were multiple sources of evidence to support a protective role of genetically proxied SGLT2 inhibition and usage of SGLT2 inhibitors on risk of prostate cancer in men with and without diabetes and/or prostate cancer. Future clinical trials should be prioritised for investigation of the long-term use of SGLT2 inhibitors in the prevention and treatment of prostate cancer.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Introduction

Diabetes is one of the most common chronic conditions that affecting 537 million individuals in 2021[1]. Among various types of anti-diabetic drugs, recent clinical trials showed the beneficial effect of sodium-glucose cotransporter 2 (SGLT2) inhibitors on reduced risk of atherosclerotic cardiovascular disease (ASCVD) in addition to improvements in $HbA_{1c}$[2–4]. Based on the trial evidence, the American Diabetes Association (ADA) and European Association for the Study of Diabetes (EASD) guidelines recommended SGLT2 inhibitors as first-line therapy for patients with or with indicators of high risk of ASCVD, heart failure, or chronic kidney disease (CKD) since 2020[5,6,7]. It has now been widely used by clinicians from endocrinology and cardiology departments.

Cancer was known to be one of the common comorbidities for type 2 diabetes[8,9,10]. Among various cancer types, prostate cancer is the second most commonly diagnosed male cancer, with nearly 1.41 million new diagnosis worldwide in 2020, and is a major cause of cancer death in men[11]. However, no clinical guideline recommended usage of anti-diabetic drugs on individuals with cancers or in high risk of cancers, especially for males with these two common diseases, diabetes and prostate cancer. A recent review summarized the anti-cancer mechanisms of SGLT2 inhibitors[12]. Observational studies have also reported a decreased risk of prostate cancer amongst men with diabetes who are taking SGLT2 inhibitors[13]. However, the largest meta-analysis of randomized controlled trials (RCTs) in individuals with T2DM suggested little difference in prostate cancer incidence between users of SGLT2 inhibitors and users of placebo or active comparators[14], although in this study power could be an issue due to limited number of incidence prostate cancer cases (N=41) included in the analysis. Correctively, existing epidemiology studies provided some clues but the evidence level to support the protective effect of SGLT2 inhibition on prostate cancer risk was still low. Whether SGLT2 inhibition can be recommended for diabetic individuals with high risk of cancers or can be even repurposed as an anti-cancer drug target need further investigation.

Evidence triangulation is the practice of obtaining more reliable answers to research questions through integrating results from several different methods[15]. These methods have different assumptions and unrelated sources of biases. If results of these methods point to a similar conclusion, this will strengthen confidence in the finding. For the causal question of identifying the effect of a drug target on a disease, human genetics, epidemiology and bioinformatics are three commonly used approaches[16,17,18,19,20,21]. Triangulating evidence from these methods in one study may provide an attractive strategy to improve evidence level for drug repurposing. Mendelian randomization (MR) is a method that uses germline genetic variants as proxy measures of exposure and estimates the causal effect of an exposure on an outcome[22]. An individual's germline genetic make-up influences their biology from conception, meaning that causal estimates from MR studies reflect lifelong exposures (e.g. lifelong SGLT2 inhibition) and that the estimates are not generally susceptible to reverse causation or confounding[23]. Observational associations of the use of a drug on disease incidence are normally estimated using Cox proportional hazard models, where a 'new user active comparators' design may reduce the influence of confounders[24]. Differential expression analysis is a commonly used approach to estimate the association of expression of a target gene on a disease, which is commonly used to identify genetic mechanisms for cancers. Due to enriched data resources to support application of all three methods[25,26,27], studying the effect of SGLT2 inhibition on prostate cancer is a preferred example for such evidence triangulation strategy.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The objective of this study was to estimate the causal effects of SGLT2 inhibition on prostate cancer and its subtypes by triangulating evidence from human genetics, electronic healthcare and biological data. The effect of $HbA_{1c}$ on prostate cancer were further estimated using human genetics and observational epidemiology approaches.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

# Methods

## Summary of study design and data sources

The **graphical abstract** and **Figure 1** present an overview of three sets of analyses conducted in this study. Each analysis aims to answer the same causal question in different subpopulations. All studies contributing data to this analysis had the relevant institutional review board approval from each country and all participants provided informed consent.

First, the human genetics analysis was applied as the discovery analysis in the general male population. We estimated the putative causal effects of SGLT2 inhibition and genetically predicted $HbA_{1c}$ on risks of prostate cancer and its subtypes using MR. The summary genetic association data from a case-control genome-wide association study (GWAS) of prostate cancer in the PRACTICAL and GAME-ON/ELLIPSE Consortium[25,26] were used (N=140,254 men from the general population).

Second, as a validation of the MR findings, two sets of analyses were conducted using two independent large-scale cohort resources: (1) the association of use of SGLT2 inhibitors on incident prostate cancer were estimated in diabetic individuals using data derived from electronic health record data in the Shanghai Link Healthcare Database (SLHD; N=81,122 men with diabetes), a representative clinical database covering electronic health-care records for over 99% of Shanghai residents since 2013[28]; (2) the association of baseline $HbA_{1c}$ levels with incident prostate cancer during 10-years of follow-up was estimated using data from the Risk Evaluation of Cancers in Chinese Diabetic Individuals (REACTION) study[8] (N=57,779 men for the general population). Both human genetics and observational analyses were related to prostate cancer risk, which are related to disease prevention.

Third, the biological analysis was conducted to understand the biological function of SGLT2 expression on prostate cancer in patients with prostate cancer. The differential gene expression analysis using data from the TCGA Program[27] was conducted (https://portal.gdc.cancer.gov/; N=639 men with prostate cancer). This analysis was conducted in prostate cancer patients, which may inform cancer treatment.

## Statistical analysis

### Discovery analyses using Mendelian randomization

#### *Genetic instrument selection for SGLT2 inhibition*

To generate genetic instruments to proxy the lifelong glucose-lowering effect of SGLT2 inhibition, summary data were obtained from a GWAS of $HbA_{1c}$ levels in UK Biobank (N=344,182). Seven genetic variants robustly and independently associated with $HbA_{1c}$ ($P<5\times10^{-8}$; square of linkage disequilibrium [LD] among instruments [$r^2$] <0.1) in the *SLC5A2* region were selected as instruments to proxy SGLT2 inhibition (noted as primary instruments; **Table 1**). Since prostate cancer is male-specific, the genetic associations of the selected instruments on $HbA_{1c}$ from 159,160 UK Biobank males were used in the MR analysis (**Appendix Table 1A**).

As a validation, we used a more stringent instrument selection pipeline, as described before[29,30], which hypothesized that a genetic variant is a valid instrument for SGLT2 inhibition if this variant showed genetic colocalization evidence[31,32] on both $HbA_{1c}$ and expression levels of the target gene of SLGT2, *SLC5A2*[33,34]. After applying this process, two variants were selected as instruments (noted as stringent instruments; **Table 1**, **Appendix Table 2**).

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We further validated the reliability of the instruments using an independent dataset. The genetic effects of the SGLT2 instruments were extracted from a European-only GWAS of $HbA_{1c}$ within the MAGIC consortium[35] (N=146,806 **Appendix Table 3**). To instrument $HbA_{1c}$ levels, 315 independent variants (LD $r^2<0.001$) associated with $HbA_{1c}$ ($P<5\times10^{-8}$) were obtained from UK Biobank male only GWAS data (**Appendix Table 1B**).

More details of the instrument selection are described in **Appendix Note 1** and **Appendix Figure 1**).

***Outcome selection for human genetics analysis***

Eight prostate cancer related phenotypes were selected as outcomes for the MR analysis: total-, aggressive-, early-onset-, high aggressive vs low aggressive-, high aggressive vs low and intermediate aggressive-, advanced stage vs localised stage prostate cancer. Advanced prostate cancer was defined as metastatic disease or Gleason score (GS) ≥ 8 or PSA > 100 or prostate cancer death; early-onset refers to prostate cancer onset before age 55; low aggressive refers to T stage from the TNM staging ≤ T1, and GS ≤ 6, and PSA<10; intermediate aggressive refers to T stage: T2, and GS=7, and PSA 10~20; and high aggressive refers to T stage: T3/T4 or N1 or M1 or GS ≥ 8 or PSA >20. PSA levels were included as they drive prostate cancer diagnoses, and we wanted to exclude an effect of the exposures on PSA that could bias the prostate cancer associations. Summary data were obtained from GWAS results in the PRACTICAL and GAME-ON/ELLIPSE consortia or Kachuri et al.[25,36]. Detail information of the prostate cancer related outcomes were listed in **Appendix Table 4**.

***Mendelian randomization analyses***

Germline genetic variants used to proxy SGLT2 inhibition were matched to prostate cancer datasets by orienting effects of the exposure and the outcome to the same effect allele. If an instrument was missing in the outcome dataset, a genetic variant with high LD ($r^2>0.8$) to the instrument was selected as a proxy instrument where possible. An inverse-variance weighted approach was used to combine variant-level Wald ratio estimates into an overall effect estimate[37]. All MR estimates (odds ratios [ORs]) were scaled to SD unit to reflect the equivalent of a one SD unit (0.62%) reduction in $HbA_{1c}$.

