Assessment of potential interaction by maternal smoking in a genome-wide association (GWAS) study of Down Syndrome-associated Atrioventricular Septal Defects (AVSD)

By

Ana María Mesa Restrepo

B.Sc.

Universidad EIA

2016

Thesis Committee Chair: Stephanie Sherman, Ph.D.

An abstract of

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

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

2020

Abstract

Assessment of potential interaction by maternal smoking in a genome-wide association (GWAS) study of Down Syndrome-associated Atrioventricular Septal Defects (AVSD)

By Ana María Mesa Restrepo

The present analysis examines the potential effect modification by maternal smoking between Down Syndrome (DS) individuals and congenital Atrioventricular Septal Defect (AVSD), using a modified GWAS dataset from several phases of a large Down Syndrome Project based on Emory University. For this analysis, maternal smoking was defined as having had >100 cigarettes before and during pregnancy of the infant with DS. The analyzed population set consisted in a total of 496 individuals with Down Syndrome, 251 were males and 245 were females. The study sample was composed by 243 subjects with DS and complete AVSD (cases) and 253 subjects with DS and structurally normal hearts (controls). The logistic regression analysis was performed in plink 2/1.9 and included maternal smoking, proband's sex and the first 5 principal components related to ancestry. The model did not provide evidence of large genome-wide significance by maternal smoking interaction effects but highlighted the importance of considering environmental interactions to characterized genetic variants associated with disorders of complex and multifactorial nature.

Assessment of potential interaction by maternal smoking in a genome-wide association (GWAS) study of Down Syndrome-associated Atrioventricular Septal Defects (AVSD)

By

Ana María Mesa Restrepo

B.Sc.

Universidad EIA

2016

Thesis Committee Chair: Stephanie Sherman, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2020

Contents

I. Background

A. Congenital Heart Defects (CHD)

i. General Epidemiology of CHD

Congenital Heart Defects (CHD) are the most common type of congenital abnormalities worldwide and the largest contributor to infant mortality and morbidity (1, 2). The mean prevalence of CHD between 1970-2017 is calculated to be 8.2 per 1,000 live born babies, ventricular septal defects were most prevalent in the general population, with 3.071 per 1,000 new born children being affected (2). Additionally, it is estimated that the infant death rate due to CHD is around 44.0 deaths per 100,000 infants with European ancestry and 56.2 deaths per 100,000 infants with African ancestry (3). Also, CHD is a major cause of early fetal demise. For example, CHD is identified in 10% of stillbirths. Taken together, about 1.35 million infants are born with CHD every year (4).

Studies report marked heterogeneity among geographical regions, with Africa having the lowest prevalence (2.3 per 1,000) and Asia the highest (9.3 per 1,000), while North America is in the intermediate (8.1 per 1,000). This difference among race/ethnic groups could be partially explained because of the availability of diagnostic technology, a limitation that is starting to change. However, particular environmental and or genetic factors most likely have an influence (2, 4).

CHDs are a heterogeneous group of defects. Locations of some of the most common heart malformations and their estimated prevalence were described by Kalisch-Smith et al. and are presented in Table 1 (5):

Table 1. Estimated prevalence of CHD per thousand live births

CHD type	Acronym	Prevalence by 1,000 births
Ventricular Septal Defect	VSD	4.0
Atrial Septal Defect	ASD	1.0
Persistent Ductus Arteriosus	PDA	0.8
Bicuspid aortic valve	BAV	14.0
Tetralogy of Fallot	TOF	0.4
Double outlet right ventricle	DORV	0.2
Atrioventricular Septal Defect	AVSD	0.3
Transposition of the Great Arteries	TGA	0.2
Hypoplastic left Heart	HLH	0.2
Persistent Truncus Arteriosus	PTA	0.1
Aortic coarctation/stenosis	CoA	0.8

CHDs have a complex etiology, involving both genetic variants and environmental exposures. The literature mentions environmental teratogens (dioxins, polychlorinated biphenyls, pesticides), maternal exposures (alcohol, smoking, isotretinoin, antiseizure medications), infectious agents (e.g. rubella) and maternal diabetes and obesity as well recognized non-genetic causes of CHD. On the other hand, chromosomal abnormalities are the first recognized genetic cause of CHD. Particularly, CHD occurs in approximately 40% to 50% for infants with trisomy 21 (or DS), in which atrioventricular septal defect (AVSD) is the prototypic lesion. Nevertheless, it is important to notice that the previous described causes explain only around 20-30% of CHD incidence; 70-80% of cases have no known associated risk factors (2, 4, 6).

