creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Master's Thesis of Public Health

Predicting *Aldehyde Dehydrogenase 2 (ALDH2)* Polymorphism: *ALDH2*-genotype and Facial-Flushing-Reaction Mismatching Cases

February 2019

Graduate School of Public Health

Seoul National University

Public Health Science Major

Eun Jin Woo

Predicting *Aldehyde Dehydrogenase 2 (ALDH2)* Polymorphism: *ALDH2*-genotype and Facial-Flushing-Reaction Mismatching Cases

알데하이드 디하이드로즈네이즈 2 다형성 예측 모형: *ALDH2* 유전형-음주 후 안면 홍조 반응 불일치 케이스

지도교수 성 주 헌

이 논문을 보건학 석사학위논문으로 제출함

2018년 11월

서울대학교 보건대학원

보건학과 보건학전공

우 은 진

우은진의 석사학위논문을 인준함

2018년 12월

위 원 장 조 성 일 (인)

부 위 원 장 정 효 지 (인)

위 원 성 주 헌 (인)

Abstract

Backgrounds: *Aldehyde dehydrogenase 2(ALDH2)* gene encodes a critical metabolic enzyme which plays an important role in breakdown of acetaldehyde. In addition, the risk of various alcohol-related diseases and cancers can be altered according to the *ALDH2* genotype. Currently *ALDH2* genotype is predicted by facial flushing reaction questionnaire since it is not possible to genotype every individual. However, some individuals, who consume relatively large amount of alcohol than average, tend to be adapted to facial flushing reaction and hence express opposite phenotype from their *ALDH2* genotype.

Objectives: This study aims to develop *ALDH2* polymorphism prediction model with alcohol related variables to increase predictive performance with current screening method, facial flushing questionnaire and also to explore genes associated with facial flushing reaction others than *ALDH2* and score polygenic effect of associated SNPs.

Methods: 1807 study participants from Korean Healthy Twin study cohort were involved in prediction modelling of *ALDH2* genotype. 2/3 of study participants were included in training dataset and significant alcohol related variables were selected as predictors with no multicollinearity by backward stepwise regression. Candidate prediction models were validated with 1/3 of study participants and predictive performance of models were measured with AUC with ROC curve, Hosmer-Lemeshow test, reclassification. In addition, *ALDH2*(rs671) conditional GWAS was

performed using family-based score test for association (FASTA) from 'GenABEL' in R software and GWAS summary statistics were used to calculate polygenic score.

Results: Selected predictors apart from facial flushing reaction were 1) amount of alcohol consumption (g/week), Hazard domain of AUDIT, AUDIT question No.8. With these variables 5 candidate models were generated and Models with facial flushing reaction, Hazard domain of AUDIT and AUDIT question No.8 as predictors were selected as the best prediction model to discriminate active and inactive *ALDH2* carriers with AUC of 0.93 which is 0.013 higher than facial flushing reaction questionnaire alone. Results from ALDH2(rs671) conditional GWAS had no significant SNPs yet clumped 17 SNPs at suggestive significant level including rs6480460 in *PRF1* were used to generate polygenic score. Regardless of *ALDH2* genotype, individuals with no flushing reaction expressed stronger polygenic effect on metabolic ability to breakdown alcohol.

Conclusion: Facial flushing questionnaire is an excellent screening tool for *ALDH2* polymorphism. However, prediction model with additional alcohol related variables as predictors can improve the predictive performance. Also, inactive *ALDH2* carriers with no facial flushing reaction had stronger polygenic effect of SNPs related to metabolic ability for alcohol intake.

Key words: *Aldehyde dehydrogenase 2 (ALDH2)*, Facial flushing reaction, prediction model, polygenic score, family-based genome wide association study, conditional genome wide association study

Student number: 2017-27710

CONTENTS

Abstract i

List of Tables v

List of Figures vi

I. Introduction

1. Background 1
2. Objectives 3

II. Methods

1. Study participants 5
2. Data measurement
   - 2.1 Alcohol Use Disorders Identification Test (AUDIT) 5
   - 2.2 Average amount of alcohol Intake 6
   - 2.3 The Facial flushing reaction questionnaire 6
   - 2.4 Genotype data 7
3. Statistical analysis 7
   - 3.1 Descriptive analysis 7
   - 3.2 Development and Validation of Prediction model 7
   - 3.3 Polygenic score for Facial flushing reaction 8

III. Results

Part.1

1. Descriptive statistical analysis
   - 1.1 Study participants by sex 9
   - 1.2 Study participants by matching and mismatching case 9
2. Prediction of *ALDH2* polymorphism

2.1 Development of prediction model 13

2.2 Validation of prediction model 14

Part.2

1. *ALDH2* Conditional Genome-wide association study of Facial Flushing Reaction 18

2. Polygenic score for facial flushing reaction 23

IV. Discussion 24

V. References 27

VI. 국문 초록 (Abstract in Korean) 30

List of Tables

Table.1 General Characteristics of study participants 11

Table.2 General Characteristics of study participants by *ALDH2* genotype and facial flushing reaction 12

Table.3 Sensitivity and Specificity of Facial Flushing Questionnaire in Korean Healthy Twin Study 13

Table.4-1 Explanatory models 14

Table.4-2 Coefficient of Explanatory models 14

Table.5 Comparison of Models' predictive performance 17

Table.6 Characteristics of Top 10 SNPs from *ALDH2*(rs671) conditional GWAS for facial flushing reaction used to generate the PS 22

Table.7 Average polygenic score by different groups 23

List of Figures

Figure.1 Diagram of objectives of Part 1 and Part 2 4

Figure.2 ROC Curves of Model.1 and Model.5 15

Figure.3 Discrimination Slope of Model.1 and Model.5 16

Figure.4 Manhattan Plot of SNPs from GWAS for facial flushing reaction 19

Figure.5 Q-Q Plot of GWAS p-value for facial flushing reaction 19

Figure.6 Manhattan Plot of SNPs from *ALDH2*(rs671) conditional GWAS for facial flushing reaction 20

Figure.7 Q-Q Plot of *ALDH2*(rs671) conditional GWAS p-value for facial flushing reaction 20

Figure.8 Regional Plot of *ALDH2*(rs671) conditional GWAS for facial flushing reaction with rs6480460 21

I. INTRODUCTION

1. Background

The Aldehyde dehydrogenase 2 (*ALDH2*) enzyme mostly governs the metabolic breakdown of acetaldehyde from alcohol consumption [1]. Aldehyde dehydrogenase 2 are classified into active and inactive forms by having single nucleotide substitution in the gene encoding Aldehyde dehydrogenase 2, *Aldehyde dehydrogenase 2*(*ALDH2*) [2]. The genetic substitution of adenine to guanine on rs671 of *ALDH2* causes inactive *ALDH2* (*1/*2 or *2/*2) by altering amino acid 487 of mature protein, with lysine substituted for glutamic acid (504 of the gene when the mitochondrial leader sequence is included) [3]. This genetic polymorphism of *ALDH2* however only appears to be prevalent in East Asian population with an incidence of approximately 30% *ALDH2* (*1/*2) heterozygosity and 5–10% *ALDH2* (*2/*2) homozygosity [11,13] including Korean, Japanese and Chinese as Caucasian, Southeast Asian and African population mostly possess active form of *ALDH2* (*1/*1) [4,5]. Inactive *ALDH2* results in reduced enzymatic activity which leads to prolonged accumulation of toxic acetaldehyde in the blood. Inactive-*ALDH2* carriers therefore express physiological response such as facial flushing reaction and other unpleasant feelings after alcohol consumption [10]. In addition, alcohol drinking behavior and alcohol-related diseases and cancers can be strongly influenced by *ALDH2* genotype [7]. Physiological responses caused by inactive *ALDH2* causes

alcohol consumption to be aversive and therefore less prone to alcoholism or alcohol dependence. As such, inactive *ALDH2* carriers are highly protective against alcohol related problems. [5]. Also, In Korean population, susceptibility to the development of gastric cancer associated with alcohol intake was modified by *ALDH2* polymorphism in Korean population especially in case of inactive *ALDH2* genotype and inactive *ALDH2* carriers are reported to have increased risk of esophageal cancer even with moderate drinking as well [9,12]. Other cancers such as oral cancer, breast cancer and colon cancer showed a moderate level of association and still remained as a matter of debate [11,13].