A set of Mendelian randomization In the main MR analyses, the effects of genetically proxied SGLT2 inhibition (using seven primary instruments) were estimated on total prostate cancer, its subtypes and PSA levels in the general male population (PRACTICAL and GAME-ON/ELLIPSE)[25]. The effect of SGLT2 inhibition on T2DM[38] was estimated as a positive control analysis. For the validation MR analyses, the effects of SGLT2 inhibition on the prostate cancer related outcomes were estimated using the stringent instruments and instruments from the independent dataset (MAGIC).

We report findings according to the STROBE-MR (Strengthening the Reporting of Mendelian Randomization Studies) guidelines[39,40] (the STROBE-MR check list attached). MR has three key assumptions: (1) the germline genetic instruments used to proxy SGLT2 inhibition are robustly associated with the exposure ("relevance"); (2) the instruments are not associated with common causes (confounders) of the instrument-outcome association (exchangeability); (3) the instruments are only associated with the outcome through the exposure under study ("exclusion restriction"). These MR assumptions were tested using the sensitivity methods, including generalized inverse variance weighted (gIVW)[41], genetic colocalization[31,32] , phenome-wide association studies (which including classic risk factors

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

associated with SGLT2 instruments) using data from the IEU OpenGWAS database[26], heterogeneity tests across instruments using Cochran's Q[42], weighted median[43] and mode-based estimate approaches[44] and Multivariable MR[45]. Extended descriptions of these sensitivity analyses are provided in the **Appendix Note 3**. Moreover, the SGLT2 instruments were associated with other 17 genes. Associations between expression of these genes and prostate cancer may reflect pleitropy, but hard to be tested. Here, we estimated whether expression levels of these genes were different in normal tissue versus prostate tumour tissue using DESeq2[46] (data from TCGA), which provided additional evidence to distinguish pleiotropy. For all MR analyses, Bonferroni corrections were applied to establish multiple testing-adjusted thresholds. All the MR analyses were conducted using the TwoSampleMR R package v0.5.6[47].

**Validation using electronic healthcare data and cohort with long-term follow-up**

To triangulate with the MR findings, we estimated the association between the use of SGLT2 inhibitors and the incidence of prostate cancer during the follow-up period (**Figure 2A**; data from the Shanghai Link Healthcare Database [SLHD]). To emulate a trial protocol using large-scale observational data[48], we setup stringent eligibility criteria and used a 'new user active comparator' design (more details in **Appendix Note 2**). To reduce the influence of potential confounders, dipeptidyl peptidase 4 (DPP4) inhibitors were selected as active comparators since it is a newer second-line anti-diabetic drug that had been compared with SGLT2 inhibitors in previous clinical trials[49]. Within SLHD, 130,817 males with T2DM newly treated with SGLT2 inhibitors (or DPP4 inhibitors) were selected. The look back period for this study was from 1st of Jan 2013 till cohort entry (which was the first time a male had an outpatient visit record in any tier-three hospital in Shanghai). After selection, 81,122 males who fitted the eligibility criteria were selected as the original cohort of this study. The follow-up period during which prostate cancer diagnosis was recorded was 1st of March 2017 (when SGLT2 inhibitors was marketed in China) till 31st of December 2021 (median follow-up period was 1.33 years). Propensity-score matching was conducted using 30 covariates, aiming to balance the baseline characteristics between the two treatment groups and reduce confounding effects: 12 comorbidities, 7 anti-diabetic drugs and 11 other medications were used in building up the propensity score (**Appendix Table 5**). An adjusted Cox proportional hazard's model was used and hazard ratio (HR) with 95% confidence intervals (CIs) were calculated to estimate the risk of incident prostate cancer among new users of SGLT2 inhibitors as compared with new users of dipeptidyl peptidase 4 (DPP4) inhibitors.

For the second observational analysis using data from the REACTION study[8], we defined the outcome as incident prostate cancer during the 10-years of follow-up period. Prostate cancer was defined using ICD 10 code "C61" for the two studies. We estimated the observational association between baseline $HbA_{1c}$ and incident prostate cancer during 10 years of follow-up in the REACTION study (223 cases and 57,556 controls)[8]. $HbA_{1c}$ levels were measured at baseline and used as a continuous variable with per SD increasement in the analysis (mean=6.04%, SD=1.11%). Age, body mass index, tobacco consumption, alcohol consumption, physical activity, and diet score were included as covariates in the Cox proportional hazards model (more details in **Appendix Note 2**).

**Biological validation using gene expression data in cancer patients**

To dig into functional mechanism, a differential gene expression analysis was carried out using DESeq2[46] to understand the expression patterns of SGLT2 (counts) in normal prostate tissue versus prostate tumour tissue (data from GDC data portal

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

https://portal.gdc.cancer.gov/). After selection, 691 prostate tissue samples from 639 individuals, involving 639 prostate tumour samples and 52 normal solid tissues, were selected from three databases: TCGA, West Coast Prostrate Cancer Dream Team (WCDT) and Count Me In (CMI) Project. The dataset batch was set as a covariate in the model to adjust for batch effect.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Results

**Discovery: MR effects of SGLT2 inhibition on prostate cancer risk**

The characteristics of the primary and stringent genetic instruments used to proxy SGLT2 inhibition are listed in **Table 1** and **Appendix Table 1-3,** respectively. Across these exposures, the F-statistics used to test the relevance MR assumption suggests that weak instrument bias was unlikely to be an issue in this study.

Genetically proxied SGLT2 inhibition (estimated by primary instruments), equivalent to a one SD (0.62%) reduction in $HbA_{1c}$, reduced the risk of total prostate cancer by 44% (OR=0.56, 95%CI=0.38 to 0.82, P=0.003; **Table 2**). This effect was consistent across the seven instruments (heterogeneity P=0.80; **Figure 3**). The other four sensitivity MR models showed similar effect estimates (**Appendix Figure 2**).

Genetically proxied SGLT2 inhibition lowered the risk of advanced (OR=0.52, 95%CI=0.27 to 0.99; P=0.049) and early-onset (OR=0.27, 95%CI=0.11 to 0.71; P=0.008) prostate cancer. Little evidence was observed to support an effect of SGLT2 inhibition on other prostate-cancer related outcomes (**Table 2**). In addition, there was little evidence to support an effect of SGLT2 inhibition on PSA levels (β=0.14, 95%CI=0.030 to -0.30, P=0.107; **Appendix Table 6**), which suggested that SGLT2 inhibition is likely to show an effect on reducing risk rather than influencing the diagnostic work up for prostate cancer. As a positive control, we confirmed the well-established effect of SGLT2 inhibition on reducing the risk of T2DM (OR=0.66, 95%CI=0.49 to 0.88, P=0.005; **Appendix Table 6**).

The validation MR analysis using the two instruments selected by the stringent approach and using SGLT2 instruments derived from the MAGIC consortium validated the effect of SGLT2 inhibition on total, advanced and advanced vs localised prostate cancer (**Figure 3** and **Appendix Table 7**).

Tests of MR assumptions

The exchangeability MR assumption was tested using genetic colocalization between SGLT2 inhibition and prostate cancer, where we observed evidence of colocalization of the two traits in the *SLC5A2* region (colocalization probability=72%; **Appendix Table 8**).

The exclusion restriction MR assumption was examined in several analyses. The PheWAS of the primary SGLT2 instruments showed that these genetic variants were associated with blood cell traits (e.g. red blood cell counts), body weight traits (e.g. waist circumference), diastolic blood pressure and low-density lipoprotein cholesterol (**Appendix Table 9**). Multivariable MR adjusting for these traits respectively (**Appendix Table 10A**) suggested that the effect of SGLT2 inhibition on prostate cancer was independent of these traits (**Appendix Table 10B**). In addition, the SGLT2 instruments showed associations with the expression of 17 genes excluding *SLC5A2*, with two genes being targets for existing drugs for coagulation and hemoglobinuria treatment. The 17 genes were not associated with glycemic traits or to have an interaction with any anti-diabetic or anti-cancer drugs[50] (**Appendix Table 11**). The differential expression analysis using data from the TCGA Program showed little evidence of differential expression levels in prostate tumour tissue versus normal tissue for 15 of the 17 genes (except *SLC5A2*) (**Appendix Table 12**), implying that pleiotropy was less likely occur via these genes.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The MR sensitivity analyses did not provide strong evidence of heterogeneity or pleiotropy for the effect of SGLT2 inhibition on prostate cancer, but the statistical power to clearly demonstrate this was low (**Appendix Table 6** and **7**).

**Validation of association of usage of SGLT2 inhibitors with prostate cancer risk using electronic healthcare data**

We identified 26,988 new users of SGLT2 inhibitors and 54,134 new users of DPP4 inhibitors who fulfilled the eligibility criteria out of 130,817 males from SLHD (**Figure 2A**). After a 1:1 propensity-score matching, we identified a cohort of 48,310 patients (24,155 in each group) with well-balanced baseline characteristics (standardized mean differences less than 1.5%) between the two treatment groups (**Appendix Table 5A**). Cox proportional hazards model showed that SGLT2 inhibitors use (compared with DPP4 inhibitors use) was associated with a 23% reduction in risk of prostate cancer (SGLT2 inhibitors use=467.4 versus DPP4 inhibitors use=492.75 per 100,000 person-years; HR=0.77, 95%CI=0.61 to 0.99, P=0.03) during a median follow-up of 1.33 years (**Figure 2B**). Sensitivity analyses lagging the outcome period between one to six months showed similar protective effects, albeit less precisely estimated (**Appendix Table 13**).