ii. Genetics of congenital heart defects

After the enormous and multilateral effort for sequencing the human genome, several studies have focused on understanding the genetics of congenital heart disease (CHD). As it was mentioned before, chromosomal abnormalities, like 22q11 deletion and trisomy 21, are highly associated with CHD. In addition, more than 100 human genes, most of them

classified as developmental genes, have been implicated with CHD. These have been identified through whole exome sequencing (WES) approaches among familial CHD as well as case/control studies. WES have been particularly useful when paired with candidate gene studies for singling out likely pathogenic variants (7). Also, genome wide association (GWAS) analysis show that a significant proportion of CHD patients present pathogenic copy number variants (CNV) (4-6, 8).

Another approach to find susceptibility genes is to study those with a specific genetic disorder, such as DS. For example, infants with DS have a 2000-fold increased risk for AVSD in comparison to infants from the general population. It is hypothesized that the statistical signals for genetic variants on the background of trisomy 21 explaining why some infants with DS have AVSD and other do not, may be large and thus, more easily detected (1, 9).

Nevertheless, to date, simple genetic causation for CHD account only for around one-third of recorded CHD cases; this is the motivation to further explore oligogenic factors, environmental factors, and/or gene–environment interaction (5).

iii. Environmental contributors to congenital heart defects

Table 2 lists extrinsic and intrinsic environmental factors that have been associated with CHD. The information is based on Kalisch-Smith et al (5).

Table 2. Environmental contributors to CHD

Extrinsic Factors		**Intrinsic Factors**
Teratogens	**Diet Supplements**	
Thalidomide	Lack of Folic Acid	Diabetes
Excess Retinoic Acid and Vitamin A	Hyperhomocysteinemia	Obesity
Alcohol	Nicotinamide Adenine Dinucleotide (NAD) deficiency	Viral infection (rubella)
Smoking	Vitamin D deficiency	Maternal Hyperthermia
Maternal hypoxia		
Anticonvulsants		
Antidepressants		
Antiarrhythmics		

B. Down Syndrome and Congenital Heart Defect

Down Syndrome (DS) is the most common and widely studied chromosomal aneuploidy that survives to term in humans. It is estimated that about 1 in 700 newborns are born with DS, accounting for around 6,037 annual births in the United States and the highest national prevalence per 10,000 live births for births defects in the country with 14.47 (10). It is caused by trisomy of chromosome 21 and is associated with mild to severe intellectual disability, weak muscle tone (hypotonia) and hyper flexibility, a characteristic facial appearance and other health complications. Individuals with DS present an increased risk of CHD, digestive abnormalities, gastroesophageal reflux, hypothyroidism, hearing and vision problems, leukemia and Alzheimer disease earlier in life (11).

Most cases of DS, close to 95%, occur as a result of an error in meiotic cell division called nondisjunction, in which a gamete ends up with an abnormal number of chromosomes. However, this condition also appears when part of chromosome 21 attaches to another chromosome during the formation of gametes in a parent or very early in fetal development, this error is known as translocation and comprises around 3% of DS cases. The remaining

2% of DS cases are known as mosaic, in which the trisomy is present in only some of the body's cells (11-13).

It has been well stablished that CHD is one of the most important clinical manifestations in children with DS, due to its significant impact on morbidity and mortality for this population (14). Around 40% to 60% of individuals with DS present some type of CHD, being atrioventricular septal defects (AVSD) the most common; it comprises 30-40% of all DS-associated CHD (15). Moreover, in the general population, 1/10,000 infants have an AVSD and among infants with AVSD, 60% have DS (4, 16). This leads to a 2000-fold increased risk of presenting AVSD compared with the euploid population (1).

a. Population based observational studies

The current literature present observational studies showing the prevalence and spectrum of the congenital cardiac malformations seen in DS. These characteristics vary in different ethnic groups and countries. However, countries or regions in close geographical proximity or similar genetic makeup exhibit comparable rates of different features of CHD in DS. That is the case for the USA and Latin American countries, and the Mediterranean populations (17, 18).

