Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Qiying Wei Date

Genetic risk score-based gene-smoking interaction association with heart failure

By

Qiying Wei

Master of Public Health

Epidemiology

_____

Yan V. Sun, Ph.D

Committee Chair

Genetic risk score-based gene-smoking interaction association with heart failure

By

Qiying Wei

B.S.

Fudan University

2018

_____

Yan V. Sun, PhD

Committee Chair

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health in Epidemiology

2020

Abstract

Genetic risk score-based gene-smoking interaction association with heart failure

By Qiying Wei

Background: Multiple genetic and environmental factors contribute to the individual's risk for developing heart failure. Previous studies have focused on either genetic susceptibility or environmental risk factors (such as cigarette smoking) for heart failure. However, the gene-environmental interaction effect has not been well studied for heart failure. The goal of this study is to examine gene-smoking interactions on heart failure risk using a genetic risk score (GRS) based on recent findings from genome-wide association study (GWAS) of heart failure.

Methods: A heart failure GRS including 12 loci were calculated in the UK Biobank study to examine gene-smoking interactions with heart failure using logistic regression models, controlling for covariates including age, sex, alcohol intake frequency, body mass index, diabetes history and socioeconomic status. Additionally, the gene-smoking interaction of individual genetic variants association with heart failure was assessed.

Results: We identified significant synergistic interaction between the GRS and smoking status for heart failure. While GRS was positively associated with heart failure within ever-smokers or never-smokers, GRS was more strongly associated with heart failure among smokers (OR= 1.74 among ever-smokers, OR = 1.39 among never-smoker). Furthermore, compared to the low GRS group, smoking showed higher risk for heart failure among the high GRS group (OR = 1.28, 95%CI: 1.15-1.42), while the smoking effect was diminished among people with low GRS for heart failure (OR = 1.02, 95%CI: 0.90-1.16). In the analysis of a single genetic variant, we identified significant gene-smoking interaction of BAG3 locus on heart failure.

Conclusions: Genetic risk modifies the impact of smoking on heart failure. Cigarette smoking poses particularly higher risk among people with high genetic risk for heart failure.

Key words: Gene-environment interaction, genetics, smoking, heart failure, genetic risk score

Length: 270 words

Genetic risk score-based gene-smoking interaction association with heart failure

By

Qiying Wei

B.S.

Fudan University

2018

Thesis Committee Chair: Yan V. Sun, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health in Epidemiology

2020

Abbreviations and Acronyms

ATXN2, *Human Ataxin-2*; AGAP5, *ArfGAP With GTPase Domain, Ankyrin Repeat And PH Domain 5*; BAG3, *Bcl2-associated athanogene 3 protein*; CDKN1A, *Cyclin Dependent Kinase Inhibitor 1A*; CI, *confidence interval*; FTO, *FTO Alpha-Ketoglutarate Dependent Dioxygenase protein*; HERMES, *Heart Failure Molecular Epidemiology for Therapeutic Targets*; HF, *heart failure*; GRS, *genetic risk score*; ICD, *International Classification of Diseases*; KLHL3, *Kelch Like Family Member 3 protein*; OR, *odds ratio*; SNP, *single nucleotide polymorphism*; SYNPO2L, *Synaptopodin 2 Like protein*; TDI, *Townsend deprivation index*;

Introduction

Epidemiology of heart failure

Heart failure is defined as a chronic and progressive syndrome caused by myocardial dysfunction of ventricular filling or blood ejection (Yancy, et al., 2013). In Europe, all-cause heart failure has been reported to be rising with a median prevalence of 11.8% (range 4.7–13.3%) over the past decade (Van Riet, et al., 2016). In contrast to vast progress being made in the treatment of other cardiac disorders, heart failure reportedly affects at least 26 million individuals globally, with an increasing projected burden of disease with the aging populations (Savarese & Lund, 2017). In addition, heart failure is an important contributor to mortality, with an increasing number of five-year mortality rates due to heart failure by 24.4 % for 60 years-olds and 54.4% for 80 year-olds (Thomas & Rich, 2007). Today, people aged over 65 are estimated to occupy more than 20% of the global population (Rudolph, et al., 2019), the prevalence and prognosis of heart failure is becoming an increasing public health concern. Furthermore, heart failure dramatically impacts patients' quality of life and ability to conduct normal functions of life, which in turn, places a heavy burden on the health care system.

While most people are at risk for heart failure, it does not affect the entire population homogeneously. There are several factors associated with the risk of heart failure including male sex, high body mass index (BMI), abdominal fat accumulation, elevated fasting blood glucose, elevated systolic blood pressure, elevated apolipoprotein B/ apolipoprotein A ratio, and cigarette smoking (Maheedhar, et al., 2015). These risk factors tend to occur in combinations in patients, and thus, it is challenging to understand the underlying mechanism of factors individually. Specifically, smoking status is a common modifiable lifestyle risk factor which has been extensively studied in relation to heart failure. According to Jackson Heart Study, 4129 black participants without a history of heart failure or cardiovascular disease at baseline, after smoking cessation of more than 15 years, the heart failure risk of former light smokers who smoked less than 32 pack-years declined to a similar level with never smokers. Also, smokers of light smoking intensity showed a lower risk of heart failure development than current heavy smokers (Kamimura, et al., 2018).

Etiology of heart failure

Heart failure is a terminal stage of various cardiovascular diseases such as monogenic, hypertrophic and dilated cardiomyopathy (Howie, et al., 2009) and has become one of the focal points of heart disease prevention and treatment strategies. Within heart failure, there are two types, the right heart failure and the left heart failure. In most cases, right heart failure occurs after left heart failure, which is caused by elevated pulmonary circulation resistance and increased compensatory pressure of the right heart that eventually leads to right heart decompensation. Moreover, patients with heart failure could be stratified into acute and chronic heart failure based on ventricular remodeling and myocardial fibrosis, which is a consequence of cardiovascular problems with impaired left ventricular function with either reduced or preserved ejection fraction (Howie, et al., 2009).