However, *ALDH2* and other genotypes linked to alcohol metabolism are not considered in alcohol drinking guideline in Korea and most study results regarding correlation of alcohol consumption and alcohol-related diseases and cancers have been concluded with non-East Asian population. Therefore, considering inactive *ALDH2* genotype in the study with East Asian population could result more precise association of *ALDH2* genotype and diseases as well as cancers.

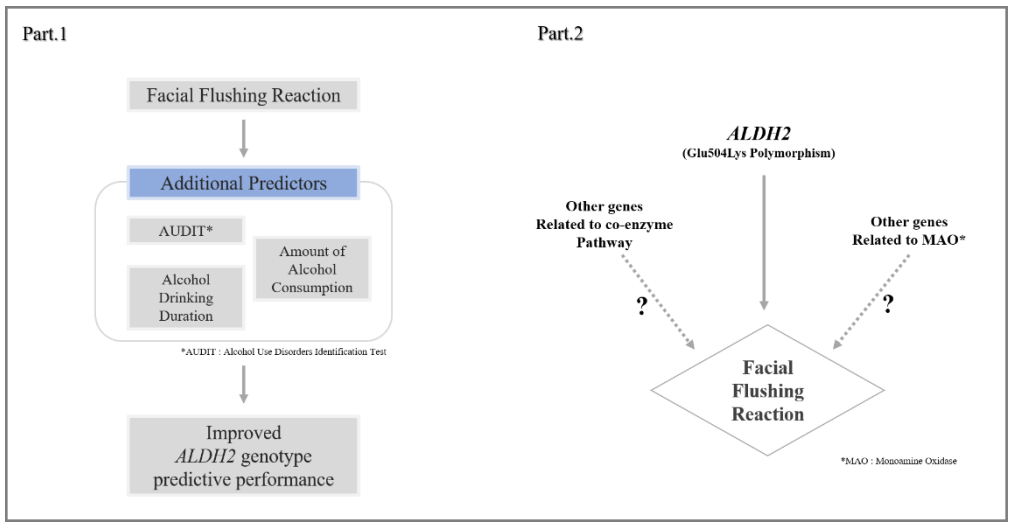
At present genotyping *ALDH2* is the most accurate way to identify individual's *ALDH2* polymorphism. However, it is not possible to test *ALDH2* genotype for every individual and therefore facial flushing reaction, the physiological response of inactive *ALDH2*, serves as a predictor of inactive *ALDH2* for both genotype of *ALDH2* *1/*2 and *ALDH2**2/*2 with two questionnaires [14]. The facial flushing reaction questionnaire was originally developed and validated by Japanese group with the questionnaire's sensitivity and specificity for identifying inactive ALDH2

were 96.1 and 79.0%, respectively [14] while sensitivity and specificity from validation of facial flushing reaction questionnaire with Korean population was 95.1% and 76.5%, respectively [15]. This results from East Asian population allowed the utility of facial flushing reaction questionnaire as a screening tool for *ALDH2* genotype in clinical practice as well as large scale studies to investigate the association of *ALDH2* genotype and other diseases or cancers [16]. However, from the study of comparison between facial flushing reaction questionnaire and *ALDH2* polymorphism for risk of upper aerodigestive tract cancer in a Japanese population showed misclassification between facial flushing and *ALDH2* genotype that facial flushing had no significant association with upper aerodigestive tract cancer, although inactive *ALDH2* showed significant association with upper aerodigestive tract cancer [17]. The discrepancy of the results was assumed to be due to misclassification of facial flushing and *ALDH2* polymorphism which was particularly obvious among heavy drinkers with inactive *ALDH2*.

2. Objectives

This study aims to provide evidence for the utility of more precise screening tool of *ALDH2* polymorphism without genotyping as illustrated in Figure 1. Part.1 of this study develops and validates the prediction model for *ALDH2* genotype with facial flushing reaction and additional alcohol related phenotype data such as absolute amount of alcohol consumption(g/week) and Alcohol Use Disorders Identification Test (AUDIT) score, and Part.2 of the study explores genes affecting facial flushing reaction with genotype data and measure polygenic score for facial flushing reaction.

Figure 1. Diagram of objectives of Part 1 and Part 2



II. METHODS

1. Study participants

The Healthy Twin Study is a cohort study of Korean twin-family including adult monozygotic and dizygotic twin pairs and their first-degree family members. Detailed information of The Healthy Twin Study was previously described [18]. Total 3500 participants were recruited since 2005 and went through general health examinations and completed a comprehensive questionnaires survey. Among total 3500 participants, lifetime abstainer and individuals who had not completed alcohol drinking related questionnaire including facial flushing reaction were excluded.

2. Data measurement

2.1 Alcohol Use Disorders Identification Test (AUDIT)

Participants were asked to complete 10 questions from Alcohol Use Disorders Identification Test (AUDIT) regarding (1) alcohol consumption (2) drinking behaviour and dependence and (3) the consequences or problems related to drinking. Questions from 1 to 3 were for screening hazardous drinking, questions from 4 to 6 were for alcohol dependence and questions 7 to 10 for harmful use. Each question was scored on three or five-point scale and added up for total AUDIT score [19].

2.2 Average amount of alcohol Intake

Average amount of pure ethanol was calculated based on the average frequency of drinking occasions and the amount of alcohol consumed for an occasion by different alcoholic beverages in Korea and denoted as gram per week (g/week).

2.3 The Facial flushing reaction questionnaire

The Facial flushing reaction questionnaire was developed and validated from Japanese studies [14] and participants completed Korean version of facial flushing reaction questionnaire. The questionnaire included two questions (i) Do you flush in the face immediately after drinking as little as a glass of beer?: “never”, “sometimes”, “often” , “always”, “unknown”; Current flushing (ii) Did you flush in the face immediately after drinking as little as a glass of beer during the first to second year you started drinking?: yes, no, or unknown; Former flushing. Participants were classified as “never” if answered “no” to question (i) and “unknown” to question (ii) or “unknown” for both questions. Current flushing status was applied to classify participants who answered “sometimes”, “often” or “always” to question (i).The remaining participants were classified “never”.

Active *ALDH2*(*1/*1) carriers who answered “sometimes”, “often” or “always” to question (i) and participants with inactive *ALDH2*(*1/*2 or *2/*2) who never experienced facial flushing reaction were defined as a facial flushing reaction mismatch case.

2.3 Genotype data

DNA was extracted from blood samples collected during health examination and genotyped using Affymetrix Genome-wide Human SNP array 6.0. The genotype data was imputed by using 1000 Genome haplotypes phase I integrated variant set release GRCh37/hg19 in Asian population and The Korean Reference Genome.

3. Statistical analysis

3.1 Descriptive analysis

Descriptive analysis was performed by different sex and *ALDH2* genotype-Facial Flushing Reaction matching and mismatching group with R software version 3.4.3 for general demographic features and alcohol related questionnaire. Mean ± standard deviation and counts with proportion were measured for continuous variables and categorical variables, respectively. The sensitivity and specificity of facial flushing reaction questionnaire was measured as follow: (i) % of inactive *ALDH2*(*1/*2 or *2/*2) carrier with facial flushing response for sensitivity (ii) % of active *ALDH2*(*1/*1) with never experienced facial flushing response for specificity.

3.2 Development and Validation of Prediction model

To generate *ALDH2* polymorphism prediction model study participants were randomly divided into training dataset (2/3 of study participants) and validation dataset (1/3 of study participants). To select appropriate predictors, variable

screening was firstly done to find how the variables are correlated with *ALDH2* and how variables are correlated with each other to check multicollinearity. Backward stepwise regression was performed including all possible variables related alcohol questionnaires including facial flushing reaction. The response variable *ALDH2* was set as 'event' for inactive *ALDH2* (*1/*2 or *2/*2) and 'no event' for active *ALDH2* (*1/*1) and considered as a binary outcome for logistic regression. Regression coefficient from the explanatory model was used to develop prediction model which brings the probability of event. To validate the performance of prediction model, several evaluation aspects were measured including overall performance, Hosmer-Lemeshow test, Integrated Discrimination Improvement (IDI) and C statistics (=AUC) with ROC curves [20,21]. Next the evaluation aspects of the prediction model were compared with predictive performance of facial flushing reaction questionnaire.

3.3 Polygenic score for Facial flushing reaction

Polygenic score was calculated as the sum of weighted risk alleles. Genetic loci associated with facial flushing reaction were explored by using Genome-wide association study which was done by using family-based score test for association (FASTA) from 'GenABEL' package in R software. However, since *ALDH2* exerts strong effect on facial flushing reaction, rs671 of *ALDH2* was adjusted for conditional analysis GWAS. Resulted SNPs from rs671 of *ALDH2* conditional analysis GWAS were used to construct polygenic score with polygenic scoring software written in R, called PRSice [22].

III. RESULTS – Part.1

1. Descriptive statistical analysis

1.1 Study participants by sex

Among 3500 participants from The Healthy Twin Study, 893 men and 914 women who satisfied inclusion criteria were involved in the study. Baseline characteristics, AUDIT and other alcohol related characteristics are shown in Table 1.1. The average score for all three domains were significantly higher in men than women and the total score of AUDIT was almost twice higher in men than women. The proportion of each category of facial flushing reaction was relatively similar in both sexes which implies that biological mechanism related to facial flushing reaction or metabolic ability dealing with alcohol intake does not differ by sex while the average amount of alcohol consumption was three times higher in men than women. The most prevalent range of drinking duration was 11-20 year for both men and women and 22% of men were carrying inactive *ALDH2* while 16% of women were inactive *ALDH2* carriers.

1.2 Study participants by matching and mismatching case

Among 1807 study participants 19.4% of individuals were inactive *ALDH2* carriers which was slightly higher % than previously reported in international study

which was around 17% in Korean population [i] but less than studies conducted in Korea which was about 25-35% [ii] as shown in Table 1.2. In active *ALDH2* carrier group, 25% of individuals were belonging to mismatch group. The average score for Hazard category of AUDIT only showed significantly lower score in mismatch group than matching group. However, the average amount of alcohol consumption by mismatch group was not significantly less than in matching group. Also, the proportion of alcohol drinking duration was not different in active *ALDH2* matching and mismatch group.

In heterozygote-inactive *ALDH2* carrier group (*ALDH2*$^{(*1/*2)}$), 5.3% of individuals showed mismatched facial flushing reaction from their *ALDH2* genotype. The average AUDIT scores between matching group and mismatching group were significantly different that scores from all three domains of AUDIT and the average total AUDIT score were significantly higher in mismatching group. Although there were few individuals who seemed to drink alcohol extensively in inactive *ALDH2* mismatch group, it was obvious that inactive *ALDH2* mismatching group drink far more than that of inactive *ALDH2* matching group. However, the most prevalent drinking duration range was the same as all other groups.

There were only 8 homozygote-inactive *ALDH2* carriers (*ALDH2*$^{(*1/*1)}$) and they all expressed facial flushing when drinking alcohol and hence no mismatching individual. As expected, they consume obviously less amount of alcohol than any other groups in the study dataset.

Table.1 General Characteristics of study participants

	Men (n=893)	Women (n=914)	*P*
Age	52.9 ± 14.1	50.3 ± 12.0	< 0.001
Current smoker	402 (45.1)	70 (7.7)	< 0.001
BMI	24.6 ± 3.0	22.7 ± 2.2	< 0.001
AUDIT			
Hazard	6.7 ± 2.8	3.8 ± 2.2	< 0.001
Dependence	1.9 ± 1.7	1.0 ± 1.1	< 0.001
Harm	3.0 ± 3.0	1.3 ± 1.9	< 0.001
Total	11.6 ± 6.6	6.1 ± 4.6	< 0.001
Facial flushing reaction			
Never	592 (66.3)	618 (67.6)	0.232
Sometimes	107 (12.0)	125(13.7)	
Often	47 (5.3)	53 (5.8)	
Always	147 (16.5)	118 (13.0)	
Drinking duration			
1-5 year	80 (9.0)	151 (16.5)	< 0.001
6-10 year	105 (11.8)	270 (29.5)	
11-20 year	298 (33.4)	351 (38.4)	
21-30 year	196 (22.0)	108 (11.8)	
31-40 year	139 (15.6)	21 (2.3)	
41 year +	75 (8.4)	13 (1.4)	
Alcohol consumption(g/week)	171.6 ± 214.0	47.18 ± 76.0	< 0.001
Inactive *ALDH2* (*1/*2 or *2/*2)	200 (22.4)	150 (16.4)	0.003

*Note: AUDIT, Alcohol Use Disorder Identification Test;
Continuous data in mean ± standard deviation and p-values were calculated on t-test.
Categorical data in counts and tested and p-values were calculated on Pearson's Chi-square test

Table.2 General Characteristics of study participants by *ALDH2* genotype and facial flushing reaction