The population based National Down Syndrome Project, conducted in 6 sites across the United States analyzed a cohort of 1,469 eligible infants with DS. The study found major cardiac defects were present in 44% of the population; of those with CHD, AVSD was present in 39%, secundum atrial septal defect (ASCII) in 42%, ventricular septal defect in 43%, and tetralogy of Fallot in 6% of the proportion with CHD (19).

A German study analyzing 1,618 patients with CHD and DS born between 1980 and 2014 found the most common phenotypes to be AVSD (51.2%), VSD (25.1%), TOF (6.7%) and ASD (8.9%) (14).

An independent study from Brazil analyzing 1,207 with DS, among whom 604 (50%) presented CHD found that the most common CHD were ASD (42.1%), complete AVSD (15.1%), ASD and VSD together (14.6%), VSD (12.7%), patent ductus arteriosus (PDA) (6.6%), patent foramen oval (PFO) (5.6%), TOF (2%), and other diseases (1.3%). Pulmonary hypertension was present in 9.4% of cases with DS+CHD. Out of the total, 150 patients (24.8%) underwent cardiac surgery (20).

A study with 275 infants with DS from the National Pediatric Institute of Mexico, found that 58% of the population presented CHD. The most common type was ASD (24%), VSD (22%), PDA (21%), Interestingly, only 8% presented with AVSD (21).

Finally, a prevalence study conducted in South Korea found their national DS prevalence to be 4.4 per 10,000 live births. Of the 394infants with DS analyzed, 56.9% had a CHD. ASD was the most common defect accounting for 30.5% of DS followed by VSD (19.3%), PDA (17.5%), and AVSD (9.4%) (22).

Other factors have been also identified, in addition to race/ethnicity, that explain some of the variation in risk. For example, a study on 1,011 mothers of infants with DS reported their use of folic acid supplementation during pregnancy. Using logistic regression techniques, they found that lack of maternal folic acid supplementation was more common among infants with DS and AVSD (OR=1.69 95% CI, 1.11-2.58, p=0.007) than among infants with DS and no heart defects (23). In a previous report from the same group, AVSD

was found to be almost twice as prevalent in female babies with DS than in male babies with DS (19, 24).

In addition, a study from Turkey found that the distribution of CHD varies according with maternal age. Babies born of mothers younger than 25 or older than 35 years old are more likely to have CHD (25).

b. Genetic Association Studies

A candidate gene study, evaluating a list of 26 genes in a group of 141 individuals with DS and complete AVSD (cases), compared against a group of 141 individuals with DS and structurally normal hearts (controls) was able to identified potentially damaging variants in almost 20% of individuals with DS+AVSDs but in only 3% of those with DS without CHD. Six genes were specifically implicated: *COL6A1*, *COL6A2*, *CRELD1*, *FBLN2*, *FRZB*, and *GATA5* (9).

A genome wide association study using a logistic regression approach was performed in a group of 452 individuals with DS. Even though no individual variants were found to reach genome-wide significance, two trisomic regions on chromosome 21 and four non-chromosome 21 disomic regions were highlighted and said to warrant further investigation. Their results suggested that common variants of large effect size do not explain the large risk for AVSD among infants with DS. Instead, the authors suggested that multiple variants of low-to-moderate effect size may be involved in risk for AVSD (1). Using the same study sample, the authors examined copy number variants (CNVs) as culprits for risk for DS+AVSD. They found that large, rare deletions increase the risk of DS+AVSD (3).

II. Manuscript

Assessment of potential interaction by maternal smoking in a genome-wide association (GWAS) study of Down syndrome-associated atrioventricular septal defects (AVSD)

Ana Mesa-Restrepo, Stephanie Sherman Ph.D.

A. Abstract

The present analysis examines the potential effect modification by maternal smoking between Down Syndrome (DS) individuals and congenital Atrioventricular Septal Defect (AVSD), using a modified GWAS dataset from several phases of a large Down Syndrome Project based on Emory University. For this analysis, maternal smoking was defined as having had >100 cigarettes before and during pregnancy of the infant with DS. The analyzed population set consisted in a total of 496 individuals with Down Syndrome, 251 were males and 245 were females. The study sample was composed by 243 subjects with DS and complete AVSD (cases) and 253 subjects with DS and structurally normal hearts (controls). The logistic regression analysis was performed in plink 2/1.9 and included maternal smoking, proband's sex and the first 5 principal components related to ancestry. The model did not provide evidence of large genome-wide significance by maternal Smoking interaction effects but highlighted the importance of considering environmental interactions to characterized genetic variants associated with disorders of complex and multifactorial nature.

B. Introduction

Congenital Heart Defects (CHD) are the most common type of congenital abnormalities worldwide and the largest contributor to infant mortality and morbidity (1, 2). The mean prevalence of CHD between 1970-2017 is calculated to be 8.2 per 1,000 live born babies, ventricular septal defects were most prevalent in the general population, with 3.071 per

1,000 new born children being affected (2). Additionally, it is estimated that the infant death rate due to CHD is around 44.0 deaths per 100,000 infants with European ancestry and 56.2 deaths per 100,000 infants with African ancestry (3). Also, CHD is a major cause of early fetal demise. For example, CHD is identified in 10% of stillbirths. Taken together, about 1.35 million infants are born with CHD every year (4).

CHDs have a complex etiology, involving both genetic variants and environmental exposures. The literature mentions environmental teratogens (dioxins, polychlorinated biphenyls, pesticides), maternal exposures (alcohol, smoking, isotretinoin, antiseizure medications), infectious agents (e.g. rubella) and maternal diabetes and obesity as well recognized non-genetic causes of CHD. On the other hand, chromosomal abnormalities are the first recognized genetic cause of CHD. Particularly, CHD occurs in approximately 40% to 50% for infants with trisomy 21 (or Down Syndrome (DS)), in which atrioventricular septal defect (AVSD) is the prototypic lesion. Nevertheless, it is important to notice that the previous described causes explain only around 20-30% of CHD incidence; 70-80% of cases have no known associated risk factors (2, 4, 6).

It has been well stablished that CHD is one of the most important clinical manifestations in children with DS, due to its significant impact on morbidity and mortality for this population (14). Around 40% to 60% of individuals with DS present some type of CHD, being atrioventricular septal defects (AVSD) the most common; it comprises 30-40% of all DS-associated CHD (15). Moreover, in the general population, 1/10,000 infants have an AVSD and among infants with AVSD, 60% have DS (4, 16). This leads to a 2000-fold increased risk of presenting AVSD compared with the euploid population (1).

The largest GWAS to date, involving 210 complete AVSD cases with DS (diagnosed with full trisomy 21) and 242 controls with DS and structurally normal hearts, did not identify any variants exceeding genome-wide significance (1). The authors report summarized their findings stating that there was no evidence for few variants of large effect (e.g., odds ratio > 2.0) explaining the increased risk of AVSD in DS, but that multiple variants of low to moderate effect may play a role (1).

The aim of this current study is to evaluate whether the effect of maternal smoking, defined as having had >100 cigarettes before and during pregnancy of the infant with DS, may interact with genetic variants and explain some of the risk for AVSD in the same population previously evaluated by Ramachandran et. al (1, 3).

C. Materials and Methods

i. Subject's description

Based at Emory University in Atlanta, GA, a large research project has been enrolling families of infants born with DS since 1989 with the aim of learning more about the cause of DS and its related medical problems. The first phase, a population-based study, referred as the Atlanta Down Syndrome Project (ADSP), enrolled individuals born with DS in the Atlanta five-county Metropolitan Area from 1989 through 1999 and control births without major congenital anomalies. The second phase, known as the National Down Syndrome Project (NDSP), expanded population-based recruiting efforts of families of infants born with DS from 2000-2004 from 6 geographic sites across the United States: the Atlanta five-county metropolitan area (GA), statewide in Arkansas (AR), Iowa (IA), and New Jersey (NJ), as well as selected geographic areas of California (CA) and New York (NY) (19). Control livebirths were also recruited from each site. Finally, the Emory Down Syndrome

Project (EDSP), a convenience sample of only individuals with DS, continues to enroll families and perform a variety of research to better understand this condition.

For this study, individuals from across all phases of the Down Syndrome Project were included. The analyzed dataset contains information from individuals with a diagnosis of full trisomy 21, their heart diagnosis, either complete atrioventricular septal defect (AVSD) or structurally normal heart, and demographic and environmental variables selected from extensive maternal interviews. The cases presented DS and complete AVSD (DS+AVSD, cases=243), the most severe heart phenotype associated with DS were compared with the controls, those with DS and structural normal hearts (DS+NH, controls=253). The heart status was extracted from medical records and surgical reports. These data overlap with the study of Ramachandran et al. (1) and only included participants whose mothers reported being of non-Hispanic European ancestry (1). In addition, information from 74 subjects with DS and either confirmed AVSD (n=47) or structurally normal hearts (n=27) were added to this dataset. Whole genome sequencing data were available on these individuals and variants identified in the GWAS array dataset of Ramachandran el al. were used.

ii. Variable selection and preparation of dataset

Mothers participating in any of the phases underwent a phone or in person interview lasting around 45 minutes to 1 hour. Questions cover a broad range of topics, including health and reproductive history, family health history, and environmental exposures. Variables related with increased risk for CHD and basic demographics were taken into consideration. It is important to mention that the way questions were asked in each stage have slight variations and some questions were not fully incorporated in each stage. The final dataset contains information on assigned phenotype; proband's sex; maternal and paternal age, race, and

education; maternal smoking habits before and during pregnancy; and family income. The most homogeneous variable related to maternal smoking was the one accounting for having >100 cigarettes before and during pregnancy with the child with DS. While there is information available on more specific behaviors accounting for maternal smoking habits, the stratification of these categories within the limited sample size of the dataset did not allow for further analysis with enough statistical power. Some important exposures such as maternal alcohol consumption, supplement use and occupation could not be investigated because of differences in the associated questions in each phase.

All the information was extracted from REDCap. Separate datasets from each phase were obtained, formatted and cleaned before merging them together with the respective GWAS information from each individual. The datasets were manipulated using SAS 9.4. The distribution of the variables of interest is shown in Table 3.

Table 3. Distribution of variables of interest

Variables of interest	Category (n) / Mean (SD)
Sex	Male (251) Female (245)
Maternal Smoking	Ever (123) Never (282) Missing (91)
Maternal Education	Less than High School (4) Some College (154) College Graduate and above (248) Missing (90)
Paternal Education	Less than High School (10) Some College (158) College Graduate and above (237) Missing (90)
Maternal Age	34.7 (5.7)
Paternal Age	35.8 (6.0)

iii. Genotyping and Quality Control

Lymphoblastoid cell lines or buffy coats from whole blood or saliva were collected from the participants and were used to isolate genomic DNA. The genotyping was carried out using the Affymetrix Genome Wide Human SNP 6.0 array in Emory University (1, 3). For the additional 75 samples, paired-ended whole genome sequencing (WGS) was performed, with an average read depth of 30x.

The quality control process was carried out in plink 2/1.9 (26). The dataset had a total of 496 samples, a filter considering minor allele frequency > 0.1, a p-value > 0.0001 relative to Hardy Weinberg Equilibrium and a missing rate < 5% was applied. 60 variants were removed due to Hardy-Weinberg exact test.

Principal Component analysis was used to exclude samples that were not clustering with non-Hispanic European ancestry. All together, these QC steps yielded a dataset with 243 cases and 253 controls, and 1,407,400 variants analyzed.

Raw genotype data are available in Gene Expression Omnibus (GEO) data repository with accession number: GSE60607.

iv. Analysis approach and model description

Given that the analyzed dataset consisted of case-control individuals, a logistic regression analysis was implemented (27) in plink 2/1.9 (26). The additive model (Equation 1) considered SNPs genotypes of each individual (*ADD*), maternal smoking (*Smk*), proband's sex (*Sex*) and the first 5 principal components (*PC1-PC5*). Maternal smoking was assessed as a potential effect modifier using the interaction term (*ADD*Smk*).

Equation 1. Additive Logistic Model considering the SNPs

$$\ln(odds\ of\ AVSD) = \beta_0 + \beta_1 ADD + \beta_2 Smk + \beta_3 Sex + \beta_4 PC1 + \beta_5 PC2 + \beta_6 PC3 + \beta_7 PC4 + \beta_8 PC5 + \beta_9 ADD * Smk + \varepsilon$$

v. Logistic regression models without considering genome-wide association

Logistic regression models were conducted in SAS 9.4 using the cleaned dataset obtained after merging the information from the 422 individuals extracted from the different phases of the Down Syndrome Project. 212 were controls and 193 were cases. For these models, the variables corresponding to maternal smoking, proband's sex, maternal age and education, and paternal age and educations were used as covariates.

D. Results and Discussion

First, we examined covariates that might explain in part the risk for DS+AVSD. Simple logistic regression models without considering the genotypes confirmed the known elevated risk for DS-associated AVSD in females (OR+2.089 95% CI: 1.398, 3.122). The rest of the covariates including maternal smoking did not give statistically significant results. The analysis was based on the 496 genotyped trisomic individuals, consisting on 253 controls and 243 cases of European Ancestry.

Although we did not have evidence of maternal smoking as a potential risk factor for DS+AVSD, we tested our hypothesis that the potential interaction of maternal smoking and the genotype, controlling for sex and population stratification, may explain in part the increased risk of DS-associated AVSD. Figure 1 shows the quantile-quantile (QQ) plots for the analysis model, with p-values obtained for the interaction term. The unadjusted QQ plot demonstrated a curved tendency, which was amended when the analysis was adjusted for significance values correcting for multiple tests, including Bonferroni, Sidak, and false discovery rate (FDR). The adjusted QQ plot showed good agreement between the expected

and observed p-values, indicating high quality data and also indicating the lack of a genotype-maternal smoking interaction of large effect to explain the increased risk of DS-associated AVSD.

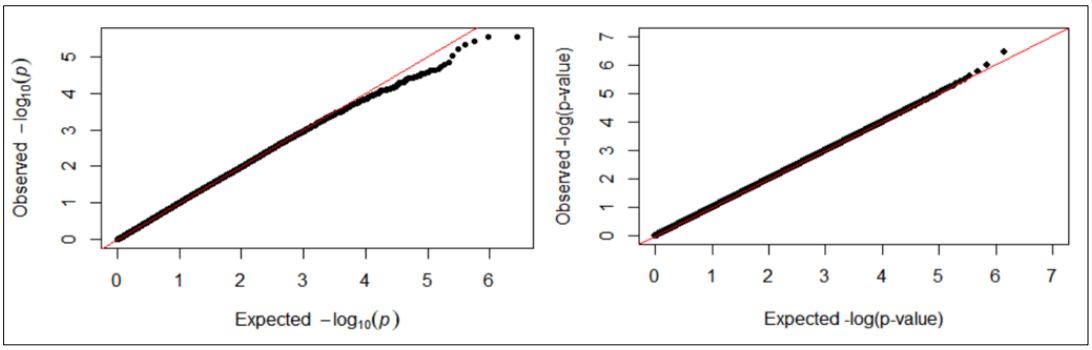

Figure 1. Quantile-quantile plot. Unadjusted (left) and Adjusted (right)

The Manhattan plot (Figure 2) was obtained from the GWAS association results from plink using R-studio. The red line represents the genome-wide statistical significance threshold level ($p<7*10^{-7}$). The blue line represents a p-value<1*10-5. After corrections, there were no statistical SNPs accounting for potential interaction. Table 4 provides a list of the top 5 SNPs. One SNP, rs10251934, was close to genome-wide significance. This SNP is located on human chromosome 7, but has no known associations with any phenotypes to date.

Table 4. Summary statistics for 5 top SNPs

SNP	CHR	A1	A2	MAF	GENO	P
rs10251934	7	T	C	0.3014	37/67/139	6.97E-07
rs10955084	8	C	T	0.4474	72/298/124	1.29E-06
rs10267308	7	G	A	0.3004	36/68/139	1.56E-06
rs512716	11	G	T	0.1667	1/163/331	2.51E-06
rs1357939	7	A	G	0.299	35/69/139	3.62E-06

SNP, single-nucleotide polymorphism; CHR, chromosome; A1, minor allele; A2, major allele; MAF, minor allele frequency; GENO, genotype counts (A1A1/A1A2/A2A2).; P, p-value

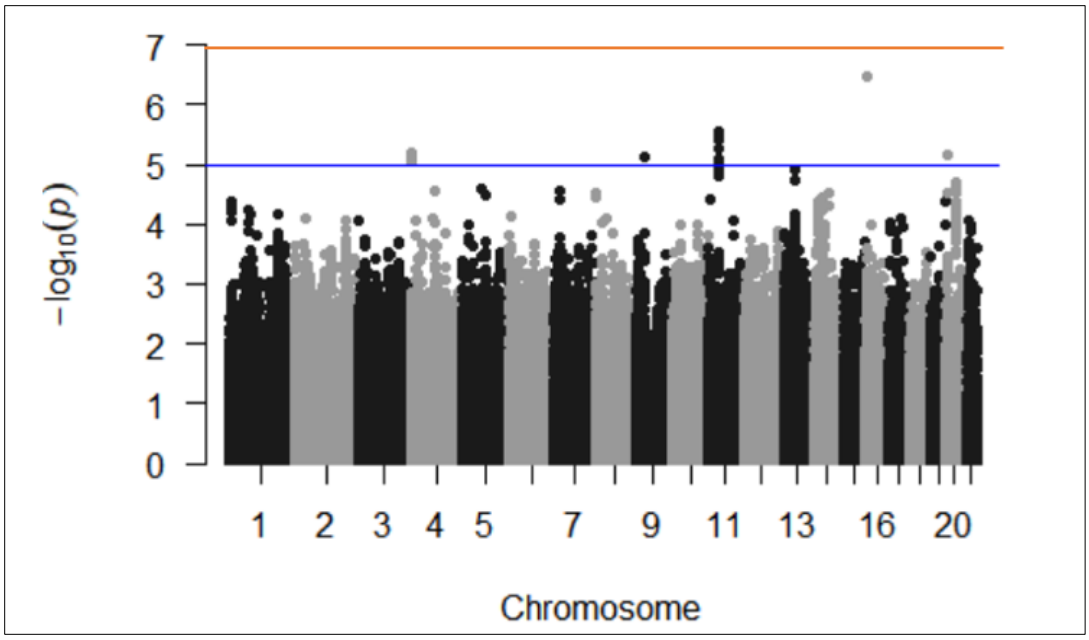

Figure 2. Manhattan plot of the genome-wide association analysis based on the case-control dataset for the non-chromosome 21 autosomal single-nucleotide polymorphisms. The horizontal lines denote the P-value thresholds for genome-wide significance.

E. Conclusion

Our study sample represents one of the largest GWAS analyses of participants who are highly susceptible for heart defect, i.e., those with DS, and specific type of CHD, namely AVSD. Nonetheless, the sample size evaluated is still underpowered to detect variants of modest to low effect sizes and even more so for gene-by-environment interaction.

The Down Syndrome Project based at Emory University offers an immense source for research of this genetic disease and its related health disorders. The project has adapted over the past decades in order to keep recruiting participants and collect the appropriate information to inform future studies. However, one limitation of this study was the heterogeneity of the maternal interviews across the different phases of the study. The interview collected similar information overall, however the exact formatting of the

questionnaire and the addition of extra sections in the newest phases restricted the ability of using other important risk variables for heart defects, those related to more specific maternal smoking behaviors, as well as additional possible environmental exposures such as maternal alcohol consumption, vitamin supplementation during pregnancy and maternal occupation. In the future, studies should better define maternal smoking exposure. Here we only used “ever” smoked in our analyses. Thus, future studies need to increase the sample size as well as refine the exposure variable related to smoking.

F. References

1. Ramachandran D, Zeng Z, Locke AE, et al. Genome-Wide Association Study of Down Syndrome-Associated Atrioventricular Septal Defects. *G3 (Bethesda)* 2015;5(10):1961-71.

2. Liu Y, Chen S, Zuhlke L, et al. Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol* 2019;48(2):455-63.

3. Ramachandran D, Mulle JG, Locke AE, et al. Contribution of copy-number variation to Down syndrome-associated atrioventricular septal defects. *Genet Med* 2015;17(7):554-60.

4. Fahed AC, Gelb BD, Seidman JG, et al. Genetics of congenital heart disease: the glass half empty. *Circ Res* 2013;112(4):707-20.

5. