보건학석사학위논문

Genetic Influences on HPV infection status in Korean women; The Healthy Twin Study

한국인 여성에서 인유두종바이러스 감염여부의 유전적 영향; 가족-쌍둥이 코호트 연구

2016 년 2 월

서울대학교 보건대학원
보건학과 유전체역학전공
한 민 지

Genetic Influences on HPV infection status in Korean women; The Healthy Twin Study

한국인 여성에서 인유두종바이러스 감염여부의 유전적 영향; 가족-쌍둥이 코호트 연구

지도교수 성 주 헌

이 논문을 보건학 석사 학위논문으로 제출함
2015년 11 월

서울대학교 보건대학원
보건학과 유전체역학전공
한 민 지

한민지의 보건학 석사 학위논문을 인준함
2015년 12월

위 원 장 조 성 일 (인)
부위원장 원 성 호 (인)
위 원 성 주 헌 (인)

Abstract

Genetic Influences on HPV infection status in Korean women; The Healthy Twin Study

한국인 여성에서 인유두종바이러스 감염여부의 유전적 영향; 가족-쌍둥이 코호트 연구

Minji Han

Department of Epidemiology, School of Public Health

The Graduate School

Seoul National University

Human papilloma virus (HPV) infection is a sexually transmitted infection (STI) and is a well-established cause of uterine cervix cancer. Although previous studies have reported that host's genetic polymorphism is associated with HPV infection, it is not well understood whether host susceptibility is a true risk factor. This study aimed to assess the overall genetic contribution to HPV infection status, the first event in the natural course of carcinogenic infection, in a twin-family cohort in Korea.

Between 2006 and 2009, cervical smears were obtained from Papanicolaou (Pap) tests of 912 women (mean age 48; 142 monozygotic twin (MZ) pairs) from 260 families that were participants of the Healthy Twin Study. HPV infection was diagnosed using two different PCR amplifications of partial sequences of HPV.

To investigate the correlation between HPV infection status of HPV types and environment factors, we used Spearman's correlation analysis. The association with HPV infection status and several environment factors was analyzed using multiple regressions with mixed model. Genetic factor which has effects on HPV infection status by HPV types was analyzed by two methods; Intraclass Correlation Coefficients (ICC) and heritability. Heritability was calculated by variance component method.

According to results of this study, correlation between HPV infection status of HPV types and environment factors, which is an important indicator of HPV infection, was oral contraceptive use. Although several environment factors were not associated with HPV infection status, genetic factor has mild association with HPV infection status. Concordance to discordance ratio within pairs was highest among MZ (3 to 19). This could be confirmed by ICC analysis using tetrachoric correlation coefficient and heritability analysis. Genetic components were estimated by heritability that was 0.24-0.31 for

overall HPV infection, and when analyses were performed by the viral type; 0.51-0.54 for lower risk strains and not significant for high-risk strains.

Our findings suggest that HPV infection status is influenced by a host's genetic factors, viral genotype as well.

==========

Keywords: HPV; reproductive tract infection; genetics; heritability; genetic susceptibility

Student Number: 2010-22098

Contents

List of Tables

List of Figures

List of Abbreviations

HPV: human papillomavirus

HR HPV: high risk human papillomavirus

LR HPV: low risk human papillomavirus

CIN: cervical intraepithelial neoplasia (a histological classification)

LSIL: low-grade squamous intraepithelial lesions

HSIL: high-grade squamous intraepithelial lesions (the SIL classification is the cytological analogue of the CIN scheme)

HLA: human leukocyte antigen

STD: sexually transmitted disease

A: Additive genetic variance

C: Shared environmental variance

E: Non-shared environmental variance

I. Introduction

Oncogenic human papillomavirus (HPV) infections has been considered the single most important and necessary, although not sufficient, cause of cervical cancer as pathological understanding continued to develop the etiology of cervical cancer [3, 4]. Cervical cancer remains the fourth most common cancer of women worldwide, and it is the primary cancer of women in most of the developing countries, where more than 85% of cases occur [5]. In 2012, with an estimated 528,000 new cases representing 7.5% of all female cancer deaths and more than 260,000 deaths occurred. Of the deaths, approximately 87% occurred in Africa, Asia, Central and South America [5].

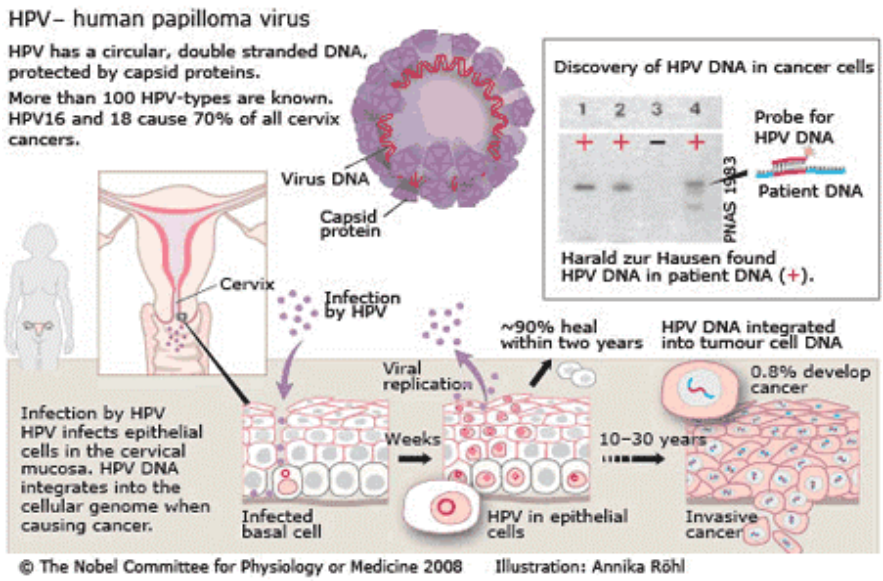

Figure 1. The human papilloma virus (HPV). Reproduc with permission from The Nobel Committee for Physiology or Medicine 200ed 8. (adapted from http://img.thebody.com/press/2008/nobel_hpv.gif. [1, 2])

HPV is a non-enveloped, double-stranded DNA virus (Figure 1.) that is mainly spread through sexual contact. There are over 170 HPV genotypes have been identified, currently [6]. Genital HPV types have been subdivided into low risk(LR) types (HPV 6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, and CP6108), which are found mainly in genital warts, and high risk(HR) types, which are known to infect the reproductive tract and be oncogenic or HR types (HPV 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53 56, 58, 59, 66, 68 and 82) [7].

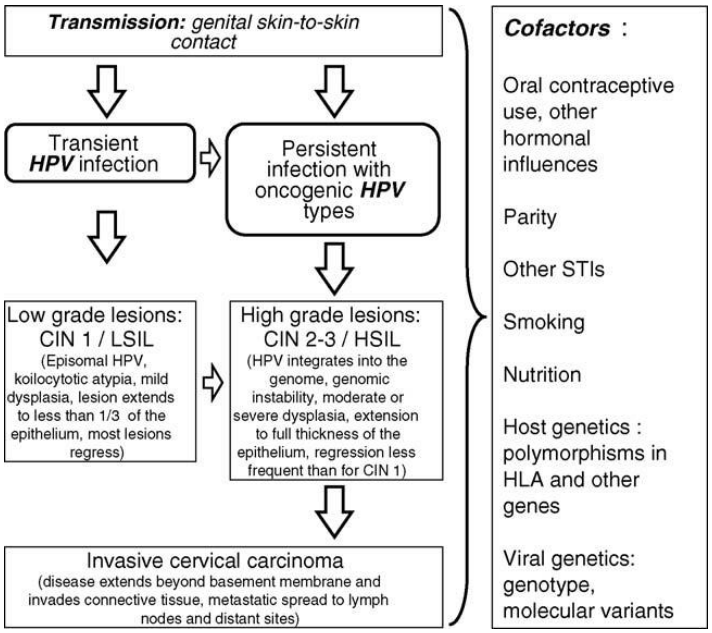

Figure 2. Etiologic model of human papillomavirus (HPV) infection as a necessary cause of cervical cancer. *Abbreviations:* CIN: cervical intraepithelial neoplasia (a histological classification); LSIL: low-grade squamous intraepithelial lesions; HSIL: high-grade squamous intraepithelial lesions (the SIL classification is the cytological analogue of the CIN scheme); HLA: human leukocyte antigen. (adapted from Trottier and Franco[8])

Figure 2 shows etiologic process from HPV infection to Invasive cervical carcinoma. Most HPV infections are transient, with the majority of new infections no longer detectable within 1 to 2 years [8]. Persistence of high risk (HR) HPV longer than 1 year strongly predicts pre-cancerous lesions of cervical cancer [9]. Once HPV enters the host, it develops an infection in the intraepithelial layer of the mucosa. Then, the cells can develop pre-cancerous properties that lead to cervical intraepithelial neoplasia (CIN) 1/2-3 or Low/high-grade squamous intraepithelial lesions (LSIL/HSIL) [2].

Figure 2 also shows cofactors of cervical cancer including the role of reproductive, lifestyle, host genetic factors, and viral genetic-factors. Current evidence for the role of host genetics in cervical cancer obtains from studies conducted large population registries in Scandinavian countries. This study included data of biological and adoptive mothers, full, half and adoptive sisters, and the Shared gene proportion (heritability) estimate for cervical tumors was 27% [10]. Another Swedish study using family data reported a heritability of 22% for invasive cervix cancer and 13% for in situ cervix cancer. A recent study with the Netherlands Twin Register reported heritability for cervix smear abnormalities [11].

The worldwide prevalence of vaginal HPV infection is estimated to average of 10.4% ranging from 8.1% in Europe to 22.1% in Africa, posing an

important public health problem [12]. HPV was responsible for 10.4% - 11.3% of infection-related cancer cases and 6% of deaths in Korea [13]. Other study reported the overall prevalence of HPV infection was as 10.4% in Korea [14].

Numerous risk factors for HPV infection that have been reported in various previous cross-sectional and prospective cohort studies; demographic, life style, social-economic status, reproductive history, Sexual activity and Genetics variants as shown in Table 1.

Table 1. Risk factors of HPV infection

Risk factors	Explanation	Reference
Demographic		
Age	Controversial ; High HPV prevalence included younger age HPV prevalence was highest in younger than 34 years, decreased in the 35–44 year-group and increased the older age-groups (45–54 years and more than 54 years) by world regions. Asia continued to decrease.	[12, 15]
Race	After adjusting number of sexual partners, only black is high HPV infection status.	[16]
Life style		
Alcohol	Increasing tendency of HPV infection	[14]
Smoking	Ever smoker had increasing tendency of HPV infection	[14, 15, 17]
social-economic status		
Education	High HPV prevalence included fewer years of education	[15]
Income	Lower income has been shown high HPV prevalence.	[15]

Table 1. *(continued)*

Risk factors	Explanation	Reference
reproductive history		
Marriage	HPV infection was more likely to be detected in those who had never married but were living with a partner versus those who were currently married. being divorced had increasing tendency of HPV infection	[16, 17]
Abortion	Related high HPV infection	[18]
menopausal status	No association with HPV infection	[15]
Sexual activity		
Age at first intercourse	Lower age at first intercourse increased the prevalence of HPV infection.	[15, 19]
Number of sexual partners	Increasing tendency of HPV infection with ≥ 3 sex partners within the past year, compared with females who had a single sex partner within the past year.	[12, 14-18]
Oral contraceptive use	Long duration of oral contraceptive use was strongly associated with HPV infection increase.	[15, 20]
Condom use	Use of condoms with regular partner was protective	[17]
hygiene habits after sexual intercourse	Significantly decreased the prevalence of HPV infection.	[20]
frequency of vaginal douching	Significantly decreased the prevalence of HPV infection.	[20]
husband's circumcision	Male circumcision is associated with a reduced risk of penile HPV infection.	[21]
having an affair of husband	Significantly increased the prevalence of HPV infection.	[20]
STD history	There is increasing tendency of HPV infection	[20]

Table 1. *(continued)*

Risk factors	Explanation	Reference
Genetics variants		
Gene	TYMS and EVPL, 2 gene regions were identified as significantly associated with type-specific HPV persistence.	[22]
Single Nucleotide Polymorphisms (SNPs)	Association between HLA-DRB1 polymorphism and HPV infection.	[23, 24]
	A SNP in the innate immune gene IRF3 (S427T) was associated with increased risk for HPV persistence.	[25, 26]
Pathway	Between HPV8 seropositivity and rs9357152, located within the major histocompatibility complex (MHC) II region at 6p21.32.	[27]

Host genetic factors are hypothesized to play a role for infection of HPV. To date, efforts in HPV infection related study have focused on understanding the role of HPV, but much remains unknown about the role of host genetic factors. Recent studies reported association between HLA DRB1 polymorphism and HPV infection [23, 24]. The innate immune genetic variants have been suggested association with HPV persistence in the population-based Guanacaste cohort in Costa Rica [25, 26].

The role of particular host genetic polymorphisms has been suggested. The relative overall importance of genetic and environmental influences, however, is not well understood for genital HPV infections. Heritability estimations can provide a clue on the role and relative importance of genetic susceptibilities in cervical HPV infections. Here, we report on genetic indices associated with HPV infection status in a Korean twin and family cohort.

II. Methods

1. Participants and Questionnaire

Study participants included 912 females from the Healthy Twin Study Korea, which is a twin family study, a part of the Korean Twin-Family Register and an underway cohort study of adult twin pairs and their family members who were willing to participate in the study through two major hospitals in Korea since 2005. The Healthy Twin recruits adult like-sex twins over the age of 30 and their family members, all of these 912 women were involved in analysis with 145 pairs of monozygotic (MZ) twins and 32 pairs of dizygotic (DZ) twins. The survey is in progress, including questionnaires, medical examinations and the collection of specimens. The questionnaires include demographics, behavioral factors with lifestyle (smoking, alcohol use, and physical activity), social-economic status (education, income) and reproductive history (marriage, childbirth, abortion, oral contraceptive use, menopausal status). This is a multicenter-based study. Every participant is provided with a physical examination at one of the three clinical centers located in different areas (Of 912 women, Samsung Medical Center, Seoul (n=618), Busan Paik Hospital, Busan (n=226), and Dankook University Hospital, Cheonan (n=68).). The protocol and methodological details are published previously [28].

