

RESEARCH

Open Access



Systematic druggable genome-wide Mendelian randomization identifies therapeutic targets for lung cancer

Wenfu Song^{1,2}, Yingying Li^{1,2}, Yaxuan Yao², Shiling Sun^{1,2*}, Xutao Guan^{1,2} and Bing Wang¹

Abstract

Background Drug repurposing provides a cost-effective approach to address the need for lung cancer prevention and therapeutics. We aimed to identify actionable druggable targets using Mendelian randomization (MR).

Methods Summary-level data of gene expression quantitative trait loci (eQTLs) were sourced from the eQTLGen resource. We procured genetic associations with lung cancer and its subtypes from the TRICL, ILCCO studies (discovery) and the FinnGen study (replication). We implemented Summary-data-based Mendelian Randomization analysis to identify potential therapeutic targets for lung cancer. Colocalization analysis was further conducted to assess whether the identified signal pairs shared a causal genetic variant.

Findings In the main analysis dataset, we identified 55 genes that demonstrate a causal relationship with lung cancer and its subtypes. However, in the replication cohort, only three genes were found to have such a causal association with lung cancer and its subtypes, and of these, HYKK (also known as AGPHD1) was consistently present in both the primary analysis dataset and the replication cohort. Following HEIDI tests and colocalization analyses, it was revealed that HYKK (AGPHD1) is associated with an increased risk of squamous cell carcinoma of the lung, with an odds ratio and confidence interval of OR = 1.28, 95%CI = 1.24 to 1.33.

Interpretation We have found that the HYKK (AGPHD1) gene is associated with an increased risk of squamous cell carcinoma of the lung, suggesting that this gene may represent a potential therapeutic target for both the prevention and treatment of lung squamous cell carcinoma.

Keywords Lung cancer, Druggable target, Mendelian randomization, Colocalization, Gene expression

Introduction

Lung cancer is the leading cause of cancer death worldwide [1], with more than one million deaths each year [2]. According to tissue type, lung cancer can be divided into

small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), among which NSCLC mainly includes lung adenocarcinoma and lung squamous cell carcinoma [3]. The pathogenesis of lung cancer is a complex process involving various risk factors [4]. At present, the main treatment methods of lung cancer include surgery, chemotherapy, radiation therapy and biological therapy.

However, the treatment of lung cancer still faces huge challenges. Recurrence and metastasis after surgery will still lead to disease progression, and chemotherapy drugs in the treatment of tumors are accompanied by a large number of toxic side effects, such as bone marrow suppression, hair loss, bleeding, liver and kidney function

*Correspondence:

Shiling Sun
sunshiling@sina.com

¹ Department of Hematology and Oncology, the First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, Henan Province, China

² The First Clinical Medical College of Henan University of Chinese Medicine, Zhengzhou, Henan Province, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

injury [5, 6]. At present, the treatment mode of lung cancer has gradually shifted to targeted therapy, and there are different therapeutic drugs for different molecular targets [7]. However, targeted therapy still has bottlenecks: On the one hand, targeted drugs are ineffective without driver gene mutation, On the other hand, after a median treatment time of 12 to 14 months, most drugs develop acquired resistance, leading to disease progression [8]. Therefore, finding new therapeutic targets for lung cancer is still the focus and difficulty of research.

Large-scale human genetics studies provide an opportunities for the development of new drugs for many sophisticated diseases, as drug targets supported by genetic evidence have a greater likelihood of success during drug development [9, 10]. In short, “medicable” genes that code for protein or gene expression can afford thread to find drug targets [11]. Currently, many large-scale genome-wide association studies (GWASs) have identified many single nucleotide polymorphisms (SNPs) associated with lung cancer risk, however, because many of the identified SNPs are located in non-coding regions or gene intervals, GWAS data do not provide clear and direct clues about disease-causing genes and drug targets.

Mendelian randomization (MR) analyses use inherited variation as an instrumental variable to enhance inferences of causality between exposure and outcomes. This approach is less susceptible to confusion and reverse causality bias than observational studies because genetic variation is randomly distributed at conception and is not modified by acquired factors and disease onset [12]. Today, MR analysis has been diffusely used to re-use confirmed drugs and to identify new remedial targets by coordinating aggregated data from disease GWAS and quantitative trait loci (eQTL) studies [13–15]. The level of expression of a gene can be well thought out a lifetime exposure, and eQTL located in genomic regions of the medicable gene is often considered a proxy [16, 17]. In this study, we implemented a systematic pharmacizable genome-wide SMR to identify therapeutic targets for lung cancer and its subtypes.

Methods

Figure 1A shows the overall design of the study. In this study, we used data sets from the Transdisciplinary Research for Cancer of Lung (TRICL) and The International Lung Cancer Consortium (ILCCO) as the primary analysis, involving 29,266 patients of European descent and 56,450 controls [18]. To verify our findings, we used the FinnGen study dataset for replication. The discovery eQTL dataset was procured from eQTLGen (<https://eqtlgen.org/>), in which eQTLs for 16,987 genes were procured from 31,684 blood samples from individuals of healthy European ancestry [19]. Meanwhile, as surrogate

data for eQTL analyses, we utilized human lung tissue and whole blood samples from the GTEx Consortium, comprising 515 and 670 samples, respectively [20].

Figure 1B show the cis-eQTL analysis. The original eQTL data consisted of 31,684 blood samples, 10,317 trait associated SNPs, 11 M SNPs ($MAF \geq 1\%$) and 19,960 genes studied. We selected data with a window size of 1 Mb and a $MAF \geq 1\%$, resulting in 16,989 (88.3%) cis-eQTL genes. The data processing methods for GTEx eQTLs are consistent with those of eQTLGen consortium.

Lung cancer outcome datasets summary

Data for lung cancer and its subtypes were obtained from the TRICL, ILCCO studies, and the FinnGen study. In the TRICL and ILCCO studies, all participants were of European ancestry, comprising 29,266 European-ancestry patients and 56,450 controls, with 11,273 cases and 55,483 controls for lung adenocarcinoma, 7,426 cases and 55,627 controls for lung squamous cell carcinoma, and 2,664 cases and 21,444 controls for small cell lung cancer [18]. Summary data on the genetic associations of lung cancer and its subtypes were derived from the publicly available R10 data of the FinnGen study, which included 5,315 cases and 314,193 controls for non-small cell lung cancer, 1,590 cases and 314,193 controls for lung adenocarcinoma, 1,510 cases and 314,193 controls for lung squamous cell carcinoma, and 717 cases and 314,193 controls for small cell lung cancer [21].

Summary-data-based MR analysis

A summary-data-based Mendelian Randomization (SMR) approach was employed to identify associations between systemic “druggable” gene eQTL data and the risk of lung cancer and its subtypes [17]. SMR can achieve higher statistical power than classical MR analyses when exposure and outcome are obtainable from two independent samples with large sample sizes based on top associated cis-QTLs [17]. Cis-eQTLs were selected within a ± 1000 kb window around the corresponding genes using a P -value threshold of 1.0×10^{-5} . SNPs with allele frequency differences greater than a specified threshold (set at 0.01 in this study) among any pair of datasets (including LD reference samples, QTL summary data, and outcome summary data) were excluded. The heterogeneity in dependent instruments (HEIDI) test was applied to distinguish pleiotropy from linkage, where $P\text{-HEIDI} < 0.05$ was considered potentially due to pleiotropy and thus discarded from the analysis. SMR and HEIDI tests were performed using the SMR software tool (SMR v1.3.1, the code is available at <https://yanglab.westlake.edu.cn/software/smr/#Overview>). The P -values was adjusted by Benjamini-Hochberg method, and to control