	Active *ALDH2*(*1/*1) (n=1457)			Inactive *ALDH2*(*1/*2) (n=342)			Inactive *ALDH2* (*2/*2) (n=8)
Facial flushing reaction	Negative	Positive		Positive	Negative		Positive
	Match(n=1093)	Mismatch(n=364)	*P*	Match(n=324)	Mismatch(n=18)	*P*	Match(n=8)
sex (male)	544 (49.8)	149 (40.9)	0.066	183 (60.0)	11 (61.1)	0.917	6(80.0)
Age	51.1 ± 13.0	54.8 ± 13.5	<0.001	50.0 ± 12.4	47.7 ± 12.4	0.505	39 ±13.6
Current smoker	288 (26.3)	87 (24.0)	0.987	86 (26.0)	11 (64.7)	0.002	3 (37.5)
BMI	23.7 ± 3.1	23.8 ± 3.3	0.485	23.4 ± 3.1	22.7 ± 3.2	0.338	23.8 ± 4.84
AUDIT							
Hazard	5.7 ± 2.9	5.0 ± 2.8	<0.001	3.6 ± 2.3	6.2 ± 2.8	<0.001	2.17 ± 1.93
Dependence	1.6 ± 1.5	1.7 ± 1.7	0.187	0.7 ± 1.0	1.7 ± 1.6	0.025	0.375 ± 0.58
Harm	2.5 ± 2.7	2.4 ± 2.7	0.914	0.8 ± 1.5	2.6 ± 3.0	0.022	0.25 ± 0.46
Total	9.8 ± 6.3	9.2 ± 6.6	0.191	5.2 ± 4.0	11.1 ± 6.9	0.002	2.79 ± 2.69
Facial flushing reaction							
Never	1093 (75.0)	-	<0.001	-	18 (5.1)	<0.001	-
Sometimes	-	285 (19.6)		42 (12.0)	-		-
Often	-	35 (2.4)		65 (18.6)	-		1 (12.5)
Always	-	44 (3.0)		225 (64.3)	-		7 (87.5)
Drinking duration							
1-5 year	124 (11.3)	40 (11.0)	0.735	60 (18.5)	2 (11.1)	0.691	4 (50.0)
6-10 year	218 (20.0)	72 (19.8)		57 (17.6)	2(11.1)		1 (12.5)
11-20 year	405 (37.1)	129 (35.4)		108 (33.3)	8 (44.5)		1 (12.5)
21-30 year	185 (16.9)	69 (19.0)		63 (19.4)	5 (27.8)		1 (12.5)
31-40 year	108 (9.9)	28 (7.7)		24 (7.4)	1 (5.6)		1 (12.5)
41 year +	53 (4.8)	26 (7.1)		12 (3.7)	0 (0)		
Alcohol consumption(g/week)	126.8 ± 177.0	105.1 ± 191.6	0.057	47.3 ± 83.4	229.9 ± 291.9	0.016	14.2 ± 18.0

*Note: AUDIT, Alcohol Use Disorder Identification Test;
Continuous data in mean ± standard deviation and p-values were calculated on t-test.
Categorical data in counts and tested and p-values were calculated on Pearson's Chi-square test

2. Prediction of *ALDH2* polymorphism

2.1 Development of prediction model

Table 3. Sensitivity and Specificity of Facial Flushing Questionnaire in Korean Healthy Twin Study

Facial Flushing Reaction Questionnaire				
ALDH2 genotype	Negative (n=1111)	Positive (n=696)	Sensitivity (%)	Specificity (%)
Active(n=1457) (*1/*1)	1093	364	94.9	75.0
Inactive(n=350) (*1/*2 or *2/*2)	18	332		

*Note: Sensitivity = % of Inactive *ALDH2* with flushing response, Specificity = % of active *ALDH2* with never flushing

Before developing a new prediction model, the sensitivity and specificity were measured with study dataset. As shown in Table 3, among 350 inactive *ALDH2* carriers, 332 responded as positive for facial flushing reaction and 1093 active *ALDH2* carriers answered as negative for facial flushing reaction out of 1457 active *ALDH2* carriers. Therefore, sensitivity and specificity for detecting *ALDH2* with facial flushing reaction questionnaire were 94.9% and 75.0% for sensitivity and specificity respectively.

In developing step, 1024 randomly selected individuals were included in training dataset. By conducting backward stepwise logistic regression, significant predictors were selected which were obviously facial flushing reaction questionnaire, amount of alcohol consumption, hazard domain (AUDIT questions of 1,2 and 3) and AUDIT

question No.8 asking “How often during the last year have you been unable to remember what happened the night before because you had been drinking?”.

As shown in Table 4-1 and Table 4-2, regression coefficients of each explanatory model were obtained to be used to construct the prediction model of *ALDH2* genotype.

Table 4-1. Explanatory models

Model.1	Inactive *ALDH2* (Y/N) =	$\beta_0 + \beta_1$*Facial flushing reaction
Model.2	Inactive *ALDH2* (Y/N) =	$\beta_0 + \beta_1$*Facial flushing reaction + β_2*Amount of alcohol consumption
Model.3	Inactive *ALDH2* (Y/N) =	$\beta_0 + \beta_1$*Facial flushing reaction + β_2*Hazard of AUDIT (Q1-3)
Model.4	Inactive *ALDH2* (Y/N) =	$\beta_0 + \beta_1$*Facial flushing reaction + β_2*Amount of alcohol consumption + β_3*AUDIT question no.8
Model.5	Inactive *ALDH2* (Y/N) =	$\beta_0 + \beta_1$*Facial flushing reaction + β_2*Hazard of AUDIT (Q1-3) + β_3*AUDIT question no.8

Table 4-2. Coefficient of Explanatory models

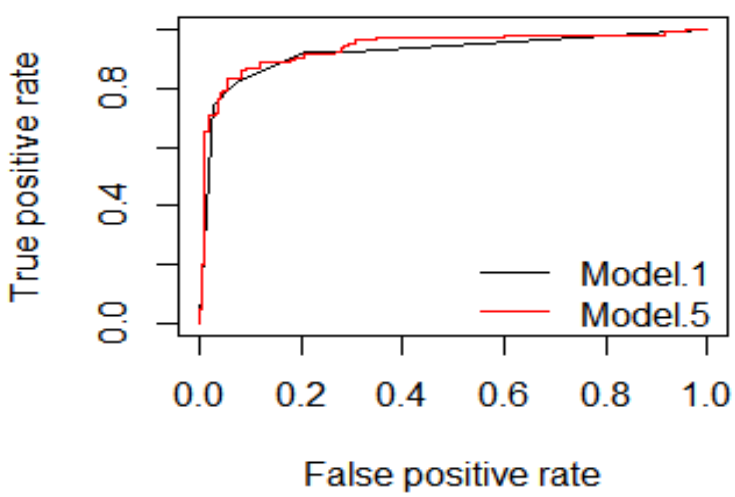
	Model.1	Model.2	Model.3	Model.4	Model.5	
Intercept	-5.9977	-5.790951	-5.71113	-5.580218	-6.07425	
Facial Flushing Reaction	1.9934	1.961977	1.96335	1.96652	2.01498	
Amount of Alcohol	-	-0.001555	-	0.000961	-	$\hat{p} = \frac{e^{\beta 0+\beta 1 x1}}{1+e^{\beta 0+\beta 1 x1}}$
Hazard of AUDIT (Question1-3)	-	-	-0.04451	-	0.11338	
AUDIT question No.8	-	-	-	-1.119428	-1.3441	

2.2 Validation of prediction model

To validate and select the final model with the best predictive performance, 5 candidate models were compared by measuring different aspects of model performance as shown in Table 5. The overall performance was measure by PseudoR^2. The overall performance of prediction model with facial flushing reaction questionnaire (Model.1) showed PsudoR^2 of 0.569 and model 2,3 also had almost

the same value of PsudoR^2 as Model.1. The highest PsudoR^2 was observed from Model.5 as shown in Table 5. Aspect of calibration was measure with Hosmer-Lemeshow statistics which examines whether the observed proportion of events are similar to the predicted probabilities of occurrences. the *p*-value of Hosmer-Lemeshow statistics from all five models were not significant which support the null hypothesis that model fits the data. Discrimination ability of each model was compared by measuring AUC from ROC curve. All 5 candidate models showed excellent discriminating ability by having above 0.9 of AUC. However, Model.5 had the highest AUC among 5 candidate models which was 0.930 while AUC of Model.1 was 0.917. As displayed in Figure 2, the ROC curve of Model.5(red line) is slightly more on left hand side than Model.1. Also, sensitivity and specificity from model 1 and model 5 was compared with their optimal cutoff from ROC curve. The sensitivity and specificity from model 1 were 0.83 and 0.90 respectively while sensitivity and specificity obtained from model 5 were 0.87 and 0.88 respectively.