We conducted additional questionnaire named “The Survey of Women’s Health for the Prevention of Cervix Cancer” to investigate risk factors of HPV infection and developing cervical cancer. The questionnaire was only asked for those who had the Pap smear test. The survey was conducted in a private room with only one participant by a staff who aware of the seriousness of dealing confidentiality of its participants. The questionnaire contained 12 questions, including 10 questions about sexual activities such as the method of contraception, condom use, hygiene habits after sexual intercourse, frequency of vaginal douching, husband’s circumcision, having an affair of husband, first age on set, the number of sexual partners and history of STD (sexually transmitted diseases).

All eligible women provided informed consent through a standardized consent form; any woman wanting to withdraw from study was able to do so at any time. Prior to the study, Ethics approval was obtained from the Institutional Review Board of Seoul Samsung Hospital, Busan Paik Hospital, and Seoul National University School of Public Health (no. 144-2011-07-11).

2. Sample collecting and HPV genotyping

Among the Healthy Twin Study, 912 females consented to undergo colposcopy and conduct a survey. Then 912 females who participated in this study have performed Papanicolaou (Pap) smear test. The samples of cervix fluid cell were acquired using endocervical brush (ThinPrep®, Hologic, USA and Surepath™, BD Diagnostics-TriPath, USA). SurePath samples were treated accordingly the process of the Prep Stain slide [29]. The used cytobrush was fixed in alcohol buffered solution at -70℃ until DNA analysis. The basic protocol was shared by the large Korean Genomic Cohort Study of adult individuals (KoGeS) by the National Genomic Research Institute of the Center for Disease Control, Korea (NGRI) [28].

The obtained cervical swab samples were prepared in commercial kit (Cyto-screen®, Roche) and tested for the presence of viral DNA. Viral DNA was extracted using the Chemagic Viral DNA/RNA kits (Chemagen®, Baesweiler). Amplification of partial sequences of HPV using two primer sets, GP5+/GP6+ and PGMY09/PGMY11, was described in previous studies [30]. Amplified PCR products were first identified by electrophoresis; the viral DNA-positive samples were further analyzed for nucleic acid sequencing of HPV. Sequence analysis was contracted out to a commercial sequencing

company (Cosmo Genetech, South Korea). The sequences were compared to those in the GenBank database using the NCBI BLAST search program, the genotypes of HPV were confirmed by phylogenetic analysis using MEGA 5.05 software (www.megasoftware.net).

3. Statistical Analysis

The environment factors contributing to HPV infection were selected on the basis of questionnaire results. Data preparation and statistical analyses were done with SAS (Version 9.2).

3.1 Intraclass correlation coefficient; Tetrachoric correlation coefficient

The ICC is a measure for absolute agreement within the familial relationship pairs before the phenotypic variance was decomposed into genetic and environmental components. The ICC of MZ twins provides a direct and unbiased estimate of heritability. Examination of ICC provides valuable insights regarding the relative importance of genetic and environmental factors. Because the phenotype, HPV infection status is dichotomous, we estimated intraclass correlation using a tetrachoric correlation matrix for HPV infection, which is supposed to be an appropriate method for dichotomous data[31]. Tetrachoric correlation coefficient is defined as:

$$r_t = \cos\left(\frac{180^\circ\sqrt{AD}}{\sqrt{AD}+\sqrt{BC}}\right)$$

In this study, ‘a’ is number of both HPV unaffected participants, ‘b’ and ‘c’ means HPV affected, and ‘d’ is number of both HPV infected.

Table 2. Tetrachoric correlation matrix in this study.

		Participant2 in intraclass pair		
	HPV	**-**	**+**	**Total**
Participant 1 in intraclass pair	**-**	a	b	a+b
	+	c	d	c+d
	Total	a+c	b+d	n

3.2 The variance component method

The overall genetic influence was estimated as heritability, the proportion of total variance attributable to genetic variance, which is assessed using the variance component method. Total phenotypic variance (Vp) can be explained by the sum of genetic (Vg) and environmental (Ve) components:

$$\mathbf{Vp = Vg + Ve}$$

Genetic variance is typically divided into additive (Va) and dominance (Vd) and sometimes epistasis variance, which means interactions among genes.

$$\mathbf{Vg = Va + Vd + epistasis\ variance}$$

Environmental variance is subdivided into common or shared (Vc) and unshared variance (Vue). Common or shared environmental variance is common influences to members of a family. Unshared environmental variance is unique to each individual and measurement error.

$$\mathbf{Ve = Vc + Vue}$$

In broad sense, heritability is the proportion of the phenotypic variance in a trait that is attributable to genetic effects.

$$\mathbf{Heritability = Vg/Vp}$$

However, most human family studies calculate narrow sense heritability, the proportion of the phenotypic variance in a trait that is attributable to the additive effects of genes, following equation:

$$\mathbf{Heritability\ (h^2) = Va/Vp}$$

For the HPV infection status, a liability function was generated of which threshold reflected the HPV infection rate among the study subjects, and heritability was estimated based on the liability function.

Being of a dualistic feature, HPV infection status has the postulation of a fundamental customary dispersal of burden. The amount of environmental and genetic effects is this burden, and its distribution has a boundary that ascertains between being infected and uninfected with HPV. From the predominance of HPV infection this boundary is approximated. Therefore, a person who is impacted by HPV infection state would transcend the boundary value for the distribution of the burden and that of the relative. [32]. On this boundary the covariates are modeled as effects. [33]

The statistical significance of heritability was assessed by means of a likelihood ratio test, in which the obtained likelihood of the model with the

stated additive genetic variance is compared with the likelihood of the model with the additive genetic variance constrained to zero. Heritability was estimated by the HPV viral types adjusting for environmental risk factors. To describe individual differences in liability to HPV infection, the genetic contribution to HPV infection was assessed by estimating the heritability of the HPV infection status using a pedigree-based likelihood approach as implemented using a variance component analysis in the SOLAR ("Sequential Oligogenic Linkage Analysis Routines") software package version 4 [32]. (http://txbiomed.org/departments/genetics)

We chose the best-fitting model centered on the maximum-likelihood estimation for the model-fitting analysis. There was a rather inconsequential variance chi-square between the two models. This is indicative of the bounds falling from the weaker model and that they were minimally different from zero. The difference between the numbers of parameters estimated was equivalent to that of the variance allocated as the degrees of freedom and a chi-square. The ratio in ambiguity explained by the fitted model was calculated by the Kullback-Leibler divergence R-squared equation. The assessment of the model's significance was done by relating with or without the covariates and the log-likelihood for models (maximize polygenic model and maximize sporadic model) [34].