Despite a wide array of signs and symptoms that present in heart failure patients, heart failure has been attributed to specific subsets and stages diagnosed clinically. These include classifying stage A and stage

B as asymptomatic patients with or without structural abnormalities such as LV hypertrophy, dilation or dysfunction (Shah, et al., 2017), and stage C and D with related symptoms such as exercise intolerance, muscle fatigue, dyspnea and edema in the form of leg swelling or ascites (Coats, et al., 1994). According to the American Heart Association and American College of Cardiology's A-to-D staging system, stage A and stage B heart failure are pre-symptomatic and are often missed at the time of diagnosis.

Pathways underlying heart failure can arise from cardiac lesions at any level, including the myocardium, vasculature, pericardium, heart valves, electrical system or a combination of cardiac abnormalities (Snipelisky, et al., 2019). The modern mechanisms involved in heart failure can be virtually conjectured by several established models: Hemodynamic model, Cardiorenal model, Neurohumoral model, Abnormal Ca^{2+} cycling model and Cell death model (Maheedhar, et al., 2015). At the genetic level, aberrant gene expression can impact contractility, calcium handling and myocardial energetics caused maladaptive changes in cardiac function (Dirkx, et al., 2013).

Genetic pathways of heart failure development

At the genetic level, multiple studies have investigated the association between clinical heterogeneity of heart failure and genetic susceptibility. Multigenerational cohorts from the Framingham study estimated that a parental heart failure prospectively leads to at least a 1.7 times increased risk of heart failure in the offspring. This study also found that 18% of the heart failure burden in the offspring was attributable to the parental heart failure. Additionally, the authors further suggested that genetic factors related to responses to stress can be transmitted from parent to child. The inheritance altered diastolic function maladaptively, increasing vascular stiffness and a propensity for sodium. (Lee, et al. 2006). Moreover, a nationwide Swedish adoption study has also investigated the heritability of heart failure. Results obtained from 21,643 adoptees, 35,016 adoptive parents and 43,286 biological parents showed that heart failure in a biological parent is a significant risk factor for child (OR=1.45, 95%CI: 1.04-2.03. OR =1.58, 95%CI: 1.03-2.42 after exclusion of cardiomyopathies) (Lindgren, et al., 2018). According to the Framingham

Offspring study (1,497 participants' routine echocardiography), heart failure is related to the rising prevalence of left ventricle (LV) systolic dysfunction when analyzed cross-sectionally. In the longitudinal analysis, the study found that inheritable factors boosted the risk of heart failure. (Lee, et al. 2006)

Risk of single nucleotide polymorphisms (SNPs) on heart failure has also been investigated by genome-wide association studies (GWAS) studies. Previous work conducted on the UK Biobank identified 14 regions of the human genome critical for the shape and function of the left ventricle. This study also determined that genes in those regions play an important role in regulation of early ventricular development and contraction. These regions function on LVEDV, LVESV, LVMVR and LVEF and influence on blood ejection in the left ventricle. In early stages, the heart has such a strong compensatory capability that patients could be asymptomatic with hypertrophy. In later stages of heart failure, the compensatory regulation is not able to sustain adequate cardiac output, resulting in weakness, shortness of breath, edema and other symptoms. Analysis of MRI images concluded that genetic factors contribute 22%-39% of pathologic LV size and function, which directly hampered the cardiac pumping into the aorta (Aung, et al., 2019). Left ventricular dilatation blocks the myocardium in contraction and pumping function, putting patients at high risk of cardiovascular diseases. Genes associated with development of heart failure are related to three major mechanisms. MYOZ1and SYNPO2L are highly related to cardiac development, BAG3 affects the protein homeostasis, and CDKN1A is associated with cellular senescence. (Shah, et al., 2020) Furthermore, according to the largest GWAS of heart failure to date, 12 independent variants from 11 independent loci were significantly associated with heart failure. Eight out of 12 of the lead SNPs were located in introns of known genes, while others were located in intergenic regions or 3'-UTR. (Shah, et al., 2020)

Research foundation and goals

While factors such as genetics and lifestyle have been extensively studied in relation to heart failure, their joint effects remain to be investigated. The interaction between biologic and lifestyle/behavioral causes of

cardiovascular disease is a very important mechanism for understanding both the etiology and public health impact of heart failure. To comprehensively describe the effects of genetic predictors of heart failure, it is necessary to place genetic risk within the context of known environmental risk factors. The joint effect of gene-environment risk factors, either additive or multiplicative, is generally greater than the contribution of either alone (Flowers, et al., 2012). Thus, there is a need to determine whether environmental factors modify the association between genetic risk factors and heart failure.

Methods

Study population

Data used in this study were from the UK Biobank, a population-based prospective cohort study. UK Biobank provided a rich collection of phenotypic data from 502,616 volunteer participants aged between 40 and 69 at the time of recruitment. Participants provided signed consent electronically and provided information on social demographics, lifestyle, physical measures and lab (Bycroft, et al., 2018).

To eliminate race and ethnicity confounding bias, we selected Caucasian subjects with adequate genetic information. In order to further refine the subset of participants, we used ethnicity information from UK Biobank to define Caucasian participants categorized as White, including British, Irish and any other White background. The Mixed, Asian or Asian British, Black or Black British, Chinese, other ethnic group or no clear answer were excluded.

Definition of heart failure

The detailed case definitions and the UK Biobank Field codes are shown in the Supplementary Table 1. The case definition referred to Aragam's article on phenotypic refinement of heart failure based on UK Biobank (Aragam, et al., 2019). We extracted diagnosis of heart failure from both Hospital Episode

Statistics (HES) and self-reported health information. A participant is defined as a heart failure case if the person was classified as heart failure and cardiomyopathy for one or more times in self-reported non-cancer illness, main ICD9/ICD10 diagnosis or secondary ICD9/ICD10 diagnosis. All other eligible controls without these data codes were classified into controls. To control for the bias from potential heart failure cases in stage A and stage B, only controls free of vascular or heart problems were included, which was figured out by doctor diagnosis (UK Biobank Field ID:6150).

Statistical analysis

We used data at baseline to assess the basic characteristics and logistic regression analysis on the gene-environment interaction between genetic risk score and smoking status to the risk of heart failure. Baseline characteristics of our study sample were calculated across heart failure cases and controls as percentages for categorical variables and mean with standard deviation for continuous variables. Empirical risk factors of heart failure such as smoking, sex, alcohol intake frequency, diabetes, age, body mass index, socioeconomic status was obtained by programs in Python 3.4 in a Linux system.

Genetic data were extracted using Plink2, which produced observations of single SNP and weighted genetic risk score for each individual. Heart failure genetic risk score (GRS) was calculated according to 12 significant genes related to heart failure (Shah, et al., 2020). The score was based on a weighted sum for genetic risks of each loci:

$$GRS_j = w_1 x_{1j} + w_2 x_{2j} + ... + w_m x_{mj}$$

Where j denotes the j_{th} individual, the weights w_1, w_2, ..., w_m reflects the effect sizes (according to Odds Ratios of each loci) and estimated in single SNP analyses of the trait. In the assessment, GRS was standardized in order to moderate the effect of gene on heart failure (Figure 1). In the initial data analysis step, genetic risk score was categorized into three levels (bottom 0%-20%, 20%-80%,80%-100%).

Logistic regression was used to examine the effect of each genetic or environmental risk factor and its efficacy. The data analysis was conducted in R 3.6.1 and SAS 9.4 in Linux system. All models were adjusted for their gender, age, alcohol intake frequency, BMI, diabetes history and socioeconomic status. The full model was originally set as:

ln (odds of Heart Failure) = $\beta 0$ + $\beta 1$ * Smoking + $\beta 2$ * GRS + $\gamma 1$ * Age + $\gamma 2$ * Sex + $\gamma 3$ * Alcohol + $\gamma 4$ * BMI + $\gamma 5$ * Diabetes + $\gamma 6$ * TDI + $\delta 1$ * Smoking * GRS + $\delta 2$ * Age * GRS + $\delta 3$ * Sex * GRS + $\delta 4$ * Alcohol * GRS + $\delta 5$ * BMI * GRS + $\delta 6$ * Diabetes * GRS + $\delta 7$ * TDI * GRS

Where smoking status was dichotomized into ever-smoked and never-smoked; alcohol intake frequency was classified into three times a week or above and below three times a week; diabetes history was dichotomized into yes or no, and GRS, age, BMI and TDI were treated as continuous variables. TDI stands for Townsend deprivation index at recruitment, which is an indicator for socioeconomic status Higher TDI represents low socioeconomic status.

Stepwise regression was used for modelling. Backward elimination was applied to obtain the final model for assessment, which starts with a model with all of the candidate covariates as predictors. The least significant covariate is dropped and the model refit. The least significant covariate in the new model is dropped until its p value is below α=0.05. The reduced model was set as:

ln (odds of Heart Failure) = $\beta 0$ + $\beta 1$ * Smoking + $\beta 2$ * GRS + $\gamma 1$ * Age + $\gamma 2$ * Sex + $\gamma 3$ * Alcohol + $\gamma 4$ * BMI + $\gamma 5$ * Diabetes + $\gamma 6$ * TDI + $\delta 1$ * Smoking * GRS

Results

The flow chart of data selection was shown in Figure 2. After excluding non-Caucasian participants (n = 56,589), participants without genetic information (n = 2,778) and controls with potential heart failure development (n = 23,280), a total of 419,969 participants were included in the statistical analysis stage.

Demographic information and distribution of risk factors to heart failure by case, control and total population with two-sample test p-values are listed in Table 2. Based on 419,969 participants, 55.13% have never smoked in their lifetime; 34.43% smoked previously but quit before data collection, and 10.09% were current smokers. 63.56% of the cases and 44.58% of the controls have smoked in their lifetime. The smoking status had a significant influence on the outcome of heart failure across cases and controls, indicating present smokers had a higher odds of developing heart failure than past smokers. Both present and past smokers had a higher chance of developing heart failure than non-smokers. ($p<0.001$) Similarly, male gender, alcohol intake frequency (above three times a week), diabetes history, elder age, obesity and low socioeconomic status were all significant risk factors for heart failure. ($p<0.001$) Heart failure risk was also elevated monotonically in accordance with genetic risk scores. ($p<0.001$)

Under the same smoking status, increased genetic risk leads to an increased risk of developing heart failure. Among the smoking population, the odds of heart failure among the high-GRS group was 1.74 (95%CI, 1.54-1.97) times the odds of heart failure among the low-GRS group; the odds of heart failure among the intermediate GRS group was 1.34 (95%CI, 1.21-1.50) times the odds of heart failure among the low-GRS group. Among the non-smoking population, the odds of heart failure among the high-GRS group was 1.39 (95%CI, 1.24-1.55) times the odds of heart failure among the low-GRS group; the odds of heart failure among the intermediate-GRS group was 1.11 (95%CI, 1.01-1.22) times the odds of heart failure among the low-GRS group. All the results indicated that for the same genetic risk level, smoking is a significant risk factor for heart failure. These results were shown in the Supplementary Table 3.

Moreover, controlling for all the environmental covariate, the risk of higher genetic risk score level was associated with risk of heart failure development, indicating that smoking status has a gradually increasing impact on the development of heart failure for individuals with higher GRS levels. Compared to the low-GRS group, smoking has a more significant influence on heart failure among the high-GRS group (OR = 1.28, 95% CI: 1.15-1.42), while the effect is almost the same between smokers and non-smokers in low-GRS level (OR = 1.02, 95% CI: 0.90-1.16). The estimates and 95% confidence intervals were included in Table 4.