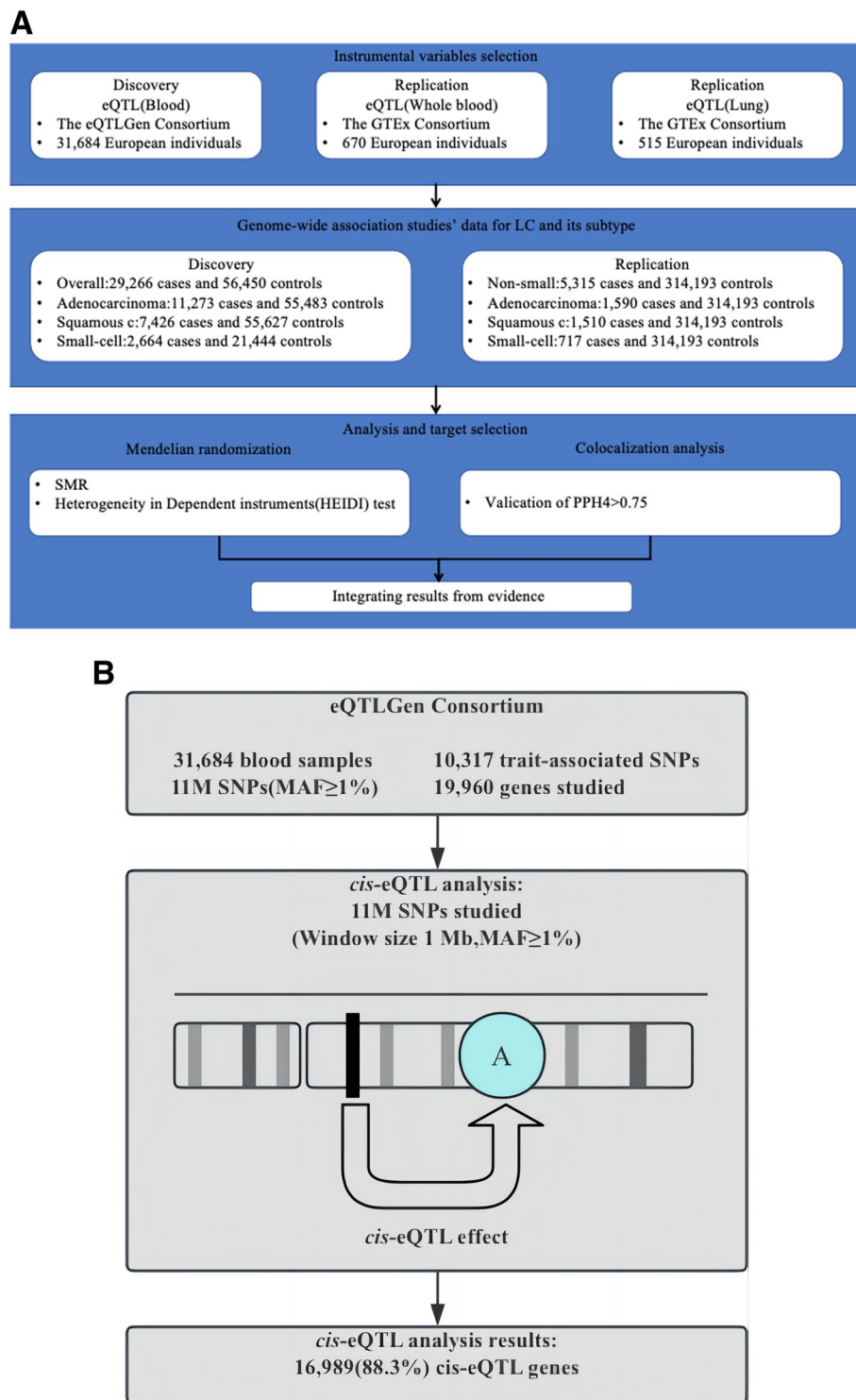


Fig. 1 **A** Study design. SMR, summary-based Mendelian randomization; QTL, quantitative trait loci; LC, Lung Cancer; PPH4, posterior probability of H4. **B** cis-eQTL analysis

the false discovery rate (FDR) at 0.05. Co-localization analysis was conducted on results with FDR-corrected P -values < 0.05 and P -HEIDI > 0.05 .

Colocalization analysis

Colocalization analysis was carried out using the coloc R package to investigate the potential causal relationship between systemic “druggable” genes and lung cancer and its subtypes [22]. There are four primary hypotheses in colocalization analysis: H0: SNPs at the selected locus are unrelated to both exposure (eQTL) and disease (lung cancer). H1: SNPs at the selected locus are related to exposure (eQTL) but not to disease (lung cancer). H2: SNPs at the selected locus are related to disease (lung cancer) but not to exposure (eQTL). H3: SNPs at the selected locus are related to either exposure (eQTL) or disease (lung cancer), but these are distinct SNPs (LD). H4: SNPs at the selected locus are related to both exposure (eQTL) and disease (lung cancer), sharing the same SNP (LD). A PPH4 value > 0.75 for H4 is considered evidence supporting colocalization [23].

Results

SMR analysis of cis-eQTLs and lung cancer and its subtype

In the main analysis dataset results, the causal relationships between blood-derived eQTLs and lung cancer and its subtypes are depicted in forest plots (Fig. 2A through E). After excluding SNPs with P -HEIDI < 0.05 and FDR-adjusted P -values > 0.05 via SMR analysis, we identified a total of 21 genes that have a causal relationship with lung cancer, including RNASET2, LPAR2, RPS6KA2, FAM76B, MAP3K20, PSMD5, GMFG, DISP2, NAF1, MPZL2, MPZL-3, CEP57, GBA1, HLA-DQB1, ARID3B, HYKK (AGPHD1), HLA-E, HLA-DQB1-AS1, CUTALP, HLA-DQB2, and MTX1LP; these findings can be found in detail in Fig. 2A. There were 12 genes with a causal relationship to lung adenocarcinoma, including RNASET2, SECISBP2L, RPS6KA2, PTGFR, MPZL3, GMFG, NTN5, NAF1, MPZL2, GALK2, HYKK (AGPHD1), and UCKL1, as shown in Fig. 2B. For lung squamous cell carcinoma, 23 genes were causally related, comprising ANKRD26, RAD51C, PEAK3, PBX3, COPA, DCAF8, CHEK1, CCDC117, NCSTN, LRRC58, CCR1, HYKK (AGPHD1), HLA-DQB1, ZNF165, HLA-DQB1-AS1, TRIM10, SUPT4H1, C4B, HLA-DQB2, ATP6V1G2,

APOM, LINC00243, and ZKSCAN8, as detailed in Fig. 2C. Six genes were found to have a causal relationship with small cell lung cancer, namely RAD52, FRS3, FLOT1, IER3, HYKK (AGPHD1), and LINC00243, which are illustrated in Fig. 2D.

In the replication cohort results, we observed that PSMA4 has a causal relationship with lung adenocarcinoma, KIAA1614-AS1 is causally linked to lung squamous cell carcinoma, and HYKK (AGPHD1) has a causal relationship with non-small cell lung cancer, lung adenocarcinoma, and lung squamous cell carcinoma; these findings are presented in Fig. 2E. Notably, only HYKK (AGPHD1) was replicated in the FinnGen study, and it was consistently replicated across non-small cell, adenocarcinoma, and squamous subtypes.

SMR analysis of replication cis-eQTLs and lung cancer and its subtype

Results from the primary analysis dataset using GTEx whole blood and lung cancer and its subtypes. After excluding SNPs with P -HEIDI < 0.05 and FDR-adjusted P -values > 0.05 via SMR analysis, we identified a total of 18 genes that have a causal relationship with lung cancer, including U919328.19, GABBR1, IER3, LINC00243, SKIV2L, HLA-DQA1, HLA-DQB1, HLA-DQB1-AS1, HLA-DQB2, RNASET2, RP1-167A14.2, PSMD5-AS1, MPZL2, MPZL3, PSMA4, UCKL1 and PRPF6. There were 5 genes with a causal relationship to lung adenocarcinoma, including RNASET2, MPZL2, MPZL3, CHRNA5 and UCKL1. For lung squamous cell carcinoma, 11 genes were causally related, comprising COPA, NCSTN, U91328.19, GABBR1, IER3, HLA-DQB1-AS1, HLA-DQB2, ANKRD26, PKD2L1, PSMA4 and RAD51C. Only one gene was found to have a causal relationship with small cell lung cancer, namely PSMA4. However, in the replication cohort results, we only observed that PSMA4 has a causal relationship with lung adenocarcinoma. Thus, the causal relationship between PSMA4 and lung cancer, as well as its subtypes, requires further investigation.

Results from the primary analysis dataset using GTEx human lung tissue and lung cancer and its subtypes. After excluding SNPs with P -HEIDI < 0.05 and FDR-adjusted P -values > 0.05 via SMR analysis, we identified a total of 14 genes that have a causal relationship with

(See figure on next page.)

Fig. 2 **A** Associations of predicted gene expression with lung cancer in Mendelian randomization analysis. **B** Associations of predicted gene expression with lung adenocarcinoma cancer in Mendelian randomization analysis. **C** Associations of predicted gene expression with lung squamous cell carcinoma cancer in Mendelian randomization analysis. **D** Associations of predicted gene expression with small cell lung cancer in Mendelian randomization analysis. **E** Associations of predicted gene expression with lung cancer (replication) in Mendelian randomization analysis. OR, odds ratio; CI, confidence interval

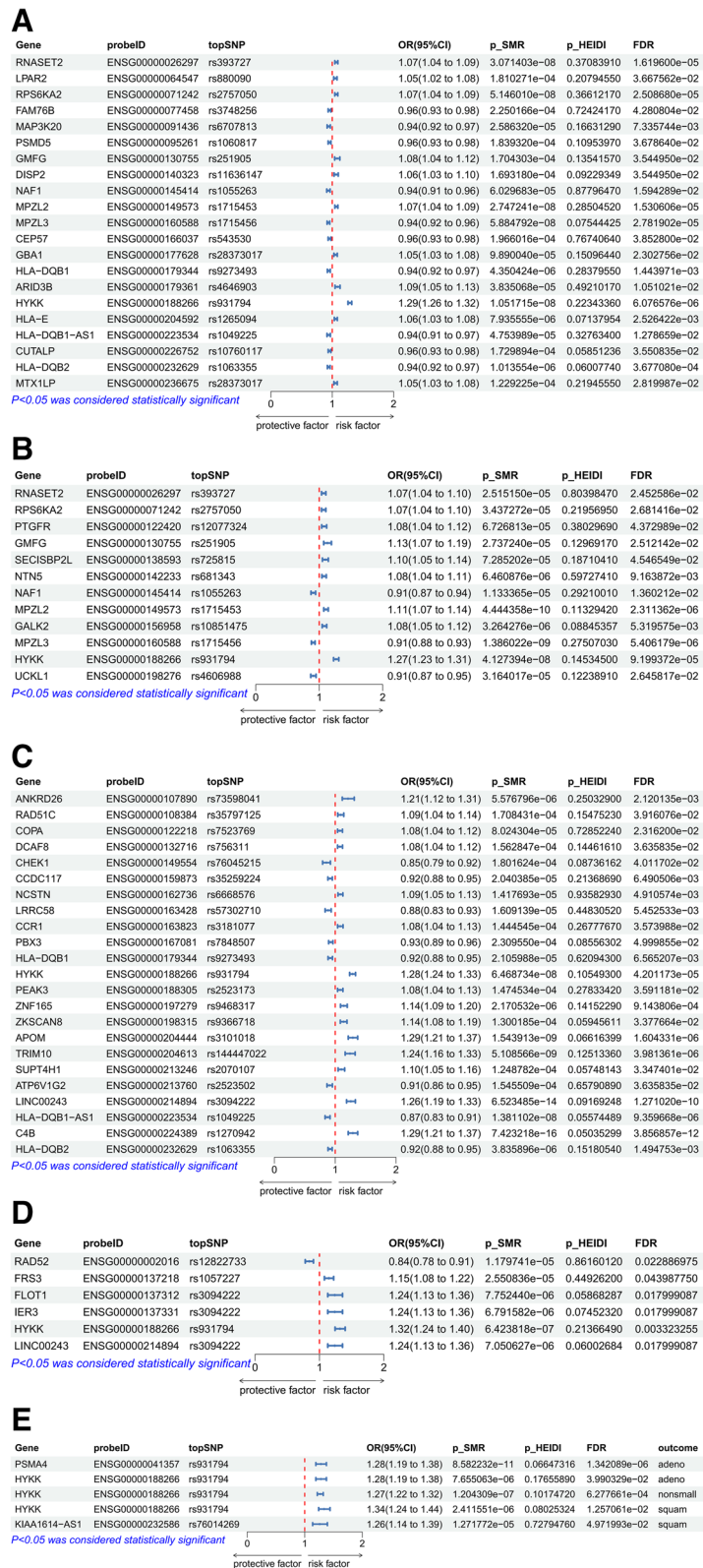


Fig. 2 (See legend on previous page.)

lung cancer, including FUBP1, RP11-218F10.3, CTD-2012J19.1, BTN2A2, TRIM31, FLOT1, LINC00243, HLA-DQA1, HLA-DQB1, HLA-DQB1-AS1, HLA-DQB2, RNASET2, AC011330.5 and SECISBP2L. There were 6 genes with a causal relationship to lung adenocarcinoma, including FUBP1, CTD-2012J19.3, CTD-2012J19.1, DCBLD1, NRG1 and SECISBP2L. For lung squamous cell carcinoma, 9 genes were causally related, comprising U91328.19, FLOT1, LINC00243, HLA-DQA1, HLA-DQB1, HLA-DQB1-AS1, HLA-DQB2, ZNF483 and BLOC1S2. They are not found to have a causal relationship with small cell lung cancer. In the replication cohort results, We have not identified any genes that show a causal association with lung cancer or its subtypes. Therefore, further research is needed to elucidate the causal relationships between eQTLs in human lung tissue and lung cancer, including its various subtypes.