III. Results

1. Study participants and the prevalence of HPV infection

The epidemiological and sociodemographic characteristics of the study population are shown in Table 3. Subjects were aged 25–79 years (mean age 48(S.D=11.50)). The overall prevalence of HPV was 7.9%. The prevalence of HPV infection was increased from 7.9% in participants aged <39 years to highest ratio 9.3%in those aged 40-49 years, and then decrease to 6.3% in in subjects aged 50-59 years, and then increase to 8.0% again aged >60 years. In LR, there was different tendency (Figure 3). About 50% of the women (n=464) were homemaker, >90% (n=823) had been never smoke and alcohol intake rate was approximately 50% (n=424). >95% of the women (n=867) had been married (728 were married status), 848 had given birth to a child, and 166 had used an oral contraceptive.

Related sexual activity factors from additional questionnaire were presented in Table 1 only sanitation after sex, possibility of husband affair, number of lifetime sexual partner, and history of STD. These factors were shown high Odds Ratio (OR) [20]. The prevalence of no sanitation after sex was 15.9% (n=7), more 4 of life time sexual partner was shown 16.7% prevalence (n=6).

Of the 912 participants, 72 (7.9 %) were found to have HPV infections.

Among those 72 cases, 44 (61.1%) were high risk HPV genotypes (HR, type 16, 18, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, and 66). 15 (20.8%) were low risk (LR, type 32, 42, 43, 54, and 70) and 13 (18.1%) were risk-undetermined strains. For subgroup analysis, HPV infection status was further classified as HR and LR genotypes, excluding risk-undetermined strains.

Table 3. Number of HPV infection cases by the viral strains according to characteristics of participants in the Healthy Twin Study

	Total	HPV infection status			
		Negative	Positive		
			Total* HPV	HR	LR
	(n=912)	(n=839)	(n=72)	(n=44)	(n=15)
	N	N	N, (%)	N, (%)	N, (%)
Age					
≤39	254	234	20 (7.9)	10 (4.0)	4 (1.6)
40 - 49	291	264	27 (9.3)	16 (6.0)	6 (2.1)
50 - 59	191	179	11 (6.3)	8 (4.2)	3 (1.6)
≥60	176	162	14 (8.0)	10 (5.7)	2 (1.1)
Education					
≤Elementary school	230	219	11 (4.8)	7(3.0)	2 (0.9)
High school	425	383	41 (9.9)	25 (5.9)	8 (1.9)
≥ College	257	237	20 (7.8)	12 (4.7)	5 (1.9)
Occupation					
Homemaker	464	390	30 (6.5)	21 (4.5)	6 (1.3)
Worker	448	449	42 (9.4)	23 (5.1)	9 (2.0)

***** Total HPV genotypes include HR types, LR types and undetermined risk types.
Abbreviation: HPV, Human papilloma virus; HR, high risk HPV types; LR, Low risk HPV types and risk; STD, sexually transmitted disease

Table 3. *(continued)*

	Total	HPV infection status			
		Negative	Positive		
			HPV total*	HR	LR
	(n=912)	(n=839)	(n=72)	(n=44)	(n=15)
	N	N	N, (%)	N, (%)	N, (%)
Smoke					
Never smoked	823	756	66 (8.1)	38 (4.62)	15 (1.8)
Ex-smoker	26	22	4 (15.4)	4 (15.4)	0
Current smoker	59	57	2 (3.4)	2 (3.4)	0
Missing	4	4	0	0	0
Alcohol					
Non-drinker	407	379	27 (6.9)	16 (1.8)	5 (1.2)
Ex-drinker	80	73	7 (8.8)	5 (6.3)	0
Current drinker	424	386	38 (9.0)	23 (5.5)	10 (2.4)
Missing	1	1	0	0	0
Marriage status					
Married	728	672	54 (7.6)	31 (4.3)	13 (1.8)
Single	45	40	5 (11.1)	4 (8.9)	1 (2.2)
Divorce, Separation, or cohabiting	139	126	13 (9.4)	9 (6.5)	1 (0.7)
Missing	1	1	0	0	0
Childbirth					
No	29	28	1 (3.5)	1 (3.5)	0
Yes	848	780	67 (8.0)	40 (4.8)	14 (1.7)
Missing	35	31	4 (11.4)	3 (8.6)	1 (2.9)

***** Total HPV genotypes include HR types, LR types and undetermined risk types.
Abbreviation: HPV, Human papilloma virus; HR, high risk HPV types; LR, Low risk HPV types and risk; STD, sexually transmitted disease

Table 3. *(continued)*

	Total	HPV infection status			
		Negative	Positive		
			HPV total*	HR	LR
	(n=912)	(n=839)	(n=72)	(n=44)	(n=15)
	N	N	N, (%)	N, (%)	N, (%)
Number of children					
1	107	97	10 (9.4)	5 (4.7)	2 (1.9)
2	416	386	30 (7.2)	18 (4.3)	7 (1.7)
≥3	328	300	27 (8.6)	17 (5.2)	5 (1.5)
Missing	61	56	5 (8.2)	4 (6.6)	1 (1.6)
Oral contraceptive					
No	733	683	50 (6.9)	34 (4.6)	9 (1.2)
Ex-user	157	140	16 (10.9)	7 (4.5)	5 (3.2)
Current user	9	5	4 (44.4)	2 (22.2)	1 (11.1)
Missing	13	11	2 (15.4)	1 (7.7)	1 (7.7)
Sanitation after sex					
Douche or shower	578	530	47 (8.3)	26 (5.0)	12 (2.1)
Paper or wet tissue	185	173	12 (6.5)	9 (4.9)	1 (0.5)
Not sanitation	44	37	7 (15.9)	5 (4.9)	1 (2.3)
Missing	105	99	6 (5.7)	4 (3.8)	1 (1.0)
Husband affair					
No	379	358	20 (5.3)	14 (3.7)	2 (0.5)
Probable	86	81	5 (5.8)	2 (2.3)	3 (3.5)
Yes	118	104	14 (11.9)	8 (6.8)	4 (3.4)
Have no idea	184	163	21 (11.4)	13 (7.1)	3 (1.6)
Missing	145	133	12 (8.3)	7 (4.8)	3 (2.1)

***** Total HPV genotypes include HR types, LR types and undetermined risk types.
Abbreviation: HPV, Human papilloma virus; HR, high risk HPV types; LR, Low risk HPV types and risk; STD, sexually transmitted disease

Table 3. *(continued)*

	Total	HPV infection status			
		Negative	Positive		
			HPV total*	HR	LR
	(n=912)	(n=839)	(n=72)	(n=44)	(n=15)
	N	N	N, (%)	N, (%)	N, (%)
No. of lifetime sexual partner					
0	2	2	0	0	0
1	664	615	48 (7.4)	30 (4.5)	11 (1.7)
2	100	92	8 (8.0)	5 (5.0)	1 (1.0)
3	50	46	4 (8.0)	2 (4.0)	1 (2.0)
≥4	36	20	6 (16.7)	2 (11.1)	2 (5.6)
Missing	60	54	6 (10.0)	3 (5.0)	0
History of STD					
No	665	616	48 (7.4)	30 (5.0)	9 (1.4)
Yes	45	38	7 (15.6)	2 (4.4)	5 (11.1)
Have no idea	22	19	3 (13.7)	3 (13.6)	0
Missing	180	166	12 (7.8)	9 (5.0)	1 (0.6)

***** Total HPV genotypes include HR types, LR types and undetermined risk types.
Abbreviation: HPV, Human papilloma virus; HR, high risk HPV types; LR, Low risk HPV types and risk; STD, sexually transmitted disease