In the logistic regression analysis, we applied the backward elimination to identify the appropriate predictive variables to fit in the regression models because most of the gene-environment interaction terms were insignificant. After the backward elimination step, only smoking status had a statistically significant interaction with GRS. The estimates of the final model (Model 6) was adopted in Table 5. Additionally, the relationship between probability of heart failure for smokers and non-smokers was presented in Figure 4. After controlling for age, sex, alcohol intake frequency, BMI, diabetes and socioeconomic status, people in the higher GRS group are more likely to get heart failure by smoking status. (β=0.18, 95% CI: 0.12-0.23)

To evaluate the more detailed gene-environmental interaction between SNPs and smoking on heart failure, we measured the interaction assessment replacing the genetic risk score with the 12 loci indicated in HERMES study. SNP rs17617337 (BAG3) showed a significant gene-smoking interaction on heart failure after adjusting for known risk factors (Table 6). Interestingly, while most SNPs reported in previous studies showed significance association with heart failure (Table 6), three SNPs were insignificantly associated with heart failure in the UK Biobank sample (rs4746140 in SYNPO2L/AGAP5, rs11745324 in KLHL3 and rs56094641 in FTO).

Discussion

Our study aimed to examine the gene-smoking interaction associated with risk of heart failure. Cigarette smoking is an established risk factor for heart failure. Both current and previous smokers have a higher possibility of developing heart failure relative to non-smokers. We found that the prevalence of heart failure between previous smokers and current smokers is not significantly different, which is consistent with previous findings by Kamimura, et al., 2018. In addition to smoking status, factors such as male gender, alcohol intake frequency (above 3 times a week), diabetes history, elder age, obesity and low socioeconomic status all contribute to higher heart failure incidence. The GRS was significantly associated with heart failure, although three individual SNPs were not significantly associated with heart failure in the present study of UK Biobank samples.

The significant GRS×Smoking interaction on heart failure can partially explain observed inter-individual variability of heart failure risk. On one hand, in the population there is a range of genetic risk profiles, with each individual adopting a unique risk spectrum, from a low to a high genetic risk. On the other hand, the genetic risk occupies a position along the environmental spectrum according to an individual's lifestyle. Meanwhile, the genetic risk and environmental risk (i.e. smoking) collaborates to affect the outcome of disease. This interaction may help explain the multifactorial pathologies that are seen from the relationship between genetic risk and smoking status. Smoking status has a gradually increasing impact on the development of heart failure for individuals with higher GRS levels than for individuals with lower GRS levels. After controlling for age, sex, alcohol intake frequency, BMI, diabetes and socioeconomic status, the probability of developing heart failure among smokers is significantly elevated in high and intermediate GRS groups, while the probability was not increased substantially among nonsmokers.

The synergistic interaction between heart failure GRS and cigarette smoking also emphasizes the benefit of eliminating smoking exposure particularly among people with higher genetic susceptibility of heart failure.

There are several strengths to this study, including the first exploration of gene-environmental interactions on heart failure using a GWAS-based GRS and a large population sample for a robust estimate. The gene-environment interaction results may help account for the unexplained genetic risk factors on developing heart failure. This study also included the use of standardization method to ensure the efficacy and accuracy as well. Several limitations of our study merit attention. Firstly, dichotomous lifestyle such as smoking and alcohol may limit detailed changes within the same classified group. By dichotomizing smoking, we assumed that previous smokers and current smokers were in the same category and therefore share the same lifestyle risk of heart failure. This may not be precise, however, because although the former smokers (HR = 1.44, 95%CI: 0.98-2.12) showed lower risk of heart failure than the current smokers (HR = 1.44, 95%CI: 0.98-2.12), and it may not differ significantly from the current smokers(HR = 2.82, 95%CI: 1.71-4.64), the range is too big to make such conclusion (Kamimura, et al., 2018). Furthermore, because we used both non-cancer illness code, which is a self-reported measurement, and ICD code, which was reported from the medical organizations for heart failure definition, the two methods may not stand consistent with each other. Additionally, the study was constructed on the genetic level, which may not take into account epigenetic and polymorphisms in gene or protein expression. Thus, further analysis should be done to analyze these effects for other genetic effects.

Our findings of gene-smoking interaction do not establish a causal relation of gene-environment interaction to the disease process. Although the genetic risk factors are hard coded in the genetic code, the environmental factors are modifiable. For people with higher susceptibility to heart failure, elimination of environmental risk exposures would reduce more risk of developing heart failure. Interventions such as smoking cessation, which was found to be a protective factor for developing heart failure, would be beneficial. In addition to this analysis on gene-smoking interaction, future studies may future explore the gene-environmental interactions of other environmental factors, as well as more broadly the context-dependent genetic effects (gene-sex, gene-age and gene-gene interactions) to explain the genetic propensity to heart failure.

Acknowledgement

This research has been conducted using the UK Biobank Resource under Application Number ‘34031’.

I would like to express my special thanks of gratitude to my instructor Dr. Yan V. Sun who helped me a lot in building the structure and finalizing the project within the limited time frame, Secondly, I would also like to thank Dr. Yunfeng Huang and Qin Hui who gave me detailed instructions on Linux system manipulation, programming and methodology.

References

1. Aragam, K. G., Chaffin, M., Levinson, R. T., McDermott, G., Choi, S. H., Shoemaker, M. B., ... & Newton-Cheh, C. (2019). Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. Circulation, 139(4), 489-501.

2. Aung, N., Vargas, J. D., Yang, C., Cabrera, C. P., Warren, H. R., Fung, K., ... & Manichaikul, A. W. (2019). Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. Circulation, 140(16), 1318-1330.

3. Bleumink, G. S., Schut, A. F., Sturkenboom, M. C., Deckers, J. W., van Duijn, C. M., & Stricker, B. H. C. (2004). Genetic polymorphisms and heart failure. Genetics in Medicine, 6(6), 465-474.

4. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... & Cortes, A. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203-209.

5. Coats, A. J., Clark, A. L., Piepoli, M., Volterrani, M., & Poole-Wilson, P. A. (1994). Symptoms and quality of life in heart failure: the muscle hypothesis. British heart journal, 72(2 Suppl), S36.

6. Dirkx, E., da Costa Martins, P. A., & De Windt, L. J. (2013). Regulation of fetal gene expression in heart failure. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1832(12), 2414-2424.

7. Flowers, E., Froelicher, E. S., & Aouizerat, B. E. (2012). Gene-environment interactions in cardiovascular disease. European Journal of Cardiovascular Nursing, 11(4), 472-478.

8. Guo, Y., Zhao, D., & Liu, J. (2015). Epidemiological study of heart failure in China. Cardiovascular Innovations and Applications, 1(1), 47-55. doi:10.15212/CVIA.2015.0003

9. Heart Failure. Maheedhar Gedela Muhammad Khan Orvar Jonsson. South Dakota journal of medicine. , 2015, Vol.68(9), p.403-5, 407

10. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genet. 5, e1000529 (2009)

11. Kamimura, D., Cain, L. R., Mentz, R. J., White, W. B., Blaha, M. J., DeFilippis, A. P., ... & Butler, J. (2018). Cigarette smoking and incident heart failure: insights from the Jackson Heart Study. Circulation, 137(24), 2572-2582.

12. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases.Nat Genet. 2018;50:390–400.

13. Lee DS, Pencina MJ, Benjamin EJ, et al. Association of parental heart failure with risk of heart failure in offspring. N Engl J Med. 2006;355 (2):138-147. doi:10.1056/NEJMoa052948

14. Lindgren, M. P., PirouziFard, M., Smith, J. G., Sundquist, J., Sundquist, K., & Zöller, B. (2018). A swedish nationwide adoption study of the heritability of heart failure. JAMA cardiology, 3(8), 703-710.

15. McMurray, J. J., & Stewart, S. (2000). Epidemiology, aetiology, and prognosis of heart failure. Heart, 83(5), 596-602.

16. Roger, V. L., Weston, S. A., Redfield, M. M., Hellermann-Homan, J. P., Killian, J., Yawn, B. P., & Jacobsen, S. J. (2004). Trends in heart failure incidence and survival in a community-based population. Jama, 292(3), 344-350.

17. Rudolph, C. W., Marcus, J., Zacher, H., Rudolph, C. W., Marcus, J., Mah, R., ... & Zacher, H. (2019). Global issues in work, aging and retirement. Aging and work in the 21st century, 292-324.

18. Savarese, G., & Lund, L. H. (2017). Global public health burden of heart failure. Cardiac failure review, 3(1), 7.

19. Shah, A. M., Claggett, B., Loehr, L. R., Chang, P. P., Matsushita, K., Kitzman, D., ... & Wright, J. D. (2017). Heart failure stages among older adults in the community: The Atherosclerosis Risk in Communities Study. Circulation, 135(3), 224-240.

20. Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjörnsson, G., Fatemifar, G., ... & Helgadottir, A. (2020). Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nature Communications, 11(1), 1-12.

21. Snipelisky D, Chaudhry S-P, Stewart GC. The Many Faces of Heart Failure.Card Electrophysiol Clin. 2019;11:11–20.

22. Talmud, P. J. (2007). Gene–environment interaction and its impact on coronary heart disease risk. Nutrition, metabolism and cardiovascular diseases, 17(2), 148-152.

23. Thomas, S., & Rich, M. W. (2007). Epidemiology, pathophysiology, and prognosis of heart failure in the elderly. Heart failure clinics, 3(4), 381-387.

24. Van Riet, E. E., Hoes, A. W., Wagenaar, K. P., Limburg, A., Landman, M. A., & Rutten, F. H. (2016). Epidemiology of heart failure: the prevalence of heart failure and ventricular

dysfunction in older adults over time. A systematic review. European journal of heart failure, 18(3), 242-252.

25. Wild PS, Felix JF, Schillert A, Teumer A, Chen M-H, Leening MJG, Völker U, Großmann V, Brody JA, Irvin MR, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function.J Clin Invest.2017;127:1798–1812.

26. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/ AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 2013;128(16):1810–52.

27. Ziaeian, B. & Fonarow, G. C. Epidemiology and aetiology of heart failure. Nature Reviews Cardiology (2016). doi:10.1038/nrcardio.2016.25

Tables

Table 1. Heart failure case definition from the UK Biobank

UK Biobank Field ID	UK Biobank Description	Code	Definition
20002	Non-cancer illness code, self-reported	1076	Heart failure/pulmonary edema
20002	Non-cancer illness code, self-reported	1079	Cardiomyopathy
41202, 41204	Diagnosis -main/secondary ICD 10	I11.0	Hypertensive heart disease with (congestive) heart failure
41202, 41204	Diagnosis -main/secondary ICD 10	I13.0	Hypertensive heart and renal disease with (congestive) heart failure
41202, 41204	Diagnosis -main/secondary ICD 10	I13.2	Hypertensive heart and renal disease with both (congestive) heart failure and renal failure
41202, 41204	Diagnosis -main/secondary ICD 10	I25.5	Ischemic cardiomyopathy
41202, 41204	Diagnosis -main/secondary ICD 10	I42.0	Dilated cardiomyopathy
41202, 41204	Diagnosis -main/secondary ICD 10	I42.5	Other restrictive cardiomyopathy
41202, 41204	Diagnosis -main/secondary ICD 10	I42.8	Other cardiomyopathies

41202, 41204	Diagnosis -main/secondary ICD 10	I42.9	Cardiomyopathy, unspecified
41202, 41204	Diagnosis -main/secondary ICD 10	I43.8	Cardiomyopathy in other diseases classified elsewhere
41202, 41204	Diagnosis -main/secondary ICD 10	I50.0	Congestive heart failure
41202, 41204	Diagnosis -main/secondary ICD 10	I50.1	Left ventricular failure
41202, 41204	Diagnosis -main/secondary ICD 10	I50.9	Heart failure, unspecified
41203, 41205	Diagnosis -main/secondary ICD 9	4254	Other primary cardiomyopathies
41203, 41205	Diagnosis -main/secondary ICD 9	4280	Congestive heart failure
41203, 41205	Diagnosis -main/secondary ICD 9	4281	Left heart failure
41203, 41205	Diagnosis -main/secondary ICD 9	4289	Heart failure, unspecified

Table 2. Demographic information and distribution of risk factors of heart failure by case, control and total population with two-sample test p-values

Covariates	Statistics	Level	HF			Parametric P-value* (Case vs Control)
			All N=419969	Case N=6273	Control N=413696	
Smoking status	N (Col %)	Never	231546 (55.13)	2286 (36.44)	229260 (55.42)	**<.001**
	N (Col %)	Previous	144613 (34.43)	2966 (47.28)	141647 (34.24)	
	N (Col %)	Current	42388 (10.09)	991 (15.8)	41397 (10.01)	
Sex	N (Col %)	Female	232893 (55.45)	1885 (30.05)	231008 (55.84)	**<.001**
	N (Col %)	Male	187076 (44.55)	4388 (69.95)	182688 (44.16)	
Alcohol intake frequency	N (Col %)	< 3/week	229898 (54.74)	3803 (60.72)	226095 (54.69)	**<.001**

	N (Col %)	3/w or more	189786 (45.19)	2460 (39.28)	187326 (45.31)	
Diabetes	N (Col %)	No	400966 (95.48)	4980 (79.39)	395986 (95.72)	**<.001**
	N (Col %)	Yes	18099 (4.31)	1252 (19.96)	16847 (4.07)	
Age	Mean		56.58	62	56.5	
	Median		58	63	58	
	Min		38	40	38	**<.001**
	Max		73	70	73	
	Std Dev		8.01	6.17	8	
BMI	Mean		27.32	29.91	27.29	
	Median		26.63	29.05	26.6	
	Min		12.12	16	12.12	**<.001**
	Max		74.68	62.29	74.68	

	Std Dev	4.75	5.77	4.72	
TDI	Mean	-1.56	-0.64	-1.57	
	Median	-2.34	-1.53	-2.35	
	Min	-6.26	-6.26	-6.26	**<.001**
	Max	10.88	10.45	10.88	
	Std Dev	2.93	3.36	2.92	
GRS(standardized)	Mean	0	0.15	0	
	Median	-0.02	0.13	-0.02	
	Min	-4.36	-3.02	-4.36	**<.001**
	Max	5.6	4.26	5.6	
	Std Dev	1	1.01	1	

*Categorical variables are reported as N(%). Continuous variables are reported as mean, median, minimum, maximum and standard deviation.

*GRS stands for genetic risk score, standardized to fit normal distribution N(0,1).

*TDI stands for Townsend deprivation index at recruitment, which is an indicator for socioeconomic status. Higher TDI represents low socioeconomic status.

Table 3. Odds Ratios of heart failure according to tercile genetic risk score on different smoking status

Smoking status	GRS group	N	Point Estimate (95%CI)	p-Value
Ever smoked	**High vs Intermediate**	28411	1.292 (1.180, 1.414)	<.0001
	High vs low	82456	1.742 (1.543, 1.966)	<.0001
	Intermediate vs low	29867	1.348 (1.211, 1.501)	<.0001
Never smoked	**High vs Intermediate**	53672	1.249 (1.145, 1.362)	<.0001
	High vs low	156207	1.389 (1.244, 1.550)	<.0001
	Intermediate vs low	57651	1.112 (1.011, 1.223)	0.0283

*Tercile GRS: 0%-20% (Low), 20%-80%(Intermediate), 80%-100%(High) .

*GRS stands for genetic risk score, standardized to fit normal distribution N(0,1).

Table 4. Odds Ratios of heart failure according to tercile genetic risk score (ever smoked vs never smoked) HF-smoke

Odds Ratio Estimates and Wald Confidence Intervals				
Odds Ratio	**Estimate**	**95% Confidence Limits**		**p-Value**
High GRS (top 20%)	1.280	1.151	1.422	<.0001
Intermediate GRS (20%-60%)	1.237	1.154	1.326	<.0001
Low GRS (bottom 20%)	1.020	0.899	1.158	0.7544

*GRS stands for genetic risk score, standardized to fit normal distribution N(0,1).

Table 5. Summary of backward elimination logistic regression coefficient estimates and p-values

Variable	Model 1				Model 2			
	Estimate	95% CI		p-value	Estimate	95% CI		p-value
Smoke	0.18	0.12	0.23	<.0001	0.18	0.12	0.23	<.0001
GRS	0.06	-0.25	0.37	0.71	0.07	-0.19	0.34	0.59
Age	0.11	0.10	0.11	<.0001	0.11	0.10	0.11	<.0001
Sex	1.05	0.99	1.11	<.0001	1.05	0.99	1.11	<.0001
BMI	0.07	0.07	0.08	<.0001	0.07	0.07	0.08	<.0001
Alcohol	-0.29	-0.35	-0.23	<.0001	-0.29	-0.35	-0.23	<.0001
Diabetes	0.94	0.86	1.01	<.0001	0.94	0.86	1.01	<.0001
TDI	0.08	0.08	0.09	<.0001	0.08	0.08	0.09	<.0001
GRS*smoke	0.06	0.01	0.12	0.02	0.06	0.01	0.12	0.02
GRS*Diabetes	-0.05	-0.12	0.03	0.20	-0.04	-0.11	0.02	0.20
GRS*Sex	0.02	-0.04	0.07	0.57	0.02	-0.04	0.07	0.57
GRS*Age	0.00	0.00	0.01	0.71	0.00	0.00	0.01	0.72
GRS*alcohol	0.01	-0.04	0.06	0.72	0.01	-0.04	0.06	0.73