Figure 2. ROC Curves of Model.1 and Model.5



To test for improvement of predicting ALDH2 polymorphism by adding more predictors traditional Net Reclassification Improvement (NRI) and an alternative measure, Integrated Discrimination Improvement (IDI) were both measured. Overall 21.9% of study participants were reclassified, but this difference was not statistically significant. However, IDI was 0.019 (p<1e-05).

Figure 3. Discrimination Slope of Model.1 and Model.5

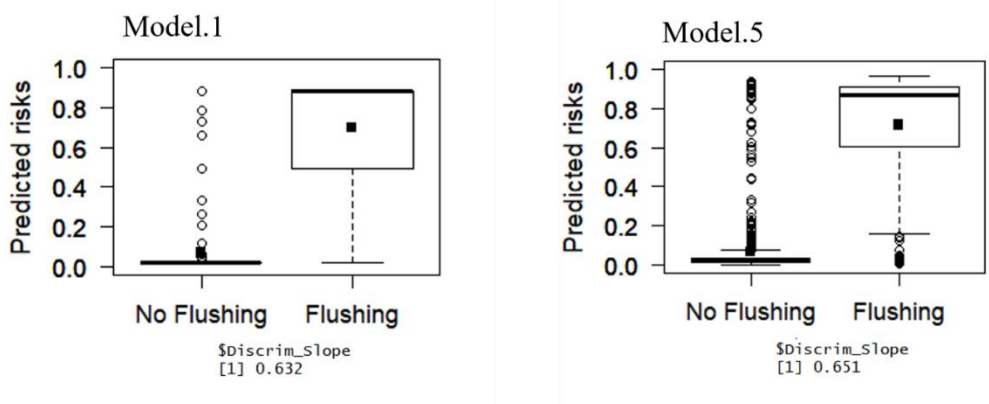


Table 5. Comparison of Models' predictive performance		Overall-performance	Calibration	Discrimination	Reclassification
	Model	PseudoR2	Hoslem GOF	AUC	Discrim-Slope
Model.1	Y~ Facial flushing reaction only	0.583	0.997	0.917	0.632
Model.2	Y~ Facial flushing reaction + Amount of alcohol consumption	0.584	0.064	0.927	0.635
Model.3	Y~ Facial flushing reaction + Hazard of AUDIT (Question 1-3)	0.583	0.469	0.921	0.633
Model.4	Y~ Facial flushing reaction + Amount of alcohol consumption + AUDIT question no.8	0.601	0.138	0.931	0.649
Model.5	Y~ Facial flushing reaction + Hazard of AUDIT only(Question1-3) + AUDIT question no.8	0.604	0.151	0.930	0.651

*Note: *P*-value of Hosmer-Lemeshow Goodness of Fit (Hoslem GOF) , Discrimination Slope from IDI (Discrim-Slope)

III. RESULTS – Part.2

1. *ALDH2* (rs671) Conditional Genome-wide association study of Facial Flushing Reaction

Genome-wide association study of for Facial Flushing Reaction without conditioning *ALDH2*(rs671) resulted strong association signal in chromosome 12, where *ALDH2* is located, with p-value of 10^{-74} as shown in Manhattan plot (Figure 4) and Q-Q plot (Figure 5). However, since *ALDH2* exerts strong mediating effect on facial flushing reaction, conditional GWAS was conducted to explore SNPs affecting facial flushing reaction other than rs671 of *ALDH2*. Results from rs671 conditional GWAS for facial flushing reaction is shown in Manhattan plot and Q-Q plot in figure 6 and Figure 7. As shown in the figures, there was no SNPs with genome-wide significance level (p=10^{-08}) after adjusting for *ALDH2*(rs671). However, 117 SNPs with suggestive significant level of association(p=10^{-05}) were detected. The base positions of those SNPs were close to each other and displayed almost the same effect and *p*-value suggesting that clumping was required to sort out the representative SNP based on empirical estimates of linkage disequilibrium between correlated SNPs. After clumping of SNPs, 17 SNPs were left to be used in polygenic scoring as shown in Table 5. Top SNP from rs671 conditional GWAS for facial flushing reaction was rs6480460 (*p*-value=2.28e-06) and the regional plot with this SNP is displayed in Figure 8.

Figure 4. Manhattan Plot of SNPs from GWAS for facial flushing reaction

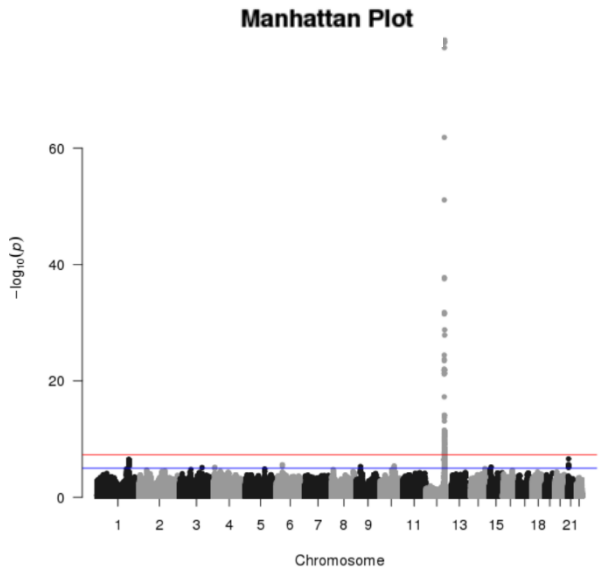


Figure 5. Q-Q Plot of GWAS p-value for facial flushing reaction

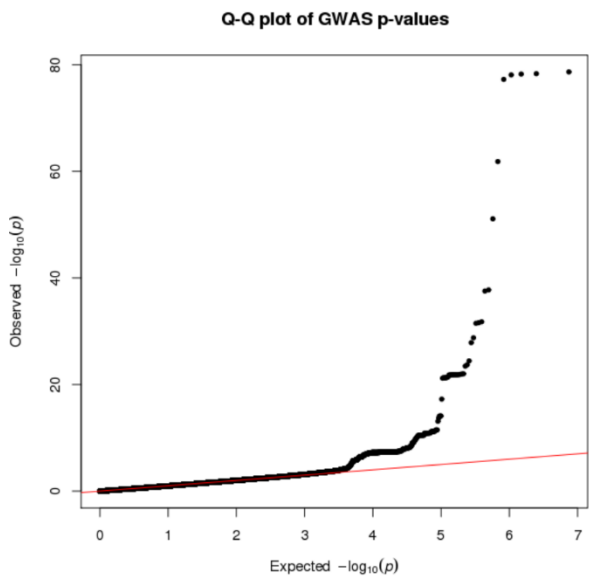


Figure 6. Manhattan Plot of SNPs from *ALDH2*(rs671) conditional GWAS for facial flushing reaction

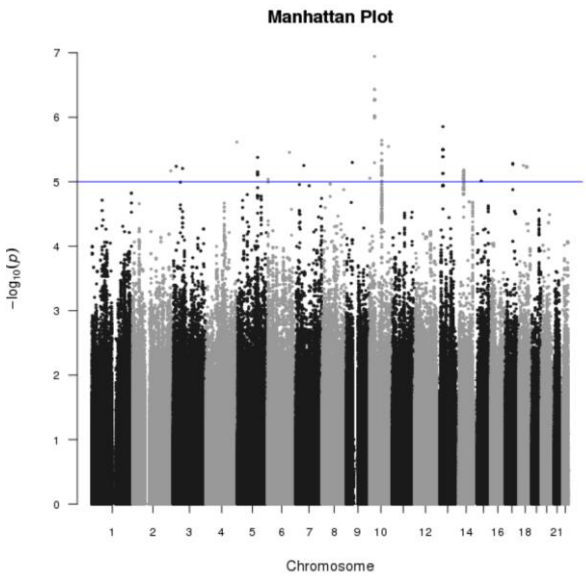


Figure 7. Q-Q Plot of *ALDH2*(rs671) conditional GWAS p-value for facial flushing reaction

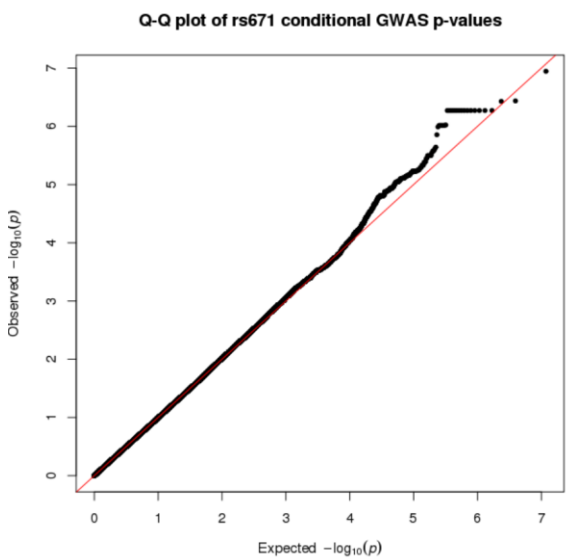


Figure 8. Regional Plot of *ALDH2*(rs671) conditional GWAS for facial flushing reaction with rs6480460

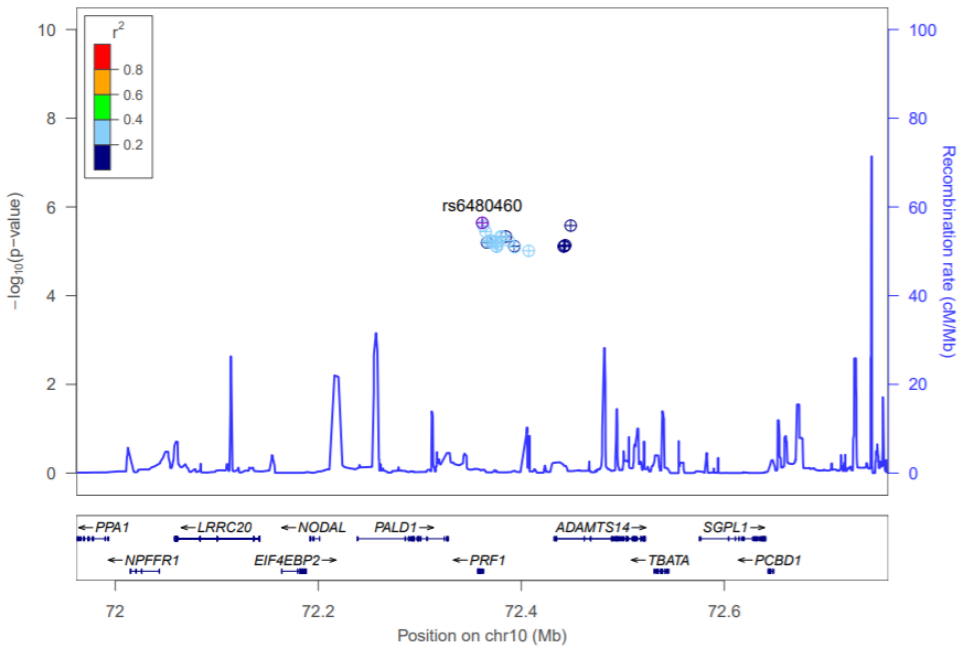


Table.6 Characteristics of Top 10 SNPs from *ALDH2*(rs671) conditional GWAS for facial flushing reaction used to generate the RPS

Chromosome	SNP	Location	Closest gene	A1	*P* *	Effect ± SE
10	rs6480460	72361596	PRF1	A	2.28E-06	0.115 ± 0.01
4	rs118059800	186770441	SORBS2	A	2.43E-06	0.573 ± 0.06
6	rs72987679	132944051	TAAR2	A	3.50E-06	0.194 ± 0.02
13	rs10507455	38098734	Unknown	G	4.10E-06	0.348 ± 0.04
9	rs189869942	35761900	NPR2	T	5.03E-06	0.355 ± 0.04
10	rs2001681	29287771	C10orf126	T	5.08E-06	0.147 ± 0.02
17	rs11570508	45228560	CDC27	A	5.38E-06	0.248 ± 0.03
7	rs145165905	48330903	ABCA13	C	5.60E-06	0.556 ± 0.06
18	rs150421202	47543922	MYO5B	G	5.91E-06	0.647 ± 0.07
3	rs983739	56823366	ARHGEF3	A	6.23E-06	0.141 ± 0.02

**P*-value was corrected for genomic inflation

2. Polygenic score for facial flushing reaction

As shown in Table 6, the effect of each risk allele was obtained from GWAS discovery and polygenic score was calculated as the sum of those weighted risk alleles with the threshold of suggestive significance (*p*-value < 1e-05) for each individual with their genotype data. The top SNP, rs6480460, was located in gene called PRF1 in chromosome 10.

The average polygenic score of 4 different groups by *ALDH2* genotype and facial flushing reaction is displayed in Table.7. In active *ALDH2* group, the average polygenic score for mismatching group, active *ALDH2* carriers with facial flushing reaction, was 4.76 which was significantly lower than that of matching group. In contrast, inactive *ALDH2* carriers without facial flushing reaction had a higher average polygenic score than inactive *ALDH2* carriers with facial flushing reaction.

Table.7 Average polygenic score by different groups

ALDH2	Active *ALDH2*	Active *ALDH2*			Inactive *ALDH2*	Inactive *ALDH2*		
Facial flushing Reaction	Positive	Negative	95% CI	*P*	Positive	Negative	95% CI	*P*
Average PRS	4.76	4.97	0.17 - 0.26	<0.001	4.92	5.08	-0.25 - - 0.07	0.001

VI. DISCUSSION

This study conducted a new prediction modelling for *ALDH2* polymorphism, which had been screened by facial flushing reaction questionnaire [14], as because individuals with genotype (*ALDH2*)-phenotype (facial flushing reaction) discordance were reported. The frequency of facial flushing reaction between male and female was not differ which implies that the biological mechanism of facial flushing reaction does not act differently by sex even though the amount of alcohol consumption was much greater in men than women. Also, regardless of the *ALDH2* genotype, it was found that the amount of alcohol consumption or Hazard domain of AUDIT, which is related to alcohol consumption, were higher in the group of positive facial flushing reaction. This suggests that applying alcohol consumption related variables as an additional predictor for screening *ALDH2* polymorphism would perform better when discriminating between active and inactive *ALDH2*. In this study, 5 explanatory models were generated with significant variables: Facial flushing reaction, Amount of Alcohol consumption, Hazard domain of AUDIT, AUDIT question No.8. Regression coefficients from explanatory models were used to generate prediction model and the final model with the best predictive performance was Model 5 which had a 0.13 higher AUC value and better reclassification than Model 1[20]. In terms of sensitivity and specificity of each model, the sensitivity of Model 5 was 0.04 higher than Model 1, but the specificity was reduced by 0.02 which implies that it is better to use Model 5 to discriminate

more inactive *ALDH2* carriers yet the predictive performance gets slightly weaker on active *ALDH2* side. In order to apply developed models in the actual prediction, the equation and coefficients on Table 4-2 can be used with the results from facial flushing reaction questionnaire and AUDIT questionnaire to get the probability of having inactive *ALDH2*.

In the second part of this study, *ALDH2* conditional genome-wide association study for facial flushing reaction was performed to find out any genes related to alcohol metabolism such as genes related to monoamine oxidase as its activity is known to be disturbed by alcohol consumption and hence triggers chronic alcohol consumption and alcohol dependence or other co-enzyme of metabolic pathway for alcohol intake [25]. In this conditional GWAS analysis, facial flushing reaction was regarded as an adapted phenotype of ability to metabolize alcohol as individuals who consume more alcohol tent to express negative facial flushing reaction. Even though there was no SNPs satisfying GWAS significance level, SNPs at suggestive significant level had a small effect on facial flushing reaction and polygenic effect of those SNPs were calculated. As found in the average polygenic score of each group, individuals with no facial flushing reaction had a stronger polygenic effect which means they genetically have greater metabolic ability for alcohol intake. There is no reported mechanism to explain *ALDH2* genotype mismatching with its physiological phenotype especially for those who carry inactive *ALDH2* and do not express facial flushing reaction. However, according to their response on facial flushing reaction, 5 out of 18 mismatching cases first had facial flushing reaction when they started drinking alcohol in the past but no longer express facial flushing

reaction at present. These answers may suggest that drinking alcohol may have affected genes related to facial flushing reaction or metabolic ability to breakdown alcohol through the epigenetic effect.

Facial flushing reaction questionnaire is still an excellent tool for ALDH2 polymorphism screening however *ALDH2* does not always correspond with Facial flushing reaction and therefore it is more efficient and precise to screen *ALDH2* polymorphism with additional simple alcohol related questionnaire without genotyping. Also, study dataset with Facial flushing questionnaire and AUDIT questionnaire can impute *ALDH2* genotype of study participants when conducting study related to *ALDH2*. In terms of efficiency for adding extra predictors in the model, the new model with additional predictors showed minimal improvement in AUC and therefore predicted risks generally showed minimal changes as well. However, new developed model still improved the predictive performance which better discriminated inactive ALDH2.However, the prediction model developed in this study has to be validated with other study with larger sample size to be used generally to determine the efficiency of adding new predictors when screening for *ALDH2* genotype.

V. REFERENCES

[1] Cederbaum, A. I. (2012). Alcohol metabolism. Clin Liver Dis, 16(4), 667-685.

[2] Boccia, S., Hashibe, M., Galli, P., De Feo, E., Asakage, T., Hashimoto, T., Boffetta, P. (2009). Aldehyde dehydrogenase 2 and head and neck cancer: a meta-analysis implementing a Mendelian randomization approach. Cancer Epidemiol Biomarkers Prev, 18(1), 248-254.

[3] Li, H., Borinskaya, S., Yoshimura, K., Kal'ina, N., Marusin, A., Stepanov, V. A., Qin, Z., Khaliq, S., Lee, M. Y., Yang, Y., Mohyuddin, A., Gurwitz, D., Mehdi, S. Q., Rogaev, E., Jin, L., Yankovsky, N. K., Kidd, J. R., … Kidd, K. K. (2009). Refined geographic distribution of the oriental ALDH2*504Lys (nee 487Lys) variant. Annals of human genetics, 73(Pt 3), 335-45.

[4] Clarke, T. K., Adams, M. J., Davies, G., Howard, D. M., Hall, L. S., Padmanabhan, S., McIntosh, A. M. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry, 22(10), 1376-1384.

[5] Lee, H., Kim, S. S., You, K. S., Park, W., Yang, J. H., Kim, M., & Hayman, L. L. (2014). Asian flushing: genetic and sociocultural factors of alcoholism among East asians. Gastroenterol Nurs, 37(5), 327-336.

[6] Li, D., Zhao, H., Gelernter, J. (2012) Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. Hum Genet; 131: 725–737

[7] Yokoyama T, Yokoyama A, Kato H, Tsujinaka T, Muto M, et al. (2003) Alcohol flushing, alcohol and aldehyde dehydrogenase genotypes, and risk for esophageal squamous cell carcinoma in Japanese men. Cancer Epidemiol Biomarkers Prev 12: 1227-1233

[8] Shin, C. M., Kim, N., Cho, S. I., Kim, J. S., Jung, H. C., & Song, I. S. (2011). Association between alcohol intake and risk for gastric cancer with regard to ALDH2 genotype in the Korean population. Int J Epidemiol, 40(4), 1047-1055.

[9] Chang, J. S., Hsiao, J. R., & Chen, C. H. (2017). ALDH2 polymorphism and alcohol-related cancers in Asians: a public health perspective. J Biomed Sci, 24(1), 19.

[10] Newman, I. M., Ding, L., Shell, D. F., & Lin, L. (2017). How Social Reactions to Alcohol-Related Facial Flushing Are Affected by Gender, Relationship, and Drinking Purposes: Implications for Education to Reduce Aerodigestive Cancer Risks. International journal of environmental research and public health, 14(6), 622.

[11] Cai, Q., Wu, J., Cai, Q., Chen, E. Z., & Jiang, Z. Y. (2015). Association between Glu504Lys polymorphism of ALDH2 gene and cancer risk: a meta-analysis. PloS one, 10(2), e0117173.

[12] Zhao, T., Wang, C., Shen, L., Gu, D., Xu, Z., Zhang, X., Chen, J. (2015). Clinical significance of ALDH2 rs671 polymorphism in esophageal cancer: evidence from 31 case-control studies. Onco Targets Ther, 8, 649-659.

[13] Zhao, Y., & Wang, C. (2015). Glu504Lys Single Nucleotide Polymorphism of Aldehyde Dehydrogenase 2 Gene and the Risk of Human Diseases. Biomed Res Int, 2015, 174050.

[14] Yokoyama, A., Muramatsu, T., Ohmori,. T, Kumagai, Y., Higuchi, S., Ishii H,. (1997). Reliability of a flushing questionnaire and the ethanol patch test in screening for inactive aldehyde Dehydrogenase-2 and alcohol-related cancer risk. Cancer Epidemiol Biomarkers Prev. 1997 Dec;6(12):1105-7.

[15] Shin, C. M., Kim, N., Cho, S. I., Sung, J., & Lee, H. J. (2018). Validation of Alcohol Flushing Questionnaires in Determining Inactive Aldehyde Dehydrogenase-2 and Its Clinical Implication in Alcohol-Related Diseases. Alcohol Clin Exp Res, 42(2), 387-396.

[16] Yokoyama, T., Yokoyama, A., Kato, H., Tsujinaka, T., Muto, M., et al. (2003) Alcohol flushing, alcohol and aldehyde dehydrogenase genotypes, and risk for esophageal squamous cell carcinoma in Japanese men. Cancer Epidemiol Biomarkers Prev 12: 1227-1233.

[17] Oze, I., Matsuo, K., Hosono, S., Ito, H., Kawase, T., Watanabe, M, et al. (2010) Comparison between self-reported facial flushing after alcohol consumption

and ALDH2 Glu504Lys polymorphism for risk of upper aerodigestive tract cancer in a Japanese population. Cancer Sci. 2010 Aug;101(8):1875-80.

[18] Sung, J., et al., Healthy Twin: a twin-family study of Korea--protocols and current status. Twin Res Hum Genet, 2006. 9(6): p. 844-8.

[19] World Health Organization. Global status report on AUDIT: The Alcohol Use Disorders Identification Test Guidelines for Use in Primary Care 2001

[20] Tzoulaki, I., Liberopoulos, G., Ioannidis, JP. (2009) Assessment of claims of improved prediction beyond the Framingham risk score. Journal of the American Medical Association. 2009; 302:2345–2352.

[21] Steyerberg, E.,Vickers, A., Cook, N., Gerds, T. (2010) Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010 January; 21(1): 128–138.

[22] Andries, T., Hilde, K., Sven, S., Florence, V., Emmanuel, C., Cynthia, M., Eske, Derks. (2018)A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27:e1608.

[23] Paik IH, Lee C, Lee SJ, Lee CU, Kim JJ, Jun TY, et al.(1988) Genetic frequencies of aldehyde dehydrogenase 2 in Korea patients with alcohol dependence. J Korean Neuropsychiatr Assoc. 1998;37:149–158.

[24] Paik K. (1995) The studies on the differences in personality trait by aldehyde dehydrogenase-I genotype variances in Korean male alcoholics. J Korean Neuropsychiatr Assoc. 1995;34:15–26.

[25] Zimatkin SM., Tsydik VF., Lelevich VV. (1997) Alcohol consumption alters monoamine oxidase activities in brain structures: relation to ethanol craving. Cas Lek Cesk. 1997 Nov 5;136(21):666-9

Ⅵ. 국문 초록

알데하이드 디하이드로즈네이즈 2 다형성 예측 모형: ALDH2 유전형-음주 후 안면 홍조 반응 불일치 케이스

서울대학교 보건대학원

보건학과 보건학전공

우은진

연구 배경: 알데하이드 디하이드로즈네이즈 2 (*ALDH2*)유전자는 알코올의 독성 대사산물인 아세트알데하이드를 분해하는 효소의 유전자로서, *ALDH2*유전형에 따라 개인의 알코올 분해능력의 차이를 보인다. *ALDH2*의 rs671 변이가 있는 경우 아세트 알데하이드는 분해되지 못하고 체내에 축적되며, 이에 따른 대표적인 생리적 현상으로 음주 후 안면 홍조 반응이 있다. 또한 *ALDH2* rs671 변이에 따라 음주 섭취와 연관된 질병이나 암에 대한 위험도가 달라질 수 있다는 결과들이 보고되었다. 따라서 *ALDH2* 유전형을 파악하고 그에 맞는 음주 행태를 취하는 것은 관련 질병을 예방하는데 도움이 될 수 있다. 현재 음주 후 안면 홍조 반응에 대한 설문을 통해 *ALDH2* 유전형을 예측할 수 있으나, *ALDH2* 유전형과 반대되는 안면 홍조 반응을 보이는 경우가 발견되었다.

연구 목적: 1) 음주 후 안면홍조 설문 문항과 더불어 음주관련 정보를 바탕으로 정확한 본인의 알코올 분해 능력을 평가할 수 있는 예측 모형을 개발하고 검증한다. 2) *ALDH2* 유전자 이외의 음주 후 안면홍조와 연관성이 있는 알코올 분해 능력에 영향을 끼치는 다른 유전자를 탐색해본다.

연구 방법: 한국 쌍둥이 가족 코호트에서 얻은 1807명의 음주 후 안면홍조 설문과 더불어 음주 관련 데이터를 포함한 *ALDH2* 유전형 예측 모형 개발을 시행하였다. AUDIT을 포함한 음주관련 변수 들 중에서 유의한 변수들로 로지스틱 예측 모형을 개발하고 예측 모형의 성능을 평가하였다. 또한 *ALDH2*를 공변량으로 포함하여 보정한 가족 기반 전장유전체 연관 분석을 통해 알코올 분해 능력에 영향을 끼치는 다른 유전자를 탐색하였다.

연구 결과: 음주 후 안면홍조 반응과 더불어 채택된 유의한 예측 변수는 다음과 같다 1) 알코올 섭취량 2) AUDIT 설문 중 Hazard domain 3) AUDIT 설문 문항 8. 유의한 예측 변수들을 이용하여 개발된 예측 모형 중 가장 뛰어난 예측 성능을 보인 모델은 음주 후 안면홍조 반응, AUDIT 설문 중 Hazard domain, AUDIT 설문 문항 8을 포함한 모델이며, 기존 음주 후 안면홍조 반응만을 포함한 예측모형보다 0.13높은 AUC값을 나타냈다. *ALDH2*(rs671)을 공변량으로 보정한 전장유전체 연관 분석에서는 음주 후 안면홍조 반응과 유의수준 $p=10^{-8}$에서 연관성을 보이는 단일 유전자 변이는 없었으나, 유의수준 $p=10^{-5}$를 기준으로 117개의 단일 유전자 변이들이 연관성을 보였다. 그 중 클럼핑을 통해 얻은 17개의 단일 유전자 변이들의 효과크기를 활용하여 개인의 다유전자성 점수를 계산하였다. 그 결과 음주 후 안면 홍조 반응을 보이지 않는 그룹에서 더 높은 다유전자성 점수가 나타났다.

결론: 음주 후 안면홍조 반응은 *ALDH2* 유전형 예측에 유용한 설문 문항

이다. 하지만 추가적인 음주 관련 변수를 예측변수에 포함시킴으로 예측 모형의 성능이 높아지는 것으로 나타났다. 본 연구에서 개발된 예측모형을 통해 유전자검사 없이 ALDH2 유전형을 파악하는데 도움이 될 수 있으나 본 연구의 표본 크기에 대한 한계점을 고려하여 다른 인구집단에서의 검증이 필요하다. 또한 음주 후 안면홍조 반응을 이용한 알코올 분해능 전장유전체 연관분석 결과는 *ALDH2* 외에 알코올 분해능력에 대한 다른 유전적 감수성의 차이가 있음을 시사한다.

주요어: 알데하이즈 디하이드로즈네이즈, 음주 후 안면 홍조, 예측 모형, 전장유전체 연관분석, 조건부 전장유전체 연관분석, 가족 기반 전장유전체 연관분석, 다유전자성 스코어

학번: 2017-27710