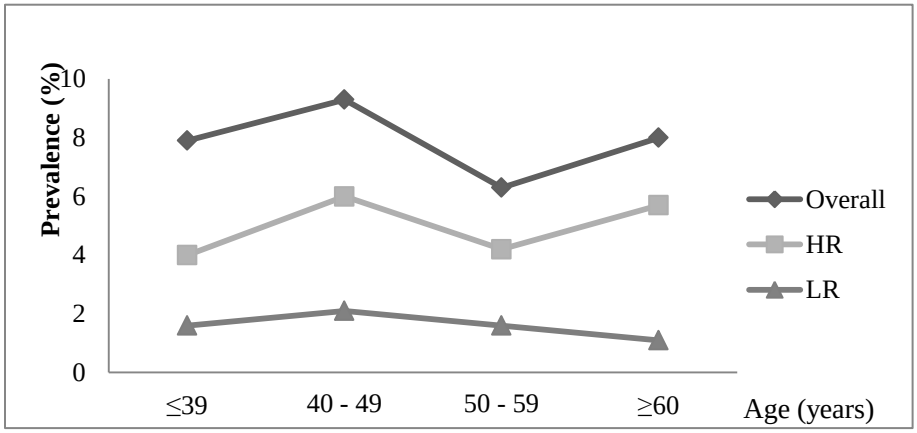

Figure 3. Age-specific prevalence of overall, HR and LR HPVs.

Abbreviation: Overall, Overall HPV types; HR, high risk HPV types; LR, Low risk HPV types

2. Familial aggregation of HPV infection

To understand family structure in participants, number of pedigrees of 912 female was calculated as shown Table 4.Total families were 260 and families with one or more female infected with overall HPV were 58. Families with HR HPV and LR HPV infection were 31 and 23 respectively. Notably, there were 164 families with one participant in this study.

Table 4. Data base of participants in the Healthy Twin Study

HPV type	No. pedigrees	No. only one participant in families	No. individuals
*Total	260	164	912
Overall HPV	58	14	72
HR	31	10	44
LR	23	4	28

* All participants; No., Number of;
HR, High risk genotype of HPV; LR, Low risk genotype of HPV

Table 5 shows concordance of HPV infection status by family relationships. Proportion of both-affected pairs was highest among MZ pairs. MZ twin pair was 142 pairs, 1st degree relative pair was 633 including 32 DZ twin pairs. 2nd and 3rd degree relative pairs were 22 and 3, respectively. No HPV infection in two groups. Both Pairwise concordance and tetrachoric correlation coefficient were highest among MZ pairs.

Table 5. Pair-wise concordance-discordance of overall HPV infection status

Relationship pair	Count	Unaffected	Discordant	Affected	*Pairwise concordance	¶Tetrachoric correlation (S.E)
MZ twin pair	142	120	19	3	0.14	0.38 (0.21)
1st degree relative	633	546	82	5	0.06	0.12 (0.13)
DZ twin pair	32	26	6	0	-	-
Sisters + DZ	388	339	46	3	0.06	0.16 (0.17)
Sister: sister	356	313	40	3	0.07	0.21 (0.17)
Mother: daughter	245	207	36	2	0.05	0.06 (0.20)
2nd degree relative	22	19	3	0	-	-
Avuncular	11	11	0	0	-	-
Half-sibling	11	8	3	0	-	-
3rd degree relative	3	3	0	0	-	-

MZ, monozygotic; DZ, dizygotic; * No. both affected pairs/ (no. both affected pairs + no. discordant pairs); no., number of.; ¶ Intraclass correlation; S.E, Standard error

3. Heritability of HPV infection

Supplementary Table 2 presents the Overall HPV heritability adjusted covariates by genetic models. The additive genetic variance (A) increased from 0.13 to 0.34 according to the adjusting covariates and genetic model. The pattern is increased for the shared environmental (C) variance, adding covariates from 0.4 to 0.12, but is not significant. When marriage status, childbirth, and the number of children were adjusted, C was 0.00 and therefore the model removed its household effect. The non-shared environmental variance (E) values ranged from 0.65 to 0.75. To explain these genetic model statistics, table 6 includes the loglikelihood and the chi-square difference tests ($\chi 2$) for comparing polygenic and sporadic models, and testing a covariate by suspending it, respectively. In all models, only oral contraceptive use was significantly covariate. Occupation and childbirth was shown to be a comparatively low p-value but finally not significant.

The estimated heritabilities are presented in Table 6. When we compared a simple genetic model (AE model), with a genetic model adding shared environmental effects (C), AE model was best-fitted based on likelihood estimation. Shared environmental effects were not significantly explained by the liability of HPV infection. The heritability of overall HPV genotype was 0.31 (S.E.= 0.01). When heritability was estimated on the viral genotype,

HR did not show significant heritability, but that of LR HPV was estimated to be 0.54 (S.E. = 0.25), slightly higher than overall measure. Accounting for age and oral contraceptive use, the heritabilities of overall and viral-type specific were slightly attenuated, but not materially changed; the heritability of total HPV strains, HR, and LR were 0.24 (S.E. = 0.18), 0.30 (S.E. = 0.20), 0.51 (S.E.=0.26), respectively.

Table 6. Heritability estimates of types of HPV using a variance component analysis.

Model	Participants	Proportion of variance		P	$R^{2¶}$
		A (S.E.)	E		
Overall HPV (N=72)	912	0.31 (0.01)	0.66	0.04	
	899	0.24 (0.18)§	0.76	0.10	0.01
HR HPV (N=44)	912	0.30 (0.10)	0.70	0.10	
	899	0.30 (0.20)§	0.70	0.20	<0.01
LR HPV (N=15)	912	0.54 (0.25)	0.46	0.04	
	899	0.51 (0.26)§	0.49	0.07	0.02

Abbreviation: Overall HPV, all types of HPV; HR, high risk type; LR, low risk type; N, The number of; A, Additive genetic variance; E, Non-shared environmental variance; ¶, Kullback-Leibler divergence R-squared equation to compute maximize polygenic model with no covariates; § Estimates were adjusted for use of oral contraceptive; Common environmental effects were not significant for all variables.

IV. Discussion and Conclusions

In this study, we evaluated the relative importance of host genetics in HPV infection status. To our knowledge, although there have been reports that specific genetic polymorphisms (SNPs) modified the HPV infection risk, it was not clear whether the overall genetic influences can explain meaningful fraction of total HPV infection risk. A heritability estimation among general population, not in clinical settings, will provide a clue for the question of "does host genes matter in HPV infection?" A heritability estimation requires families of general population. Our study confers an advantage of involving both family members and twins. We believe the heritability assessment was reliable because we could have compared multiple patterns of familial aggregation according to genetic distances. We could demonstrate that cervical HPV infection status, the first event of carcinogenic infection, does have moderate degree of host genetic component (heritability of around 0.25~0.30).