GRS*BMI	0.00	0.00	0.01	0.88				
Variable	**Model 3**				**Model 4**			
	Estimate	**95% CI**		**p-value**	**Estimate**	**95% CI**		**p-value**
Smoke	0.18	0.12	0.23	0.58	0.18	0.12	0.23	<.0001
GRS	0.07	-0.19	0.34	<.0001	0.12	0.07	0.17	<.0001
Age	0.11	0.10	0.11	<.0001	0.11	0.10	0.11	<.0001
Sex	1.05	0.99	1.11	<.0001	1.05	0.99	1.11	<.0001
BMI	0.07	0.07	0.08	<.0001	0.07	0.07	0.08	<.0001
Alcohol	-0.29	-0.34	-0.23	<.0001	-0.29	-0.34	-0.23	<.0001
Diabetes	0.94	0.86	1.01	<.0001	0.94	0.86	1.01	<.0001
TDI	0.08	0.08	0.09	0.02	0.08	0.08	0.09	<.0001
GRS*smoke	0.06	0.01	0.12	0.72	0.06	0.01	0.12	0.01
GRS*Diabetes	-0.05	-0.11	0.02	0.20	-0.05	-0.11	0.02	0.19
GRS*Sex	0.02	-0.04	0.07	0.19	0.02	-0.04	0.07	0.53
GRS*Age	0.00	0.00	0.01	0.52				
GRS*alcohol								

GRS*BMI								
Variable	**Model 5**				**Model 6**			
	Estimate	**95% CI**		**p-value**	**Estimate**	**95% CI**		**p-value**
Smoke	0.18	0.12	0.23	<.0001	0.18	0.12	0.23	<.0001
GRS	0.13	0.10	0.17	<.0001	0.13	0.09	0.16	<.0001
Age	0.11	0.10	0.11	<.0001	0.11	0.10	0.11	<.0001
Sex	1.05	0.99	1.11	<.0001	1.05	0.99	1.11	<.0001
BMI	0.07	0.07	0.08	<.0001	0.07	0.07	0.08	<.0001
Alcohol	-0.29	-0.34	-0.23	<.0001	-0.29	-0.34	-0.23	<.0001
Diabetes	0.94	0.86	1.01	<.0001	0.93	0.86	1.00	<.0001
TDI	0.08	0.08	0.09	<.0001	0.08	0.08	0.09	<.0001
GRS*smoke	0.07	0.01	0.12	0.01	0.06	0.01	0.12	0.01
GRS*Diabetes	-0.04	-0.11	0.02	0.21				
GRS*Sex								
GRS*Age								
GRS*alcohol								

GRS*BMI		

*TDI stands for Townsend deprivation index at recruitment, which is an indicator for socioeconomic status. Higher TDI represents low socioeconomic status.

*GRS stands for genetic risk score, standardized to fit normal distribution N(0,1).

Table 6. Analysis of Maximum Likelihood Estimates for single SNP or SNP×Smoking interaction associated with prevalence of heart failure.

rsid	Estimate	Std	P-value	Interaction term	Estimate	Std	P-value
rs11745324	-0.02	0.03	0.61	smoke*rs11745324	-0.01	0.05	0.89
rs140570886	0.29	0.07	<.0001	smoke*rs140570886	0.18	0.14	0.20
rs1556516	0.12	0.02	<.0001	smoke*rs1556516	0.05	0.04	0.20
rs17042102	0.13	0.03	<.0001	smoke*rs17042102	0.01	0.06	0.84
rs17617337	-0.11	0.02	<.0001	smoke*rs17617337	-0.09	0.05	0.04
rs4135240	-0.07	0.02	0.00	smoke*rs4135240	-0.03	0.04	0.52
rs4746140	-0.05	0.03	0.05	smoke*rs4746140	-0.04	0.05	0.46
rs4766578	-0.04	0.02	0.02	smoke*rs4766578	-0.04	0.04	0.31
rs55730499	0.15	0.03	<.0001	smoke*rs55730499	-0.05	0.07	0.42
rs56094641	-0.02	0.02	0.42	smoke*rs56094641	0.01	0.04	0.89
rs600038	0.10	0.02	<.0001	smoke*rs600038	0.07	0.05	0.10
rs660240	0.09	0.02	<.0001	smoke*rs660240	0.01	0.05	0.77