**Biological validation: association of *SLC5A2* expression on prostate cancer**

Differential expression analysis showed that the expression levels of *SLC5A2* was 2.02 folds higher in prostate cancer tissue than in normal prostate tissue (P=0.006; **Figure 4**). Two additional genes, *TGFB1I1* and *ITGAD*, also showed differential expression in normal versus tumour tissues (*TGFB1I1*: log2FoldChange=-1.65, $P=1.95\times10^{-7}$; *ITGAD*: log2FoldChange=1.47, P=0.003; **Appendix Table 12**). Although these two genes could be considered as potential anti-prostate cancer drug target, they were not reported to be associated with any diabetic or glycemic traits or considered as anti-diabetic drugs. Therefore, they were not likely to be pleiotropic exposures that linking SGLT2 instruments with prostate cancer.

**Validating the influence of glucose: MR and observational association of $HbA_{1c}$ with prostate cancer**

We estimated the association of $HbA_{1c}$ with prostate cancer risk using MR and observational analyses, which aimed to investigate whether the effect of SGLT2 inhibition on prostate cancer is partly via lowering $HbA_{1c}$ levels. Little evidence was observed to support the effect of genetically proxied $HbA_{1c}$ on change prostate cancer risk (OR=0.98, 95%CI=0.92 to 1.05, P=0.63; **Table 3**). Sensitivity MR analyses in which we removed variants within the *SLC5A2* region showed similar effects to those seen in our analyses of $HbA_{1c}$ on prostate cancer (**Appendix Table 14**). Observational analysis in the REACTION study also provided little evidence to support the effect of baseline $HbA_{1c}$ levels on incident prostate cancer after 10 years of follow-up (HR=0.93, 95%CI=0.80 to 1.10, P=0.40), the findings barely change after excluding individuals using anti-diabetic drugs (**Table 3**).

## Discussion

In this study, we triangulated human genetics, electronic healthcare and biological evidence to answer the same causal question: the effect of SGLT2 inhibition on prostate cancer. In the discovery, we observed that genetically proxied lifelong SGLT2 inhibition reduced total-, advanced- and early-onset prostate cancer in the general male population by 44%, 48% and 73%, respectively. Validation using various selection processes and datasets confirmed the protective effect of SGLT2 inhibition on risk of prostate cancer and its subtypes, rather than an effect on PSA biasing the diagnosis of prostate cancer. In the validation using electronic

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

healthcare data, we showed that SGLT2 inhibitor use reduced risk of prostate cancer by 23% in men with T2DM. In the biological validation, we showed that expression levels of SGLT2 were 2.02 times higher in prostate tumour tissue than that in normal prostate tissue. In the analyses validating the influence of glucose, we found little genetic and observational evidence to support an association of $HbA_{1c}$ on prostate cancer, which implies a possible non-glucose mechanism of SGLT2 inhibition on prostate cancer prevention (see **Appendix Figure 3**). Correctively, we provide three strands of evidence to prioritise SGLT2 inhibition as a target for prostate cancer prevention and adjunctive treatment.

According to the U.S. Centers for Disease Control and Prevention (CDC), adults aged between 45 and 64 receive the greatest number of new diagnoses of diabetes, which was also the age group that men are likely to receive diagnoses of prostate cancer. However, there was little evidence to support setting up of clinical guidelines concerning modification of SGLT2 inhibitor treatment among diabetic patients with co-existing prostate cancer till now. A few epidemiology studies supported the protective role of SGLT2 inhibitors on prostate cancer risk[13,51]. A recent systematic review of RCTs provided weak evidence of an effect of SGLT2 inhibitors on cancers[52]. Only one phase I trial was registered in clinicaltrial.gov (NCT04887935), which aims to investigate the safety of dapagliflozin, one type of SGLT2 inhibitors, for men considered at high risk of prostate cancer. In the present study, we observed robust human genetics and electronic healthcare evidence to support the effect of SGLT2 inhibition on reducing risk of prostate cancer, both in the general male population and in males with diabetes. Our results further support that SGLT2 inhibition may have better efficacy on prevention of early-onset prostate cancer than on total- and advanced prostate cancer. Our evidence supports the prioritization of future clinical trials of SGLT2 inhibitors in diabetic men at high risk of prostate cancer, which may have the potential to influence clinical guidelines/standards for diabetes.

It has been hypothesized that the primary mechanism of a beneficial effect of SGLT2 inhibitors on cancer is through inhibiting glycolysis in tumor cells, thus reducing tumor cell proliferation and tumorigenesis[53]. Another study showed that Canagliflozin, one type of SGLT2 inhibitor, inhibits mitochondrial complex-I and cellular proliferation in prostate cancer cells[54]. However, the lack of MR and observational evidence of a role for $HbA_{1c}$[29] suggests that $HbA_{1c}$ may not be driving the observed association of SGLT2 inhibition with prostate cancer. Recent *in-vitro* and *in-vivo* studies have shown that SGLT2 antibodies are expressed in prostate adenocarcinomas but are rarely expressed in normal prostate cells surrounding cancer cells[55]. But these studies were conducted in a small number of prostate samples, raising the question of generalizability of the findings. In our study, we harmonized data from over 600 prostate tumour tissues and over 50 normal prostate tissues. Our analysis supports an association of *SLC5A2* expression and prostate cancer in males with prostate cancer. Correctively, our genetic and biological evidence imply that SGLT2 inhibition may have a direct effect on prostate cancer prevention and treatment, which could be independent to its glucose control effect. Some well-designed clinical trials have also provided evidence to support that SGLT2 inhibitors have good tolerance and safety profiles to be used in individuals without diabetes[56,57]. Further functional and clinical studies are warranted to better understand the anti-cancer mechanism of SGLT2 inhibitors and test their anti-prostate cancer efficacy in individuals without diabetes.

**Strengths and limitations**

Our study has several strengths. First, we estimated the effects of SGLT2 inhibition on prostate cancer prevention and treatment using genetic, epidemiological and biological

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

approaches, which have different assumptions, key source of biases (e.g. pleiotropy for MR and confounders for observational analysis)[15] and different subgroup of population (i.e. the general male population, males with diabetes and males with prostate cancer). Triangulation of evidence suggests that SGLT2 inhibition is likely to have a protective effect on prostate cancer in all subpopulation groups, which strengthens confidence in this finding. Second, the instruments for SGLT2 were selected using two widely applied pipelines[18,29]. The reliability of these instruments has been tested thoroughly in this study. Third, we paid special attention to the potential influence of our genetic variant-exposure estimates on our MR results and only used male-specific instruments in this study. Fourth, the results from colocalization analysis, PheWAS, multivariable MR, and other sensitivity MR analyses suggested that the effect of SGLT2 inhibition on prostate cancer is unlikely to violate the exchangeability and the exclusion restriction assumptions of MR. More interestingly, we extended the scope of differential gene expression analysis to distinguish pleiotropy from causality, which the strategy can be widely applied to other drug target genes and complex diseases.

This study has several limitations. First, our MR estimates of the effect of SGLT2 inhibition were scaled to represent the on-target reductions in $HbA_{1c}$ levels rather than the direct effect of SGLT2 inhibitors. This assumes that SGLT2 inhibition has a proportional impact on lowering of $HbA_{1c}$. Second, caution is needed to interpret the causal effect estimate from this study. This is because the MR estimate reflects the long-term modulation of drug targets on disease risk, which may suggest different levels of risk reductions per unit change in drug target compared with those observed from clinical trials/observational studies over a relatively short duration, which would explain the attenuated effect estimate of our observational analysis. Furthermore, the estimated effect of SGLT2 inhibition on prostate cancer could at least in part be influenced by different ancestries, disease status and survival bias given the relatively late age-at-onset of prostate cancer. Fourth, the MR analyses presented assumes no gene-environment interaction of the association of genetic proxies for drug targets and prostate cancer. Fifth, SGLT2 inhibitors have been marketed in China since March 2017, the median follow-up time for the observational analysis was therefore only 1.33 years. Therefore, we consider this result as a validation for evidence triangulation rather than a stand-alone finding.

## Conclusions

Genetic, epidemiological, and biological evidence with different assumptions and used different subpopulations support the role of SGLT2 inhibition on reducing prostate cancer risk. Further clinical trials should be prioritised to establish whether there is a similar effect with the long-term prescription of SGLT2 inhibitors, at what age chemoprevention/treatment would need to commence, whether high-risk men should be targeted and the potential harms.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Acknowledgements**

J.Z. is supported by the Academy of Medical Sciences (AMS) Springboard Award, the Wellcome Trust, the Government Department of Business, Energy and Industrial Strategy (BEIS), the British Heart Foundation and Diabetes UK (SBF006\1117). J.Z. is funded by the Vice-Chancellor Fellowship from the University of Bristol. J.L.L., M.X., G.N., and Y.B. are supported by the National Natural Science Foundation of China (82088102, 81970691, 81970728, 81930021 and 81941017) and Shanghai Clinical Research Center for Metabolic Disease (19MC1910100). J.L.L., M.X., Y.X., T.W., M.L., Z.Z., S.W., H.L., G.N., W.W., and Y.B. are members of the Innovative Research Team of High-level Local Universities in Shanghai. G.N. and Y.B. are supported by the Shanghai Shenkang Hospital Development Center (SHDC12019101, SHDC2020CR1001A, SHDC2020CR3064B). V.W., G.D.S. and T.R.G. are supported by the UK Medical Research Council Integrative Epidemiology Unit (MC_UU_00032/03). R.M.M, S.H. and P.D. received support from a Cancer Research UK (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Grant). R.M.M., T.R.G. & G.D.S. are also supported by the NIHR Bristol Biomedical Research Centre which is funded by the NIHR and is a partnership between University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. Department of Health and Social Care disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. JY is supported by a Cancer Research UK Population Research Postdoctoral Fellowship (C68933/A28534). R.M.M. is a National Institute for Health Research Senior Investigator (NIHR202411). GDS reports Scientific Advisory Board Membership for Relation Therapeutics and Insitro.

We thank the individual patients who provided the sample that made data available; without them the study would not have been possible. We thank for PRACTICAL and GAME-ON/ELLIPSE consortia provided key genetic data to support this study.

**Author contributions**

J.Z., G.N., R.M., W.W., and Y.B. designed the study, wrote the research plan, and interpreted the results. J.Z. undertook the main, replication and sensitivity MR analyses with feedback from Y.Q., O.D., J.Y., and J.R. B.C. and J.Q. collected data from the Shanghai Link Healthcare Database and conducted the survival and linear regression analyses, with C.S.L.C., S.L.A.Y., S.L., and J.Y. provided critical suggestions. H.L.Z. conducted the computational biology analyses using TCGA, with supports from Y.Y.Q. J.Z. and J.L. wrote the first draft of the manuscript with critical comments and revision from M.X., Y.X., T.W., M.L., Z.Z., R.Z., S.W., H.L., C.Y.H., C.S.L.C., S.L.A.Y., S.L., O.D., P.D., S.H., Y.L., J.R., J.Y., P.H., J.Y., S.L., T.R.G., G.D.S., R.M., W.W., Y.B., and G.N. J.Z. is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Data and materials sharing**

The data, analytic methods, and study materials will be made available to other researchers for purposes of reproducing the results. In more details, the genetic association data of the selected risk factors are available in Appendix Tables. The summary level GWAS statistics for the primary and secondary outcomes are available from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). UK Biobank received ethical approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/0382). The analytical script of the MR analysis that had been used in this study is available via the GitHub repository of the TwoSampleMR R package (https://github.com/MRCIEU/TwoSampleMR/).

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Ethical approval**

UK Biobank has received ethical approval from the UK National Health Service's National Research Ethics Service (ref 11/ NW/0382). All other studies contributing data to this analysis had the relevant institutional review board approval from each country and all participants provided informed consent.

**Transparency**

The lead author (J.Z.) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Reference


1 International Diabetes Federation. IDF Diabetes Atlas, 10th edition. https://diabetesatlas.org/atlas/tenth-edition/.

2 GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; **390**: 1151–210.

3 Neal B, Perkovic V, Mahaffey KW, *et al.* Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med* 2017; **377**: 644–57.

4 Wiviott SD, Raz I, Bonaca MP, *et al.* Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2019; **380**: 347–57.

5 Buse JB, Wexler DJ, Tsapas A, *et al.* 2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2020; **43**: 487–93.

6 American Diabetes Association. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 2021; **44**: S111–24.

7 American Diabetes Association. Introduction: Standards of Medical Care in Diabetes-2022. *Diabetes Care* 2022; **45**: S1–2.

8 Bi Y, Lu J, Wang W, *et al.* Cohort profile: risk evaluation of cancers in Chinese diabetic individuals: a longitudinal (REACTION) study. *J Diabetes* 2014; **6**: 147–57.

9 Goodwin PJ, Chen BE, Gelmon KA, *et al.* Effect of Metformin vs Placebo on Invasive Disease-Free Survival in Patients With Breast Cancer: The MA.32 Randomized Clinical Trial. *JAMA* 2022; **327**: 1963–73.

10 Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. *JAMA* 1995; **273**: 1605–9.

11 Sung H, Ferlay J, Siegel RL, *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; **71**: 209–49.

12 Dutka M, Bobiński R, Francuz T, *et al.* SGLT-2 Inhibitors in Cancer Treatment-Mechanisms of Action and Emerging New Perspectives. *Cancers* 2022; **14**. DOI:10.3390/cancers14235811.

13 Murtola TJ, Tammela TLJ, Lahtela J, Auvinen A. Antidiabetic medication and prostate cancer risk: a population-based case-control study. *Am J Epidemiol* 2008; **168**: 925–31.

14 Tang H, Dai Q, Shi W, Zhai S, Song Y, Han J. SGLT2 inhibitors and risk of cancer in type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. *Diabetologia* 2017; **60**: 1862–72.

15 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; **45**: 1866–86.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

16 Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat Rev Cardiol* 2021; **18**: 435–53.

17 Liu J, Lichtenberg T, Hoadley KA, *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 2018; **173**: 400-416.e11.

18 Yarmolinsky J, Bull CJ, Vincent EE, *et al.* Association Between Genetically Proxied Inhibition of HMG-CoA Reductase and Epithelial Ovarian Cancer. *JAMA* 2020; **323**: 646–55.

19 Haycock PC, Burgess S, Nounu A, *et al.* Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol* 2017; **3**: 636–51.

20 Soni PD, Hartman HE, Dess RT, *et al.* Comparison of Population-Based Observational Studies With Randomized Trials in Oncology. *J Clin Oncol* 2019; **37**: 1209–16.

21 Garnett MJ, Edelman EJ, Heidorn SJ, *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012; **483**: 570–5.

22 Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.

23 Sanderson E, Glymour MM, Holmes MV, *et al.* Mendelian randomization. *Nature Reviews Methods Primers* 2022; **2**: 1–1.

24 Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep* 2015; **2**: 221–8.

25 Schumacher FR, Al Olama AA, Berndt SI, *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018; **50**: 928–36.

26 Elsworth B, Lyon M, Alexander T, *et al.* The MRC IEU OpenGWAS data infrastructure. 2020; published online Aug 10. DOI:10.1101/2020.08.10.244293.

27 Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015; **163**: 1011–25.

28 Qi J, He P, Yao H, *et al.* Cancer risk among patients with type 2 diabetes: A real-world study in Shanghai, China. *J Diabetes* 2019; **11**: 878–83.

29 Au Yeung SL, Schooling CM. Impact of glycemic traits, type 2 diabetes and metformin use on breast and prostate cancer risk: a Mendelian randomization study. *BMJ Open Diabetes Res Care* 2019; **7**: e000872.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

30 Luo S, Schooling CM, Wong ICK, Au Yeung SL. Evaluating the impact of AMPK activation, a target of metformin, on risk of cardiovascular diseases and cancer in the UK Biobank: a Mendelian randomisation study. *Diabetologia* 2020; **63**: 2349–58.

31 Giambartolomei C, Vukcevic D, Schadt EE, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10**: e1004383.

32 Zheng J, Haberland V, Baird D, *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020; **52**: 1122–31.

33 GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020; **369**: 1318–30.

34 Võsa U, Claringbould A, Westra H-J, *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021; **53**: 1300–10.

35 Chen J, Spracklen CN, Marenne G, *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* 2021; **53**: 840–60.

36 Kachuri L, Hoffmann TJ, Jiang Y, *et al*. Leveraging genetic determinants of prostate-specific antigen levels towards improving prostate cancer screening. bioRxiv. 2022; published online April 21. DOI:10.1101/2022.04.18.22273850.

37 Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; **37**: 658–65.

38 Mahajan A, Taliun D, Thurner M, *et al*. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**: 1505–13.

39 Skrivankova VW, Richmond RC, Woolf BAR, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* 2021; **326**: 1614–21.

40 Skrivankova VW, Richmond RC, Woolf BAR, *et al.* Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* 2021; **375**: n2233.

41 Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genet Epidemiol* 2017; **41**: 714–25.

42 Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**: 1783–802.

43 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016; **40**: 304–14.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

44 Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; published online July 12. DOI:10.1093/ije/dyx102.

45 Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* 2018; published online Dec 10. DOI:10.1093/ije/dyy262.

46 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.

47 Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7**. DOI:10.7554/eLife.34408.

48 Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016; **183**: 758–64.

49 Zheng SL, Roddick AJ, Aghar-Jaffar R, *et al.* Association Between Use of Sodium-Glucose Cotransporter 2 Inhibitors, Glucagon-like Peptide 1 Agonists, and Dipeptidyl Peptidase 4 Inhibitors With All-Cause Mortality in Patients With Type 2 Diabetes: A Systematic Review and Meta-analysis. *JAMA* 2018; **319**: 1580–91.

50 Cotto KC, Wagner AH, Feng Y-Y, *et al.* DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* 2018; **46**: D1068–73.

51 Haring A, Murtola TJ, Talala K, Taari K, Tammela TLJ, Auvinen A. Antidiabetic drug use and prostate cancer risk in the Finnish Randomized Study of Screening for Prostate Cancer. *Scand J Urol* 2017; **51**: 5–12.

52 Marilly E, Cottin J, Cabrera N, *et al.* SGLT2 inhibitors in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials balancing their risks and benefits. *Diabetologia* 2022; published online Aug 4. DOI:10.1007/s00125-022-05773-8.

53 Wright EM. SGLT2 and cancer. *Pflugers Arch* 2020; **472**: 1407–14.

54 Villani LA, Smith BK, Marcinko K, *et al.* The diabetes medication Canagliflozin reduces cancer cell proliferation by inhibiting mitochondrial complex-I supported respiration. *Mol Metab* 2016; **5**: 1048–56.

55 Scafoglio C, Hirayama BA, Kepe V, *et al.* Functional expression of sodium-glucose transporters in cancer. *Proc Natl Acad Sci U S A* 2015; **112**: E4111-9.

56 Packer M, Anker SD, Butler J, *et al.* Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med* 2020; **383**: 1413–24.

57 Anker SD, Butler J, Filippatos G, *et al.* Empagliflozin in Heart Failure with a Preserved Ejection Fraction. *N Engl J Med* 2021; **385**: 1451–61.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

## Figure legends

**Figure 1. Genetic instrument selection, data sources, and analysis strategy in a triangulation study of the effect of SGLT2 inhibition on prostate cancer.**
For human genetic analyses, the effect of sodium–glucose cotransporter 2 (SGLT2) inhibition on risk of prostate cancer and its subtypes were estimated using Mendelian randomization. For observational analyses, the effect of use of SGLT2 inhibitors on incident prostate cancer risk were estimated in males with diabetes. Use of DPP4 inhibitors were used as an active comparator. For computation biology analyses, the differential expression of SGLT2 in prostate tumour tissue vs normal prostate tissue was estimated. Then the association of SGLT2 expression on prostate cancer progression (proxied by Gleason score) and survival time were estimated. More details of instrument selection and analysis strategies were listed in **Appendix Note 1-4**.

**Figure 2. Flowchart of patient inclusion and association between use of sodium-glucose cotransporter 2 (SGLT2) inhibitors and risk of incident prostate cancer or at high risk of prostate cancer.**
(A) flowchart of patient inclusion in the study population. SLGT2i=sodium glucose cotransporter 2 inhibitors; DPP4i=Dipeptidyl peptidase 4 inhibitors; TPSA=total prostate specific antigen. A patient could be excluded for more than one reason. (B) The association between use of SGLT2 inhibitors compared with DPP4 inhibitors and risk of prostate cancer or with total PSA > 10 ng/mL (which indicated high risk of prostate cancer). The covariates been used in this analysis including demographic data (age); comorbidities (benign prostatic hyperplasia, hypertension, dyslipidemia, diabetic complications, ischemic heart disease, peripheral vascular disease, heart failure, cerebrovascular disease, chronic lung disease, moderate or severe kidney disease, moderate or severe liver disease, and other cancers); antidiabetic drugs (metformin, insulin, glucagon-like peptide-1 receptor agonist, sulfonylurea, glinide, α-glucosidase inhibitor, and thiazolidinedione); and other medications (angiotensin converting enzyme inhibitor, angiotensin receptor blocker, calcium channel blocker, α/β-blockers, diuretic, statin, fibrate, aspirin, other antiplatelet drugs, non-steroidal anti-inflammatory drug, and 5α-reductase inhibitor). The unit of the incidence rate was 100,000 person years.

**Figure 3. Mendelian randomization estimates of the effects of sodium–glucose cotransporter 2 (SGLT2) inhibition on prostate cancer risk in the general European population.**
Two sets of genetic instruments were used in this analysis. Primary instruments included seven genetic variants that were associated with $HbA_{1c}$ ($P<5\times10^{-8}$) in the *SLC5A2* region. Stringent instruments were two genetic variants associated with both expression levels of *SLC5A2* and $HbA_{1c}$ levels (with colocalization probability>0.7 between the two) in the *SLC5A2* region.

**Figure 4. Differential expression analysis results of *SLC5A2* expression in prostate tumor vs adjacent normal tissue.**

**Table 1.** Characteristics of genetic variants associated with $HbA_{1c}$ (per 0.62% lowering) or expression levels of *SLC5A2* gene and used as proxies for SGLT2 inhibition in the general population.

| Genetic variant | Gene | Effect allele/ Non effect allele | Effect allele frequency | Effect (95%CI) | P value |
|---|---|---|---|---|---|
| **SGLT2 (primary)** | | | | | |
| rs1232538 | *SLC5A2* | G/T | 0.73 | -0.014 (-0.009 to -0.019) | $4.0\times10^{-08}$ |
| rs28675289 | *SLC5A2* | T/C | 0.04 | -0.038 (-0.027 to -0.049) | $1.5\times10^{-11}$ |
| rs28692853 | *SLC5A2* | A/C | 0.50 | -0.015 (-0.010 to -0.019) | $2.8\times10^{-10}$ |
| rs45625038 | *SLC5A2* | C/T | 0.97 | -0.041 (-0.028 to -0.055) | $1.2\times10^{-09}$ |
| rs55766044 | *SLC5A2* | C/T | 0.72 | -0.018 (-0.013 to -0.023) | $3.9\times10^{-12}$ |
| rs557720784 | *SLC5A2* | C/T | 0.95 | -0.026 (-0.016 to -0.037) | $6.1\times10^{-07}$ |
| rs8050500 | *SLC5A2* | C/T | 0.45 | -0.027 (-0.022 to -0.031) | $1.2\times10^{-30}$ |
| **SGLT2 (stringent)** | | | | | |
| rs9930811 | *SLC5A2* | G/A | 0.37 | -0.016 (-0.021 to -0.012) | $8.7\times10^{-12}$ |
| rs35445454 | *SLC5A2* | T/C | 0.34 | -0.013 (-0.018 to -0.008) | $1.2\times10^{-08}$ |

Notation: Two sets of instruments that proxying SGLT2 inhibition using different instrument selection processes were listed here. For the main analysis, primary instruments selected genetic variants that were robustly associated with $HbA_{1c}$ ($P<5\times10^{-8}$) in the *SLC5A2* region. Stringent instruments selected genetic variants that were associated with both expression of *SLC5A2* gene and $HbA_{1c}$ levels and showed colocalization evidence between the two (colocalization probability>0.7) in the *SLC5A2* region, which were used in the main analysis. Two pairs of primary and stringent instruments were in moderate LD ($r^2$ between rs9930811 and rs8050500=0.56, $r^2$ between rs35445454 and rs1232538=0.23), which suggested that the two different selection processes picked two shared genetic signals as instruments in this region.



medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

**Table 2.** Effect estimates of genetically proxied SGLT2 inhibition and genetic proxied $HbA_{1c}$ levels on total, aggressive and early onset prostate cancer among men in general population using data from PRACTICAL and GAME-ON/ELLIPSE consortium.

| Exposure | Outcome | No. cases | Model | Odds ratio (95%CI) | P value |
|---|---|---|---|---|---|
| Genetically proxied SGLT2 inhibition | Total prostate cancer | 79,148 | Inverse variance weighted MR | 0.56 (0.38 to 0.82) | 0.003 |
| | Advanced prostate cancer | 15,167 | Inverse variance weighted MR | 0.52 (0.27 to 1.18) | 0.049 |
| | Early-onset prostate cancer | 6,988 | Inverse variance weighted MR | 0.27 (0.11 to 0.71) | 0.008 |
| | Advanced vs non-advanced | 14,160 | Inverse variance weighted MR | 0.86 (0.35 to 2.13) | 0.75 |
| | High vs low aggressive | 15,561 | Inverse variance weighted MR | 1.14 (0.38 to 3.39) | 0.81 |
| | High vs low + intermediate aggressive | 20,658 | Inverse variance weighted MR | 0.69 (0.37 to 1.28) | 0.24 |

Notation: advanced prostate cancer was defined as metastatic disease or Gleason score (GS) ≥ 8 or PSA > 100 or prostate cancer death; early-onset refers to prostate cancer onset before age 55; low aggressive refers to T stage from the TNM staging ≤ T1, and GS ≤ 6, and PSA<10; intermediate aggressive refers to T stage: T2, and GS=7, and PSA 10~20; and high aggressive refers to T stage: T3/T4 or N1 or M1 or GS ≥ 8 or PSA >20.



medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

**Table 3.** Effect estimates of genetic proxied $HbA_{1c}$ levels on total, aggressive and early onset prostate cancer among men in general population using data from PRACTICAL consortium and effects of observed $HbA_{1c}$ levels on incident prostate cancer among men in general population using data from the REACTION study.

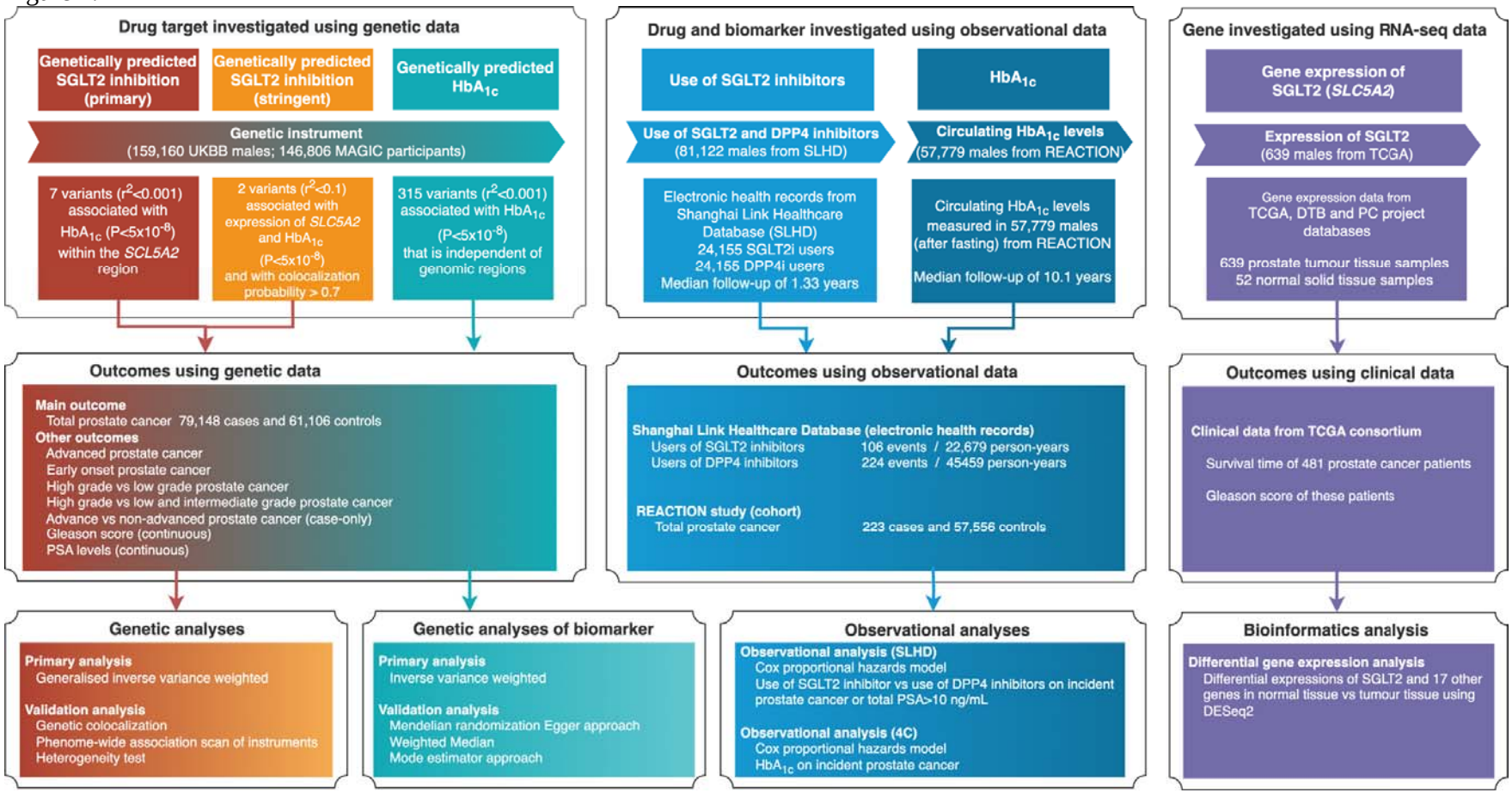
| Exposure | Outcome | No. cases | Model | Odds ratio (95%CI) | P value |
|---|---|---|---|---|---|
| Genetically proxied $HbA_{1c}$ levels | Total prostate cancer | 79,148 | Inverse variance weighted MR | 0.98 (0.92 to 1.05) | 0.63 |
| | Aggressive prostate cancer | 15,167 | Inverse variance weighted MR | 0.99 (0.92 to 1.07) | 0.81 |
| | Early-onset prostate cancer | 6,988 | Inverse variance weighted MR | 0.94 (0.82 to 1.08) | 0.37 |
| **Exposure** | **Outcome** | **No. cases** | **Model** | **Hazard ratio (95%CI)** | **P value** |
| Observed $HbA_{1c}$ levels (One SD unit=1.11%) | Incident prostate cancer (Including all 57,779 males) | 223 | Cox proportional hazard model | 0.93 (0.80 to 1.10) | 0.40 |
| Observed $HbA_{1c}$ levels (One SD unit=0.91%) | Incident prostate cancer (Excluding users of anti-diabetic drugs) | 201 | Cox proportional hazard model | 0.95 (0.80 to 1.12) | 0.53 |

Notation: aggressive prostate cancer, defined as Gleason score ≥8, PSA>100 ng/mL, metastatic disease (M1) or death from prostate cancer, and early onset prostate cancer, defined as participants diagnosed with prostate cancer before the age of 55 years. SD refers to standard deviation.



medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 1.



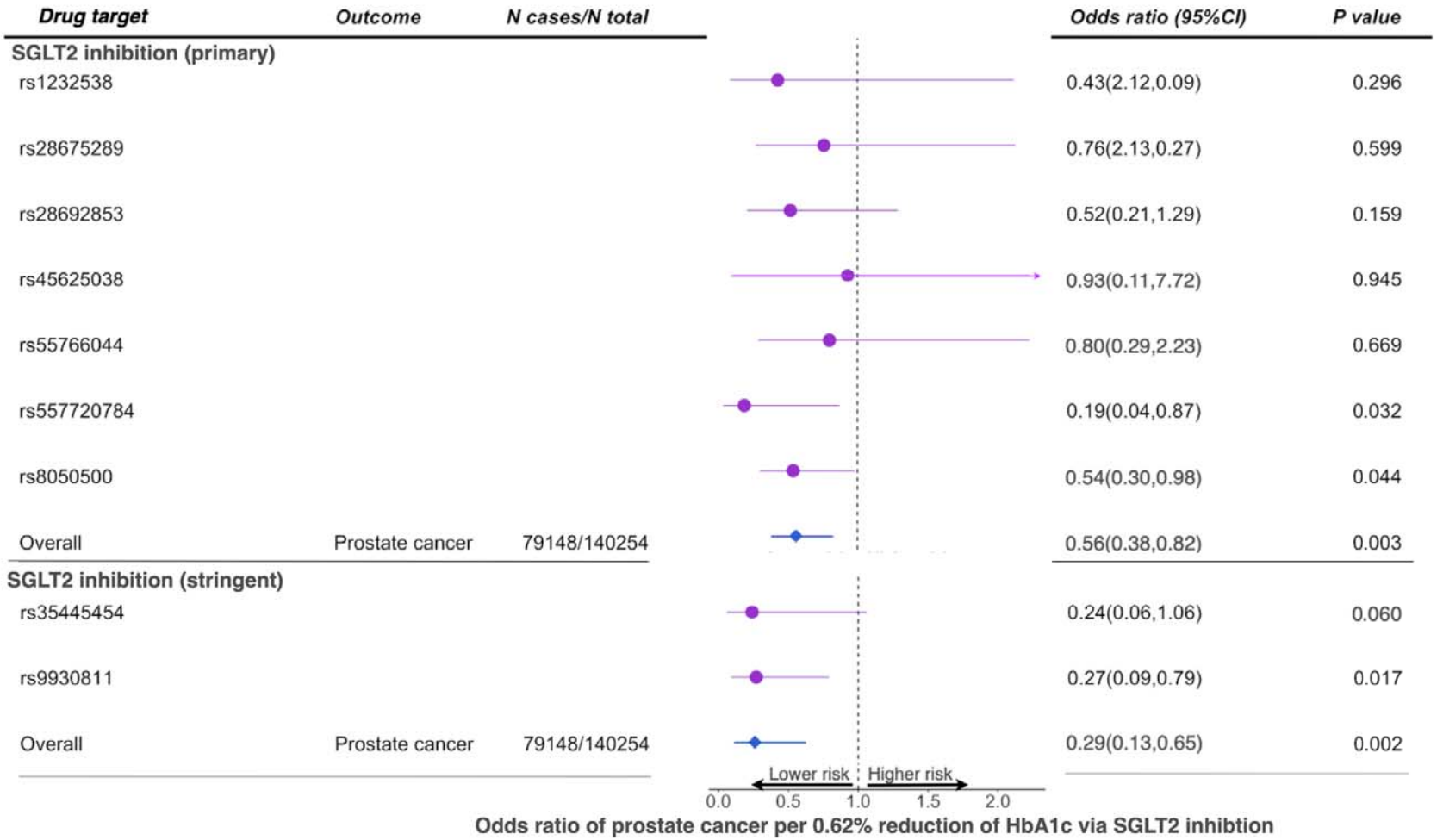



medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2.

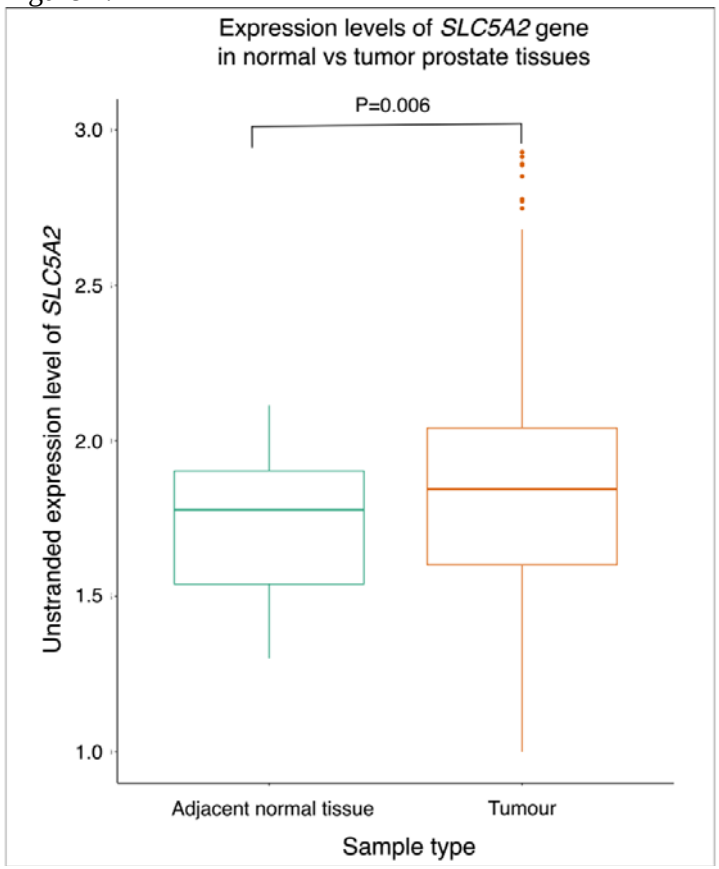
Figure 3.





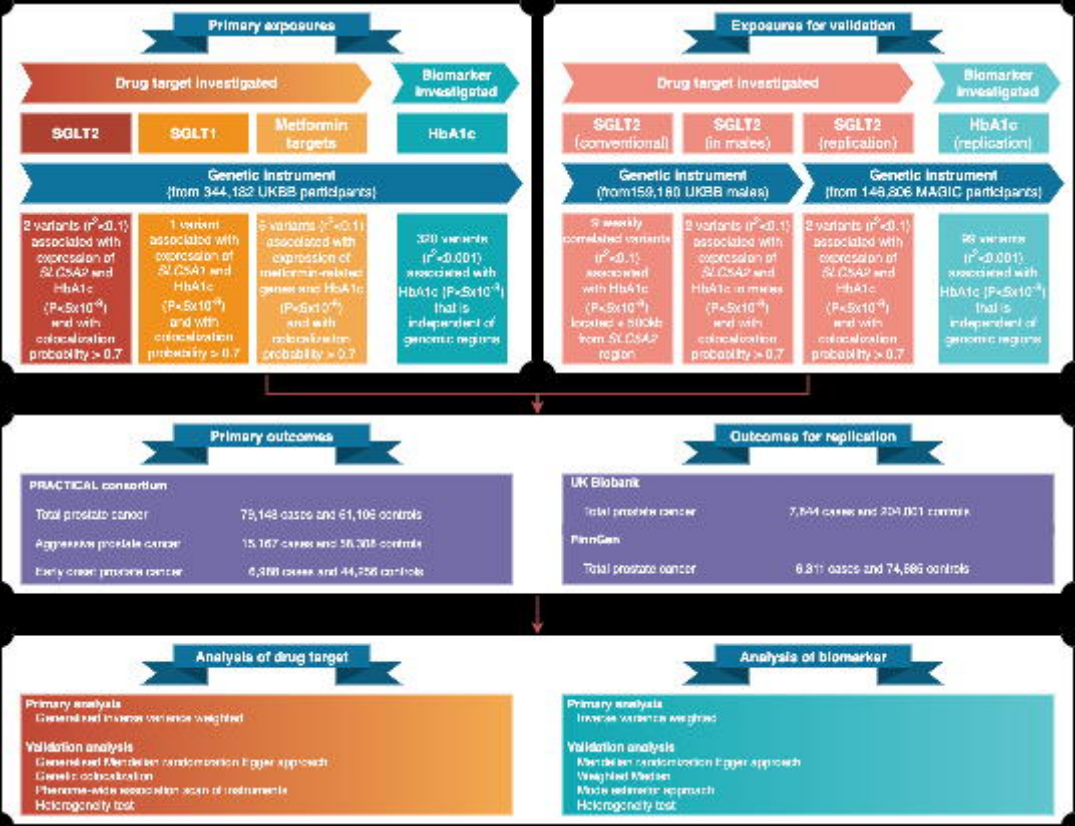

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 4.

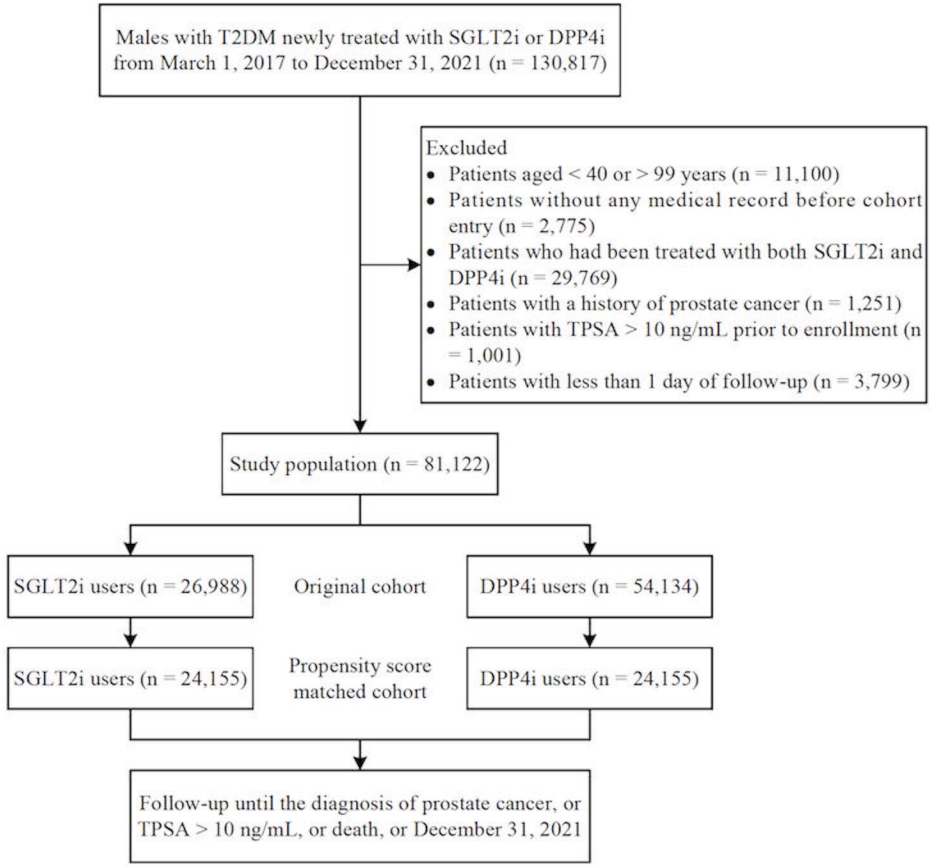





medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296790; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

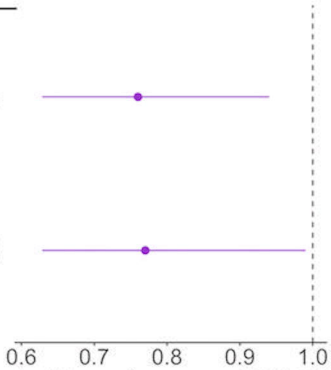

Primary exposures
Drug target investigated
Biomarker investigated
SGLT2
SGLT1
Metformin targets
HbA1c
Genetic instrument (from 344,182 UKBB participants)
2 variants ($r^2$<0.1) associated with expression of SLC5A2 and HbA1c ($P<5x10^{-8}$) and with colocalization probability > 0.7
1 variant associated with expression of SLC5A1 and HbA1c ($P<5x10^{-8}$) and with colocalization probability > 0.7
5 variants ($r^2$<0.1) associated with expression of metformin-related genes and HbA1c ($P<5x10^{-8}$) and with colocalization probability > 0.7
320 variants ($r^2$<0.001) associated with HbA1c ($P<5x10^{-8}$) that is independent of genomic regions
Exposures for validation
Drug target investigated
Biomarker investigated
SGLT2 (conventional)
SGLT2 (in males)
SGLT2 (replication)
HbA1c (replication)
Genetic instrument (from159,180 UKBB males)
Genetic instrument (from 146,806 MAGIC participants)
9 weakly correlated variants ($r^2$<0.1) associated with HbA1c ($P<5x10^{-8}$) located ± 500kb from SLC5A2 region
2 variants ($r^2$<0.1) associated with expression of SLC5A2 and HbA1c in males ($P<5x10^{-8}$) and with colocalization probability > 0.7
2 variants ($r^2$<0.1) associated with expression of SLC5A2 and HbA1c ($P<5x10^{-8}$) and with colocalization probability > 0.7
98 variants ($r^2$<0.001) associated with HbA1c ($P<5x10^{-8}$) that is independent of genomic regions
Primary outcomes
PRACTICAL consortium
Total prostate cancer 79,148 cases and 61,106 controls
Aggressive prostate cancer 15,167 cases and 58,308 controls
Early onset prostate cancer 6,988 cases and 44,256 controls
Outcomes for replication
UK Biobank
Total prostate cancer 7,844 cases and 204,031 controls
PrinCon
Total prostate cancer 6,911 cases and 74,686 controls
Analysis of drug target
Primary analysis
Generalised inverse variance weighted
Validation analysis
Generalised Mendelian randomisation Egger approach
Genetic colocalisation
Phenome-wide association scan of instruments
Heterogeneity test
Analysis of biomarker
Primary analysis
Inverse variance weighted
Validation analysis
Mendelian randomisation Egger approach
Weighted Median
Mode estimator approach
Heterogeneity test

A



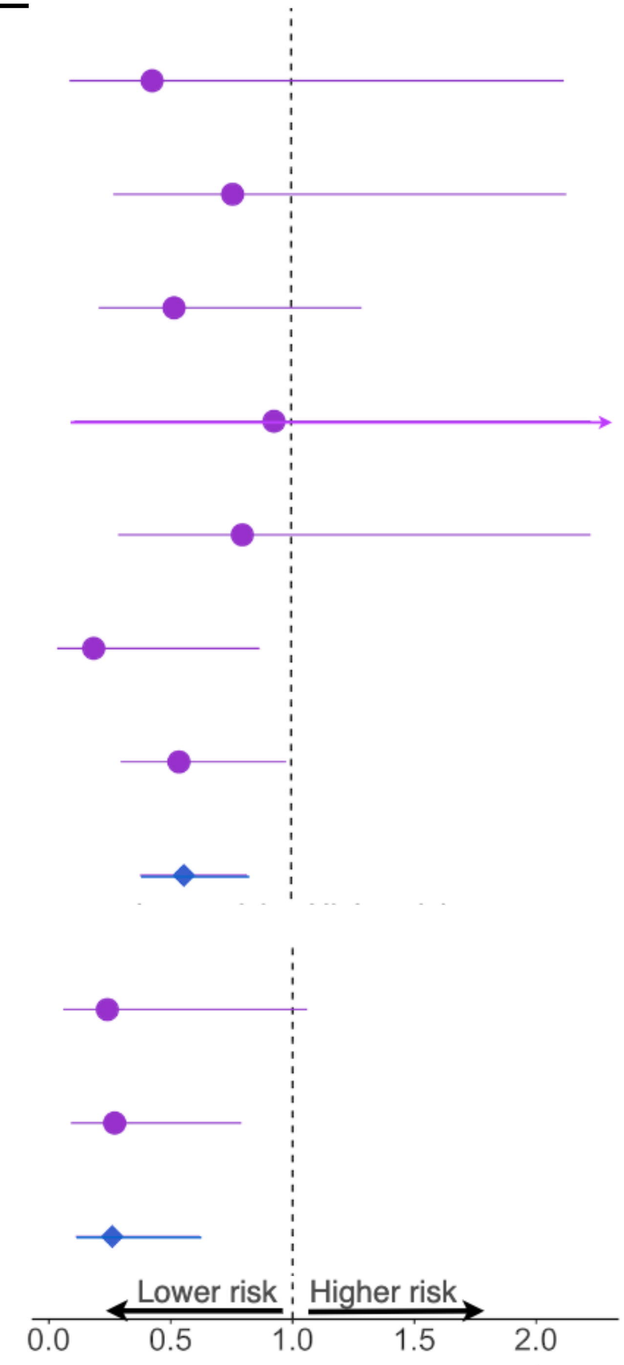

B

| Cohort | No events / No person years | | Incidence rate | | Hazard ratio (95%CI) |
|---|---|---|---|---|---|
| | SGLT2i | DPP4i | SGLT2i | DPP4i | |
| Original | 114 / 25171 | 574 / 102209 | 452.9 | 561.6 | 0.76(0.61,0.94) |
| PS-matched | 106 / 22679 | 224 / 45459 | 467.4 | 492.8 | 0.77(0.61,0.99) |



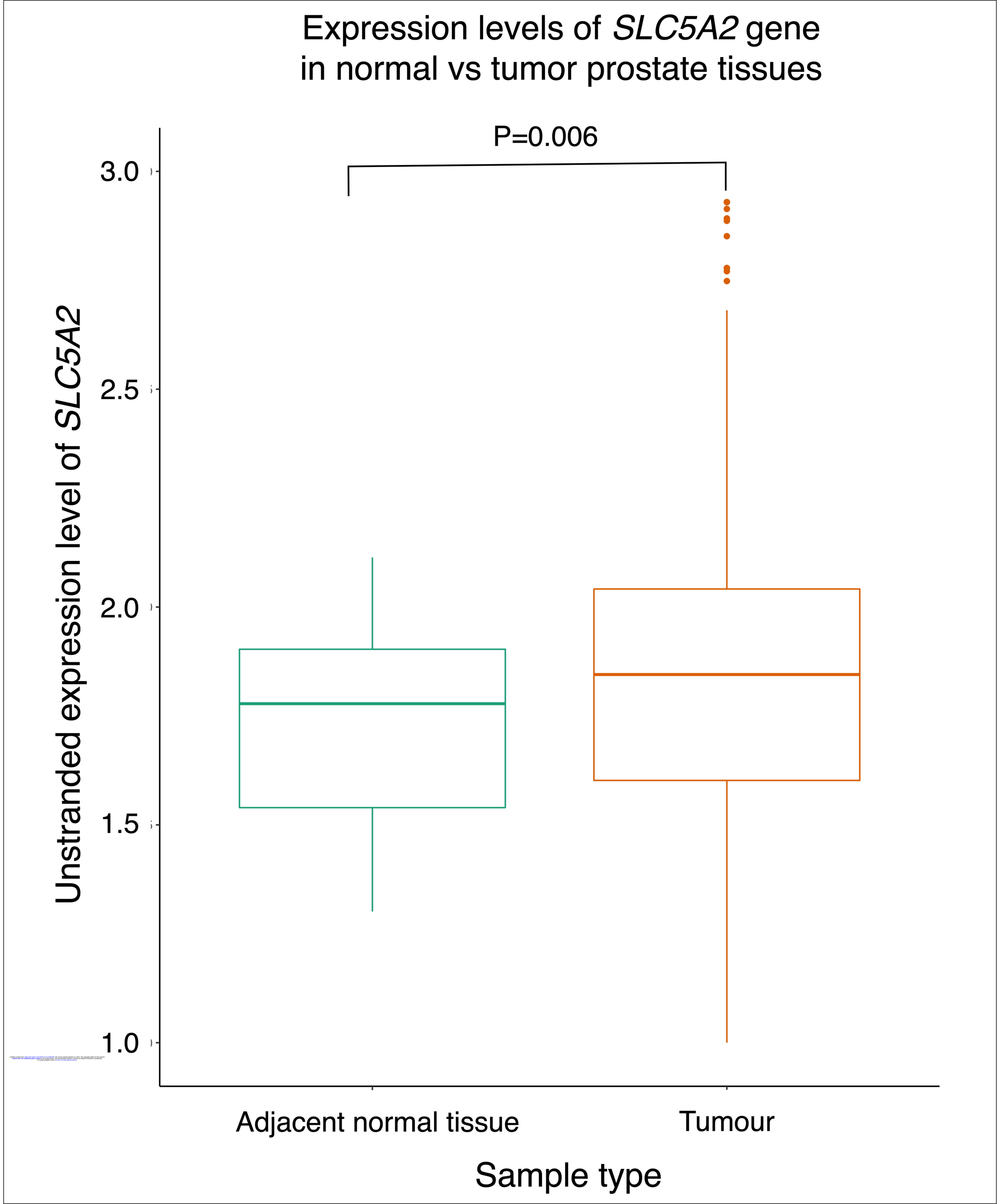

Hazard ratio of prostate cancer incidence for users of SGLT2i vs users of DPP4i

| Drug target | Outcome | N cases/N total | Odds ratio (95%CI) | P value |
|---|---|---|---|---|
| **SGLT2 inhibition (primary)** | | | | |
| rs1232538 | | | 0.43(2.12,0.09) | 0.296 |
| rs28675289 | | | 0.76(2.13,0.27) | 0.599 |
| rs28692853 | | | 0.52(0.21,1.29) | 0.159 |
| rs45625038 | | | 0.93(0.11,7.72) | 0.945 |
| rs55766044 | | | 0.80(0.29,2.23) | 0.669 |
| rs557720784 | | | 0.19(0.04,0.87) | 0.032 |
| rs8050500 | | | 0.54(0.30,0.98) | 0.044 |
| Overall | Prostate cancer | 79148/140254 | 0.56(0.38,0.82) | 0.003 |
| **SGLT2 inhibition (stringent)** | | | | |
| rs35445454 | | | 0.24(0.06,1.06) | 0.060 |
| rs9930811 | | | 0.27(0.09,0.79) | 0.017 |
| Overall | Prostate cancer | 79148/140254 | 0.29(0.13,0.65) | 0.002 |




**Odds ratio of prostate cancer per 0.62% reduction of HbA1c via SGLT2 inhibtion**

medRxiv preprint doi: https://doi.org/10.1101/2023.10.10.23296760; this version posted October 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made freely available under a CC-BY-4.0 International license.


Expression levels of SLC5A2 gene
in normal vs tumor prostate tissues
P=0.006
3.0
2.5
2.0
1.5
1.0
Unstranded expression level of SLC5A2
Adjacent normal tissue
Tumour
Sample type