Our results suggest host genetic characteristics play a moderate role in HPV infection status, which is supported by the higher concordance to discordance ratio in MZ, compared with other first degree family pairs (3:19 versus 5:82). These results are in agreement with highest tetrachoric correlation for MZ (0.38) among all family relationships (Table 5). This is because the kinship

between MZ and first degree is 1 and 0.5 respectively In addition, (0.38-0.12)/0.5 is the slope between MZ and the first degree measurement of the tetrachoric correlation coefficient which is 0.52. However, none of these values were statistically significant.

A heritability of 0.31 is considered moderate as it is lower than those of complex traits such as height (0.8–0.9)[33], obesity (0.5–0.7)[34, 35], hyperlipidemia (0.5–0.7)[35], and similar to those of most cancers (0.25–0.35)[36], and notably cervical cancer (0.27; range 0.26–0.29)[10]. Considering that HPV infection is widely known as a disease caused by viral factors and sexual behaviors, our findings can add to existing body of evidence explaining the complex nature of the cervical HPV infection. The heritability does not directly imply the presence of essential host genetic susceptibilities. Supposing some behavioral traits that modify the HPV infection status might be also influenced by genetics, the unmeasured behavioral traits related to viral infection process could be reflected to heritability. However, it is less likely because adjustment for oral contraceptive use did not alter the estimates. In previous study, oral contraceptive use was reported as a risk factor for HPV infection and cervical cancer [20, 30, 37, 38]. However, it remains controversial for the association with long-term oral contraceptive use and HPV infection. Rather, these factors might be involved in the transition from HPV infection to neoplastic cervical

lesions [39]. In our study, oral contraceptive use was significant covariate in genetic variance component analysis (Table 6). Variance due to final covariates was 0.01 which is explained by the Kullback-Leibler divergence R-squared equation to compute maximize polygenic model with no covariates. Although the measurement of covariate for oral contraceptive use is small, it is likely that there are mild family clusters in oral contraceptive use. Family cluster is defined as the "epidemiology a grouping of disorders found in more than 2 members of family". In our data, among those ever using oral contraceptives (N=166, Table 3), 60 women were found in more than 2 members in same family, 28 families. According to the family relationship, there was not both use in MZ twins. First degree pairs in both use oral contraceptive and second degree pairs were 6 and 1, respectively (Supplementary Table 2). Therefore, we adjusted only the oral contraceptive use in our final genetic model analysis.

Interestingly, our findings suggested that genetic influences in LR type HPV are stronger than that of HR HPV. This might implicate stronger genetic contribution would exert on LR infections with their generally lower infectivity and persistence than HR strains. This finding could suggest that different loci or number of genes might play roles between HR and LR HPV infection process. However, the smaller size of each viral genotype subgroups requires further replication before confirming the findings.

Our study has some limitations. First, the cross-sectional nature of the study does not allow discrimination between transient and persistent HPV infection. Thus, the heritability in this study may reflect the effects of both process involving initiation and persistence infection status. Second, the study subjects are insufficient to confirm the viral-type specific heritabilities. In addition, sexual behaviors such as oral contraceptive use, was obtained from self-reports which are subject to certain degree of measurement errors and under-reporting; however, the information was obtained before the participant's HPV infection status was reported, and it should not have differential and unlikely to affect the estimates in this study.

In conclusion, Susceptibility to HPV infection has moderated genetic heritability. These findings support previous studies which reported positive associations between specific genetic polymorphism and HPV infection status.

Supplementary Table 1. Effects of genetic and environmental factors by covariates for overall HPV infection

Model	Proportion of variance			Fit of model by covariate				
	A (S.E.)	C (S.E.)	E	Covariate	Loglikelihood	χ^2 (df)	p	$R^{2¶}$
AE	0.34 (0.16)	-	0.66		-252.37	**3.84 (1)**	**0.26**	0.00
				Age	-252.41	0.08 (1)	0.78	
ACE	0.30 (0.50)	0.04 (0.36)	0.66		-252.36	**0.37 (1)**	**0.27**	0.00
				Age	-252.40	0.08 (1)	0.77	
AE	0.35 (0.17)	-	0.65		-246.27	**4.14 (1)**	**0.02***	0.02
				Age	-247.06	1.57 (1)	0.21	
				Education	-247.17	1.81 (1)	0.17	
				Occupation	-248.56	4.59 (1)	0.03*	
ACE	0.28 (0.55)	0.05 (0.39)	0.67		-246.26	**0.33 (1)**	**0.28**	0.02
				Age	-247.04	1.55 (1)	0.21	
				Education	-247.15	1.78 (1)	0.18	
				Occupation	-248.56	4.60 (1)	0.03*	

Abbreviation: HPV, Human papilloma virus; S.E., standard error; A, Additive genetic variance; C, Shared environmental variance; E, Non-shared environmental variance; df, degree of freedom; ¶, Kullback-Leibler R-squared to compute maximize polygenic model with no covariates; §, Loglikelihood of polygenic model for heritability; Bold is statistics in comparing polygenic and sporadic models; *, significant value.

Supplementary Table 1. *(continued)*

Model	Proportion of variance			Fit of model by covariate				
	A (S.E.)	C (S.E.)	E	Covariate	Loglikelihood	χ^2 (df)	p	$R^{2¶}$
AE	0.31 (0.17)	-	0.69		-249.19	**3.00 (1)**	**0.04***	0.01
				Age	-249.20	0.02 (1)	0.88	
				Oral contraceptive	-251.78	5.18 (1)	0.02*	
				Smoke	-249.32	0.26 (1)	0.61	
				Alcohol	-249.53	0.68 (1)	0.41	
ACE	0.22 (0.55)	0.07 (0.38)	0.71		-249.17	**0.17 (1)**	**0.34**	0.01
				Age	-249.18	0.02 (1)	0.88	
				Oral contraceptive	-251.77	5.20 (1)	0.02*	
				Smoke	-249.31	0.28 (1)	0.60	
				Alcohol	-249.51	0.69 (1)	0.41	
AE	0.33 (0.17)	-	0.67		-246.70	**3.40 (1)**	**0.03***	0.02
				Age	-246.79	0.19 (1)	0.67	
				Marriage status	-246.93	0.47 (1)	0.49	
				Childbirth	-249.94	6.49 (1)	0.01*	
				Number of children	-246.73	0.06 (1)	0.81	

Abbreviation: HPV, Human papilloma virus; S.E., standard error; A, Additive genetic variance; C, Shared environmental variance; E, Non-shared environmental variance; df, degree of freedom; ¶, Kullback-Leibler R-squared to compute maximize polygenic model with no covariates; §, Loglikelihood of polygenic model for heritability; Bold is statistics in comparing polygenic and sporadic models; *, significant value.