Figures

Figure 1. Distribution of weighted genetic risk score

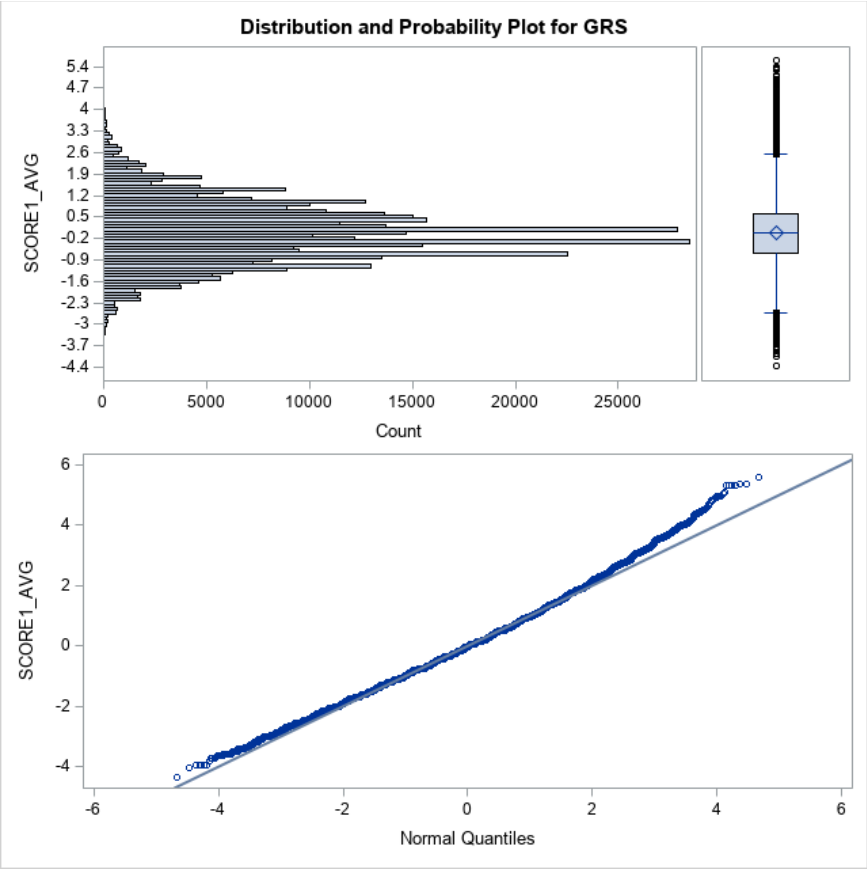
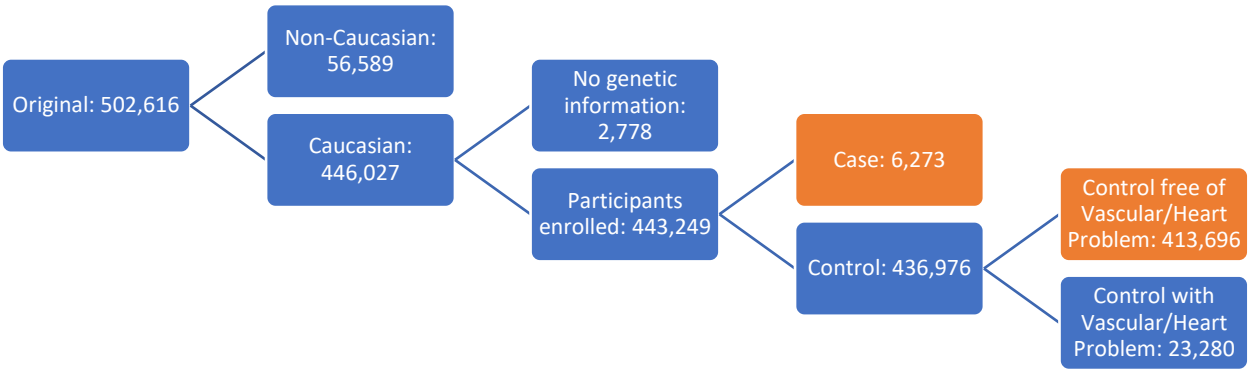
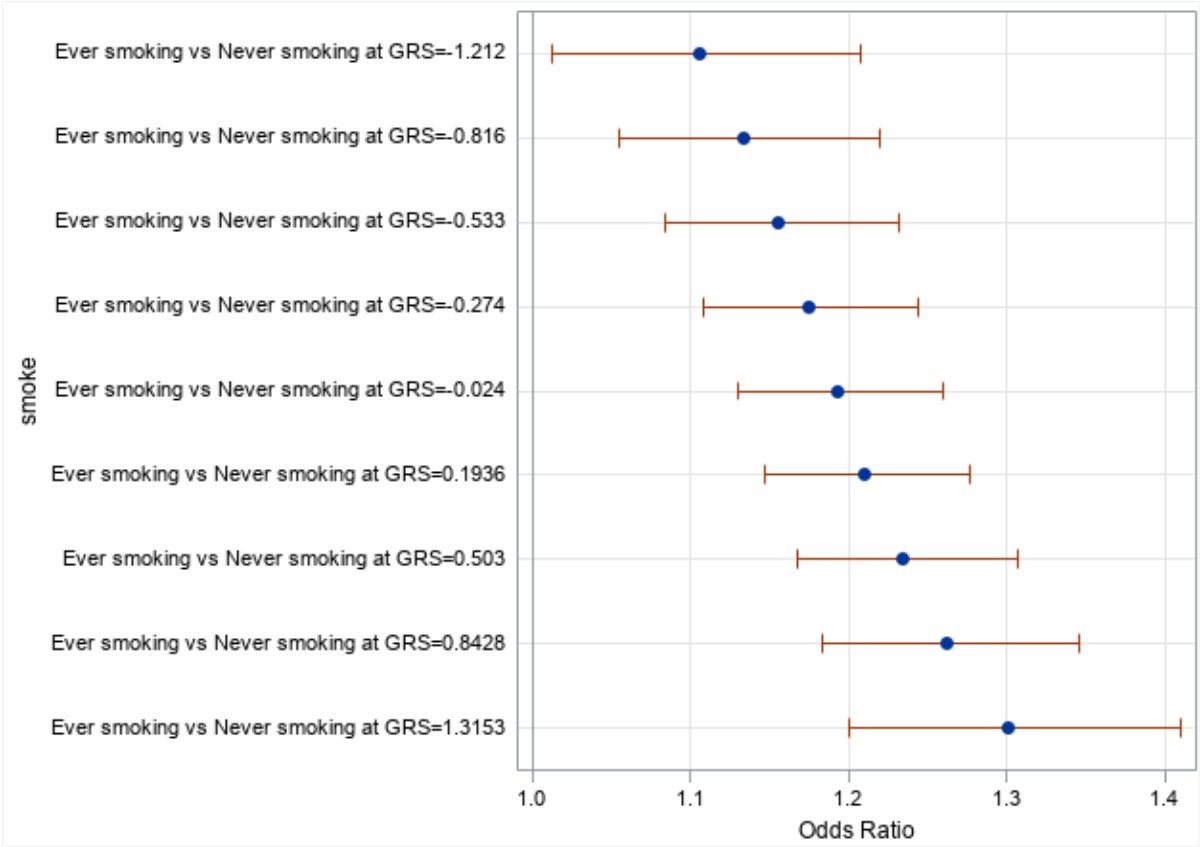


Figure 2. Flow chart of sample selection from the UK Biobank study.



*Heart Failure Cases: Non-cancer illness code, self-reported (1076,1079); Diagnosis -main/secondary ICD 10: I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.8, I42.9, I43.8, I50.0, I50.1, I50.9; Diagnosis -main/secondary ICD 9: 4254, 4280, 4281, 4289.

Figure 3. OR (95%CI) of heart failure among different smoking status (ever-smoking vs never-smoking) in GRS deciles



* GRS values listed in the figure ranked from 10% to 90%.

Figure 4. Predicted probabilities (95%CI) of developing heart failure among different smoking status (ever-smoking vs never-smoking)