Colocalization analysis

In the results of the SMR analysis, we discovered that HYKK (AGPHD1) was consistently identified in lung cancer data from both TRICL and ILCCO studies, and it was replicated specifically in non-small cell lung cancer, lung adenocarcinoma, and lung squamous cell carcinoma within the FinnGen dataset. To substantiate these findings, colocalization analysis was performed to estimate the posterior probability of common variants being causally associated with the disease outcomes. For the significant MR results, we retrieved all SNPs located within a ± 100 kb window upstream and downstream of each SNP for colocalization analysis between eQTLs and GWAS hits. In the main analysis dataset results, HYKK (AGPHD1) showed 100% posterior probability of shared causal variant (PP.H4=100%) across all four subtypes considered. In the replication cohort results, HYKK (AGPHD1) displayed varying degrees of colocalization evidence: for the non-small subtype, PP.H4 was 28.5%;

for the lung adenocarcinoma subtype, PP.H4 was 35.2%; and for the lung squamous cell carcinoma subtype, PP.H4 was 98.9% (indicating strong evidence for colocalization since PP.H4>0.75). These detailed colocalization probabilities are presented in Table 1. This suggests that HYKK (AGPHD1) is a promising candidate gene for further investigation into its role as a potential causal factor in multiple subtypes of lung cancer, particularly in lung squamous cell carcinoma, based on the high degree of colocalization likelihood observed. The colocalization results for the remaining genes are presented in Table S2.

Integrating evidence

After combining the results of different methods, we finally found a potential causal association between HYKK (AGPHD1) and lung squamous cell carcinoma. In order to clarify the causal relationship between HYKK (AGPHD1) and lung squamous cell carcinoma, a conditional Logistic regression model was used to calculate the OR value and 95%CI of lung squamous cell carcinoma associated with the use of candidate drugs, and the results were visualized by YangLab SMR visualization method (Fig. 3). The results showed that HYKK (AGPHD1) was associated with an increased risk of lung squamous cell carcinoma in the primary analysis set (OR: 1.28; 95% CI: 1.24–1.33), which is consistent with the results in the replication queue (OR: 1.34; 95% CI: 1.24–1.44; Fig. 2C and E).

Indeed, it's intriguing that while a causal link between PSMA4 and lung adenocarcinoma was not evident in the primary datasets, replication cohorts from eQTLGen and GTEx analyses involving whole blood tissue have suggested a potential causal relationship. Utilizing the YangLab SMR visualization approach has facilitated the illustration of these findings, pointing towards PSMA4 potentially acting as a protective factor against lung adenocarcinoma (Fig. 4). Nonetheless, additional studies are imperative to confirm

Table 1 Results of colocalization of HYKK (AGPHD1) with lung cancer and its subtypes

Outcome	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4	PP for shared variant
Discovery						
Over all	1.83e-96	3.64e-06	5.04e-94	4.15e-06	1.00	100%
Adenocarcinoma	1.44e-44	3.64e-06	3.98e-42	5.01e-06	1.00	100%
Squamous	2.61e-37	3.64e-06	7.19e-35	4.96e-06	1.00	100%
Small cell	5.79e-19	3.65e-06	1.60e-16	7.35e-06	1.00	100%
Replication						
Non small	7.31e-27	2.58e-03	2.02e-24	7.12e-01	2.85e-01	28.5%
Adenocarcinoma	4.14e-06	2.34e-03	1.14e-03	6.45e-01	3.52e-01	35.2%
Squamous	1.28e-13	4.22e-05	3.54e-11	1.06e-02	9.89e-01	98.9%

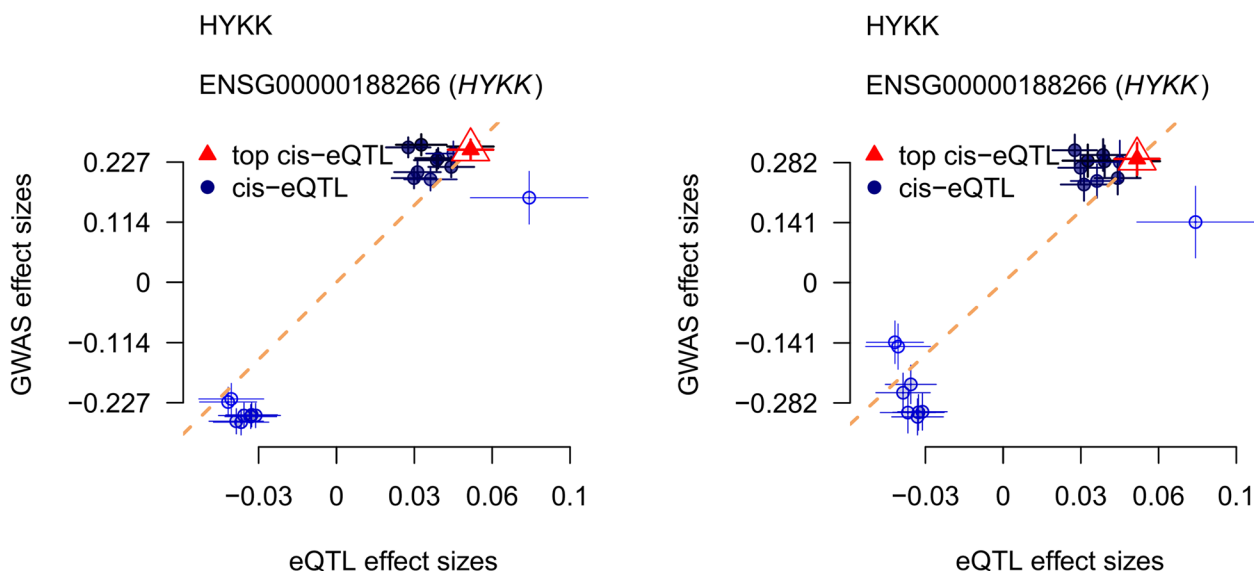


Fig. 3 Scatter plot for associations between HYKK (AGPHD1) gene and lung squamous cell carcinoma