Supplementary Table 1. *(continued)*

Model	Proportion of variance			Fit of model by covariate				
	A (S.E.)	C (S.E.)	E	Covariate	Loglikelihood	χ^2 (df)	p	$R^{2¶}$
ACE	0.33 (0.17)	0	0.67		-246.70	**3.40 (1)**	**0.03***	0.02
				Age	-246.79	0.19 (1)	0.67	
				Marriage status	-246.93	0.47 (1)	0.49	
				Childbirth	-249.94	6.49 (1)	0.01*	
				Number of children	-246.73	0.06 (1)	0.81	
AE	0.29 (0.18)	-	0.71		-250.22	**2.51 (1)**	**0.06**	0.01
				Age	-250.34	0.25 (1)	0.61	
				Sanitation after sex	-250.50	0.57 (1)	0.45	
				Husband affair	-251.04	1.64 (1)	0.20	
				History of STD	-250.72	1.01 (1)	0.31	
ACE	0.13 (0.54)	0.12 (0.37)	0.75		-250.17	**0.06 (1)**	**0.40**	0.01
				Age	-250.29	0.25 (1)	0.62	
				Sanitation after sex	-250.48	0.63 (1)	0.43	
				Husband affair	-250.96	1.59 (1)	0.21	
				History of STD	-250.71	1.08 (1)	0.30	

Abbreviation: HPV, Human papilloma virus; S.E., standard error; A, Additive genetic variance; C, Shared environmental variance; E, Non-shared environmental variance; df, degree of freedom; ¶, Kullback-Leibler R-squared to compute maximize polygenic model with no covariates; §, Loglikelihood of polygenic model for heritability; Bold is statistics in comparing polygenic and sporadic models; *, significant value.

Supplementary Table 1. *(continued)*

Model	Proportion of variance			Fit of model by covariate				
	A (S.E.)	C (S.E.)	E	Covariate	Loglikelihood	χ^2 (df)	p	$R^{2¶}$
AE	0.29 (0.25)	-	0.71		-235.77	**2.29 (1)**	**0.65**	0.07
				Age	-236.85	2.17	0.14	
				Oral contraceptive	-238.54	5.56	0.02*	
				Education	-237.62	3.71	0.05	
				Occupation	-237.32	3.09	0.08	
				Smoke	-235.83	0.13	0.72	
				Alcohol	-235.87	0.20	0.65	
				Marriage status	-236.32	1.12	0.29	
				Childbirth	-238.75	5.98	0.01	
				Number of children	-235.77	0.01	0.91	
				Sanitation after sex	-235.82	0.10	0.75	
				Husband affair	-236.15	0.77	0.38	
				History of STD	-236.40	1.27	0.26	

Abbreviation: HPV, Human papilloma virus; S.E., standard error; A, Additive genetic variance; C, Shared environmental variance; E, Non-shared environmental variance; df, degree of freedom; ¶, Kullback-Leibler R-squared to compute maximize polygenic model with no covariates; §, Loglikelihood of polygenic model for heritability; Bold is statistics in comparing polygenic and sporadic models; *, significant value.

Supplementary Table 1. *(continued)*

Model	Proportion of variance			Fit of model by covariate				
	A (S.E.)	C (S.E.)	E	Covariate	Loglikelihood	χ^2 (df)	p	$R^{2¶}$
ACE	0.29 (0.25)	0	0.71		-235.77	**2.29**	**0.65**	0.07
				Age	-236.85	2.17	0.14	
				Oral contraceptive	-238.54	5.56	0.02*	
				Education	-237.62	3.71	0.05	
				Occupation	-237.32	3.09	0.08	
				Smoke	-235.83	0.13	0.72	
				Alcohol	-235.87	0.20	0.65	
				Marriage status	-236.32	1.12	0.29	
				Childbirth	-238.75	5.98	0.01	
				Number of children	-235.77	0.01	0.91	
				Sanitation after sex	-235.82	0.10	0.75	
				Husband affair	-236.15	0.77	0.38	
				History of STD	-236.40	1.27	0.26	

Abbreviation: HPV, Human papilloma virus; S.E., standard error; A, Additive genetic variance; C, Shared environmental variance; E, Non-shared environmental variance; df, degree of freedom; ¶, Kullback-Leibler R-squared to compute maximize polygenic model with no covariates; §, Loglikelihood of polygenic model for heritability; Bold is statistics in comparing polygenic and sporadic models; *, significant value.

Supplementary Table 2. Pair-wise concordance-discordance of oral contraceptive use

Relationship pair	Count	Both non-use	Discordant	Both use
MZ twin pair	142	136	6	0
1st degree relative	633	598	19	6
2nd degree relative	22	8	13	1
3rd degree relative	3	3	0	0

MZ, monozygotic

V. References

1. Medicine, T.N.P.i.P.o., *Stockholm, Sweden: The Nobel Assembly at Karolinska Instituet; .* 2008.
2. Nour, N.M., *Cervical cancer: a preventable death.* Rev Obstet Gynecol, 2009. **2**(4): p. 240-4.
3. Bosch, F.X., et al., *The causal relation between human papillomavirus and cervical cancer.* J Clin Pathol, 2002. **55**(4): p. 244-65.
4. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12-9.
5. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
6. Ghittoni, R., et al., *Role of human papillomaviruses in carcinogenesis.* Ecancermedicalscience, 2015. **9**: p. 526.
7. Munoz, N., et al., *Epidemiologic classification of human papillomavirus types associated with cervical cancer.* N Engl J Med, 2003. **348**(6): p. 518-27.
8. Trottier, H. and E.L. Franco, *The epidemiology of genital human papillomavirus infection.* Vaccine, 2006. **24 Suppl 1**: p. S1-15.
9. Castle, P.E., et al., *Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study.* BMJ, 2009. **339**: p. b2569.
10. Magnusson, P.K., P. Lichtenstein, and U.B. Gyllensten, *Heritability of cervical tumours.* Int J Cancer, 2000. **88**(5): p. 698-701.
11. Vink, J.M., et al., *Cervix smear abnormalities: linking pathology data in female twins, their mothers and sisters.* Eur J Hum Genet, 2011.

19(1): p. 108-11.

12. de Sanjose, S., et al., *Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis.* Lancet Infect Dis, 2007. **7**(7): p. 453-9.
13. Shin, A., et al., *Population attributable fraction of infection-related cancers in Korea.* Ann Oncol, 2011. **22**(6): p. 1435-42.
14. Shin, H.R., et al., *Prevalence of human papillomavirus infection in women in Busan, South Korea.* Int J Cancer, 2003. **103**(3): p. 413-21.
15. Bauer, H.M., et al., *Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon.* Sex Transm Dis, 1993. **20**(5): p. 274-8.
16. Hariri, S., et al., *Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006.* J Infect Dis, 2011. **204**(4): p. 566-73.
17. De Sanjose, S., et al., *Cervical human papillomavirus infection in the female population in Barcelona, Spain.* Sexually Transmitted Diseases, 2003. **30**(10): p. 788-793.
18. Deacon, J.M., et al., *Sexual behaviour and smoking as determinants of cervical HPV infection and of CIN3 among those infected: a case-control study nested within the Manchester cohort.* Br J Cancer, 2000. **83**(11): p. 1565-72.
19. Kim, Y.T., *Current status of cervical cancer and HPV infection in Korea.* J Gynecol Oncol, 2009. **20**(1): p. 1-7.
20. Lee, H., et al., *Risk factors associated with human papillomavirus infection status in a Korean cohort.* Epidemiol Infect, 2014. **142**(8): p. 1579-89.
21. Castellsague, X., et al., *Male circumcision, penile human papillomavirus infection, and cervical cancer in female partners.* N Engl J Med, 2002. **346**(15): p. 1105-12.

22. Safaeian, M., et al., *Single nucleotide polymorphisms in the PRDX3 and RPS19 and risk of HPV persistence and cervical precancer/cancer.* PLoS One, 2012. **7**(4): p. e33619.
23. Chan, P.K., et al., *Association between HLA-DRB1 polymorphism, high-risk HPV infection and cervical neoplasia in southern Chinese.* J Med Virol, 2007. **79**(7): p. 970-6.
24. Guzalinuer, A., et al., *[Association between HPV infection and HLA-DQB1 alleles polymorphism in the cervical carcinogenesis in Uyghur women in southern Xinjiang].* Zhonghua Zhong Liu Za Zhi, 2010. **32**(7): p. 492-6.
25. Wang, S.S., et al., *Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer.* J Infect Dis, 2009. **199**(1): p. 20-30.
26. Wang, S.S., et al., *Common genetic variants and risk for HPV persistence and progression to cervical cancer.* PLoS One, 2010. **5**(1): p. e8667.
27. Chen, D., et al., *Genome-wide association study of HPV seropositivity.* Hum Mol Genet, 2011. **20**(23): p. 4714-23.
28. 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.
29. Mehal, W.Z., et al., *Role of human papillomavirus in determining the HLA associated risk of cervical carcinogenesis.* J Clin Pathol, 1994. **47**(12): p. 1077-81.
30. Castellsague, X. and N. Munoz, *Chapter 3: Cofactors in human papillomavirus carcinogenesis--role of parity, oral contraceptives, and tobacco smoking.* J Natl Cancer Inst Monogr, 2003(31): p. 20-8.
31. Muthen, B.O., *Dichotomous Factor-Analysis of Symptom Data.* Sociological Methods & Research, 1989. **18**(1): p. 19-65.
32. Almasy, L. and J. Blangero, *Variance component methods for*

analysis of complex phenotypes. Cold Spring Harb Protoc, 2010. **2010**(5): p. pdb top77.

33. Silventoinen, K., et al., *Heritability of adult body height: a comparative study of twin cohorts in eight countries.* Twin Res, 2003. **6**(5): p. 399-408.
34. Zhou, B., et al., *Genetic and Environmental Influences on Obesity-Related Phenotypes in Chinese Twins Reared Apart and Together.* Behav Genet, 2015. **45**(4): p. 427-37.
35. Sung, J., K. Lee, and Y.M. Song, *Heritabilities of the metabolic syndrome phenotypes and related factors in Korean twins.* J Clin Endocrinol Metab, 2009. **94**(12): p. 4946-52.
36. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med, 2000. **343**(2): p. 78-85.
37. Moreno, V., et al., *Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study.* Lancet, 2002. **359**(9312): p. 1085-92.
38. Ghanem, K.G., et al., *The association of current hormonal contraceptive use with type-specific HPV detection.* Sex Transm Infect, 2011. **87**(5): p. 385-8.
39. Vaccarella, S., et al., *Reproductive factors, oral contraceptive use, and human papillomavirus infection: pooled analysis of the IARC HPV prevalence surveys.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2148-53.

VI. Abstract in Korean

국문초록

인유두종 바이러스 (HPV) 감염은 성매개 감염이고 자궁경부암의 원인으로 잘 알려져있다. 자궁경부암은 전세계적으로는 여성에게서 네 번째로 발생하는 암이고 그 추세가 점점 줄고 있지만, 여전히 개발도상국 등의 지역에서는 높은 유병률과 많은 사망자수를 내는 암이다. HPV 감염의 주요 원인들로 환경적 요인들은 많이 연구 된 바가 있고, 유전적 요인으로는 선행연구에서 숙주의 유전polymorphism이 보고된 바가 있다. 하지만, polymorphism 자체가 숙주의 HPV 감염 감수성의 직접적 유전요인이라 보기는 아직 어렵다. 따라서 이 연구에서는 HPV 감염상태에 따른 유전적 요인이 있는지 가족-쌍둥이 코호트연구 참여자의 가족관계 분석을 통하여 알아보고자 한다.

2006년부터 2009년까지 이 연구에 참여한 여성 중 912명의 여성이 이 Papnicolaou(Pap) smear 검사를 받아, 자궁경부세포시료를 취합하였다. 채취한 Pap smear시료에서 PCR을 통해 HPV DNA를 증폭하였고 HPV 16S rRNA 분석을 통해 분리하였다. 이 연구에 참여한 가족 수는 260 가족이었고, 그 중 일란성 쌍둥이가 142쌍

포함 되어 있다. 참여자 평균나이는 48였으며 표준편차는 11.1세였다. 연구에 참여한 모든 여성은 연구 참여에 동의하였고, 역학적 설문지와 자궁경부암 예방을 위한 성생활 관련 설문지를 모두 비밀이 보장되는 개인적인 공간에서 스스로 작성하였다.

HPV 감염의 환경적 요인은 설문지 문항 분석을 통해 이루어졌고, 유전적 요인은 가족관계 별 상관도 분석으로 실시하였다. 유전적 요인 분석은 크게 두 가지로 실행하였는데, 첫 번째 방법은 가족관계 안의 HPV 감염상태 일치도 분석이다. 유적적으로 가장 가까운 일란성 쌍둥이와 유전적으로 좀 더 멀리 떨어져 있는 이란성 쌍둥이, 자매간, 모녀간, 그리고 그 보다 좀 더 멀리 떨어져있는 사촌간 분석을 실시하였다. 두 번째 방법은 유전율 분석으로, 유전적 요인과 환경적 요인을 전체 분산의 부분분산으로 계산하여 전체 분산 중 유전적 요인 분산의 비율이 어느 정도인지 분석하는 방법이다.

이 연구의 결과에 따르면, 912명의 연구 참여자 중 HPV 바이러스 감염 발생율은 7.9% 였고, HPV 감염의 환경적 요인 중 가장 중요한 요인은 경구피임약의 복용이었다. 다른 환경적 요인들은 큰 상관도를 보이지 않았지만, 유전적인 영향은 있다라는 것을 확인

할 수 있었다. 가족 관계 별 분석을 통해서 알 수 있었던 것은, 유전적으로 가장 가까운 일란성 쌍둥이에서 HPV 감염 일치도가 가장 높았고, 유전적으로 멀어질수록 줄어드는 양상이 있다는 것이었다. 유전적 분산 분석 결과, HPV 감염의 전제척인 유전율은 0.24-0.31이었고, 특히 가장 중요한 환경적 요인인 경구피임약 복용을 보정한 후의 값이 0.24였다. 바이러스의 종류별 HPV 분석결과, 자궁경부암 위험이 높은 HPV는 통계적으로 유의미하지 않았고, 자궁경부암 위험이 낮은 HPV의 유전율은 0.51-0.54로 매우 높게 나오는 것을 관찰 할 수 있었다.

이것으로 이 연구에서는 HPV의 감염위험요인 중 유전적인 요인이 있는 것을 확인 할 수 있었다. 또한 바이러스의 종류별 결과값이 다른 것을 볼 때, 자궁경부암 바이러스의 감염에 대한 이해를 더욱 깊이 할 수 있고 나아가 건강적 중재와 백신개발에도 바이러스 종류별 대처를 다르게 하는 발전을 기대해 볼 수 있다.

학번: 2010-22098