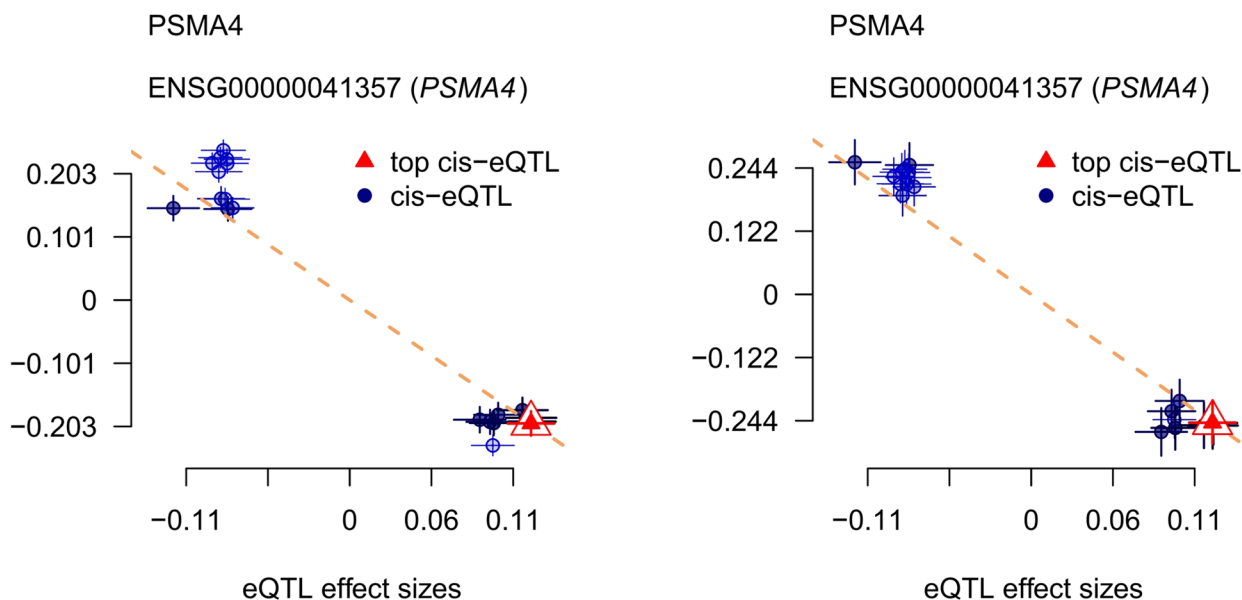


Fig. 4 Scatter plot for associations between PSMA4 gene and lung adenocarcinoma

this causal relationship and to ascertain the exact nature and magnitude of its protective effect. Further research should involve more granular analyses, larger sample sizes, and possibly experimental models to validate the observed association and explore the underlying biological mechanisms.

Discussion

In an effort to identify opportunities for drug repurposing and development targeting lung cancer and its subtypes, we conducted a large-scale MR analysis that leverages the underlying genetic mechanisms of gene expression as robust evidence. Our study identified hydroxylysine

kinase (HYKK), also known as amino-glycoside phosphotransferase 1 (AGPHD1), as a potentially causally related druggable target gene for squamous cell carcinoma of the lung.

In our research, we found that rs931794 in the HYKK (AGPHD1) gene is a risk factor for lung squamous cell carcinoma. However, Y. Bossé et al. utilized human lung tissue eQTL and lung cancer GWAS data, employing TWAS analysis through S-PrediXcan and FUSION methods, which identified HYKK (AGPHD1) as a protective factor for lung cancer, a finding that differs from the results of our study [24]. It should be noted that genes uncovered by TWAS analyses do not necessarily imply causality, rather, TWAS analysis are more aptly interpreted as prioritized or ordered candidate causal genes at specific loci. Furthermore, TWAS analyses are unable to distinguish causality from pleiotropy. Consequently, in the study by Y. Bossé et al., no definitive conclusions were drawn regarding the relationship between HYKK (AGPHD1) and either lung cancer in general or specifically lung squamous cell carcinoma, and the replication data cohort within their investigation failed to replicate the reported association between HYKK and lung cancer. Thus, the study by Y. Bossé et al. is not contradictory to our research, but rather, our investigation employs the use of the SMR analysis, which offers a stronger assessment of causal relationships compared to TWAS methodologies.

Moreover, numerous studies align with the findings of our investigation. Wang Hong et al. conducted a meta-analysis using STATA software incorporating nine studies and suggested that the rs8034191 variant in the HYKK (AGPHD1) gene might not alter the genetic risk for lung cancer in Asian populations [25]. However, multiple meta-analyses have shown that the rs8034191 variant in the HYKK (AGPHD1) gene is a risk factor for lung cancer, with individuals carrying the HYKK (AGPHD1) rs8034191 (A>G) variant having a higher likelihood of developing lung cancer [26–29]. Qi Wang through case-control studies involving 611 cases and a meta-analysis, established a relationship between the rs931794 variant and increased lung cancer risk in the Chinese population, with stratified analysis revealing a significant association between the G allele and lung cancer risk among smokers [30].

We propose that HYKK (AGPHD1) may function as a disease-causing gene specifically for lung squamous cell carcinoma, although the exact physiological role of HYKK (AGPHD1) remains unclear. Qi Wang discovered that the HYKK (AGPHD1) gene interacts with the CHRNA5-CHRNA3-CHRN4 cluster and is implicated in the risk of lung cancer development [30]. The CHRNA5-CHRNA3-CHRN4 cluster encodes nicotinic

acetylcholine receptor subunits primarily expressed in aspiratory epithelium, pneumonic neuroendocrine cells, and lung cancer cells [31]. Studies have shown that these receptors can bind to nicotine, N'-nitrosonornicotine, and other potential carcinogens, leading to gene mutations and tumor formation [32]. Furthermore, Tournier found that the injury-induced acetylcholine receptors regulate intracellular calcium ions in the process of respiratory epithelial wound repair [33], while Kraiss demonstrated that nicotinic acetylcholine receptors play a crucial role in tumorigenesis by regulating adhesion and motility of respiratory epithelial cells [34]. Therefore, we hypothesize that SNPs within the CHRNA5-CHRNA3-CHRN4 cluster could affect subjective cancer risk by altering receptor nicotine binding, normal cell proliferation, cell migration, and wound healing processes. Similarly, HYKK (AGPHD1) may also potentially influence the risk of developing lung squamous cell carcinoma through this same pathway.

The primary strengths of this study lie in the use of Summary-data-based MR analysis and Colocalization analysis, which together utilize genetic variations to estimate causal effects between gene expression and lung cancer risk. The SMR and MR analyses both utilize SNPs as instrumental variables, thereby minimizing confounding factors and biases from reverse causality, thus providing a more accurate assessment of the causal relationship between exposures and outcomes. The difference lies in the fact that SMR utilizes QTLs (quantitative trait loci, such as gene expression levels, protein expression levels, etc.) as exposures. Since genes and proteins are upstream elements in the central dogma and constitute an inherent process in the development of all diseases, SMR inherently possesses greater computational power when identifying drug targets compared to MR. Moreover, SMR effectively performs colocalization analysis as well. Secondly, we employed HEIDI tests and colocalization methods to mitigate biases due to linkage disequilibrium and horizontal pleiotropy. SMR is also equivalent to performing colocalization analysis, but unlike other colocalization methods, it can simultaneously assess the causal relationship between exposure and outcome. Therefore, by utilizing both SMR and colocalization methods, our conclusions become even more robust. Thirdly, the GWAS datasets used in our main analysis had large sample sizes, enhancing the statistical power of our study. Fourthly, replication across different datasets provided additional support for our findings. Moreover, to minimize bias due to various genetic backgrounds, we restricted our sample selection to individuals of European ancestry. Lastly, by leveraging comprehensive gene expression datasets, we adopted a stringent

selection process to evaluate actionable drug targets and uncovered HYKK (AGPHD1) as a treatment target associated with the risk of lung squamous cell carcinoma.

However, our study has certain limitations. Firstly, the eQTL dataset we used was insufficient to cover the entire druggable genome, lacking data on gene expression genetic variations related to the X and Y chromosomes. Secondly, we only utilized cis-eQTL instrumental variables, which minimizes the potential bias from horizontal pleiotropy but may limit our ability to detect additional potential causal associations. Finally, we utilized the lung cancer data from the FinnGen database R10 as our replication cohort. However, within the GWAS data for this lung cancer, there was a substantial disparity in case numbers versus control numbers across the four subtypes, leading to reduced statistical power for detecting associations between genes and phenotypes. This resulted in a relatively low replication rate between our discovery and replication phases. Thus, the causal relationships of the remaining genes with lung cancer require further scrutiny and investigation.

In conclusion, our MR analysis revealed a relationship between the HYKK (AGPHD1) gene and the risk of developing lung squamous cell carcinoma, consistent with prior randomized controlled trials and meta-analysis studies. This gene represents a promising therapeutic target for both prevention and treatment of lung squamous cell carcinoma, yet its functional mechanisms require further experimental investigation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12449-6>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank all participants and the eQTLGen consortium, The GTEx consortium, Transdisciplinary Research for Cancer of Lung (TRICL) and The International Lung Cancer Consortium (ILCCO), the FinnGen study for sharing data.

Authors' contributions

Wenfu Song was responsible for writing the article, collecting and organizing data; Shiling Sun oversaw the quality control of the paper, taking overall responsibility and supervising its management. Wenfu Song, Yingying Li and Yaxuan Yao performed statistical processing, results analysis, and interpretation. Bing Wang and Xutao Guan contributed to the revision of the manuscript.

Funding

This research received funding from Mr. Guan Xutao. Foundation of Henan Educational Committee (24A360006)

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Ethical approval was not required for this study as all data used were sourced from publicly available databases.

Consent for publication

All authors and their work units agree to the publication of the article.

Competing interests

The authors declare no competing interests.

Received: 10 March 2024 Accepted: 29 May 2024

Published online: 04 June 2024

References

1. Ferlay J, et al. GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC cancer base no. 11. Lyon: International Agency for Research on Cancer; 2013.
2. Miller KD, et al. Cancer treatment and survivorship statistics, 2022. *CA Cancer J Clin.* 2022;72:409–36.
3. Reck M, Heigener DF, Mok T, Soria JC, Rabe KF. Management of non-small-cell lung cancer: recent developments. *Lancet.* 2013;382:709–19.
4. Chen S, Wu S. Identifying lung cancer risk factors in the elderly using deep neural networks: quantitative analysis of web-based survey data. *J Med Internet Res.* 2020;22:e17695.
5. Legha SS, Muggia FM, Carter SK. Adjuvant chemotherapy in lung cancer: review and prospects. *Cancer.* 1977;39(4):1415–24.
6. Novello S, Torri V, Grohe C, et al. International Tailored Chemotherapy Adjuvant (ITACA) trial, a phase III multicenter randomized trial comparing adjuvant pharmacogenomic-driven chemotherapy versus standard adjuvant chemotherapy in completely resected stage II-IIIa non-small-cell lung cancer. *Ann Oncol.* 2022;33(1):57–66.
7. Wu YL, Zhou C, Liam CK, et al. First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. *Ann Oncol.* 2015;26(9):1883–9.
8. He Q, Qu M, Bao H, et al. Multiple post-translational modifications ensure EGFR functionality: potential therapeutic targets to overcome its drug-resistance mutations. *Cytokine Growth Factor Rev.* 2023;70:41–53.
9. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 2019;15(12):e1008489.
10. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015;47:856–60.
11. Finan C, Gaulton A, Kruger FA, et al. The Druggable genome and support for target identification and validation in drug development. *Sci Transl Med.* 2017;9:383.
12. Sekula P, Del Greco MF, Pattaro C, et al. Mendelian randomization as an approach to assess causality using observational data. *JASN.* 2016;27:3253–65.
13. Gaziano L, Giambartolomei C, Pereira AC, et al. Actionable Druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med.* 2021;27:668–76.
14. Storm CS, Kia DA, Almrathi MM, et al. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the Druggable genome. *Nat Commun.* 2021;12:7342.
15. Swerdlow DL, Holmes MV, Kuchenbaecker KB, et al. The Interleukin-6 receptor as a target for prevention of coronary heart disease: a Mendelian randomisation analysis. *Lancet.* 2012;379:1214–24.
16. Schmidt AF, Finan C, Gordillo-Marañón M, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun.* 2020;11:3255.
17. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48:481–7.

18. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet.* 2017;49(7):1126–32.
19. Vösa U, Claringbould A, HarmJan W, 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(9):1300–10.
20. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, NY).* 2020;369(6509):1318–30.
21. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature.* 2023;613:508–18. <https://doi.org/10.1038/s41586-022-05473-8>.
22. 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(5):e1004383.
23. Gan-Or Z, Dion PA, Rouleau GA. Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. *Autophagy.* 2015;11(9):1443–57.
24. Bossé Y, Li Z, Xia J, et al. Transcriptome-wide association study reveals candidate causal genes for lung cancer. *Int J Cancer.* 2020;146(7):1862–78.
25. Hong W, Zhan-di L. Genetic association between AGPHD1 variant and lung cancer risk. *Cell Biochem Biophys.* 2014;70(3):1963–8.
26. Gu M, Dong X, Zhang X, et al. Strong association between two polymorphisms on 15q25.1 and lung cancer risk: a meta-analysis. *PLoS One.* 2017;7(6):e37970.
27. Le Z, Tian-Bo J, Ya G, et al. Meta-analysis of the association between the rs8034191 polymorphism in AGPHD1 and lung cancer risk. *Asian Pac J Cancer Prev.* 2015;16(7):2713–7.
28. Liu C, Cui H, Gu D, et al. Genetic polymorphisms and lung cancer risk: evidence from meta-analyses and genome-wide association studies. *Lung Cancer.* 2017;113:18–29.
29. Sun H, Pan Y, He B, et al. Different effects of the three polymorphisms on 15q25.1 on lung cancer risk: evidence from published literatures. *J Cancer Res Ther.* 2016;12(1):12–9.
30. Qi W, Juntao K, Qibin S, et al. The SNP rs931794 in 15q25.1 is associated with lung cancer risk: a hospital-based case-control study and meta-analysis. *PLoS One.* 2015;10(6):e0128201.
31. Minna JD. Nicotine exposure and bronchial epithelial cell nicotinic acetylcholine receptor expression in the pathogenesis of lung cancer. *J Clin Invest.* 2003;111:31–3. PMID: 12511585.
32. Improgo MR, Scofield MD, Tapper AR, Gardner PD. From smoking to lung cancer: the CHRNA5/A3/B4 connection. *Oncogene.* 2010;29:4874–84. <https://doi.org/10.1038/onc.2010.256>.
33. Tournier JM, Maouche K, Coraux C, et al. alpha3alpha5beta2-Nicotinic acetylcholine receptor contributes to the wound repair of the respiratory epithelium by modulating intracellular calcium in migrating cells. *Am J Pathol.* 2006;168:55–68. PMID:16400009.
34. Kraus AM, Hautefeuille AH, Cros MP, et al. CHRNA5 as negative regulator of nicotine signaling in normal and cancer bronchial cells: effects on motility, migration and p63 expression. *Carcinogenesis.* 2011;32:1388–95. <https://doi.org/10.1093/carcin/bgr090>. PMID: 21586512.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.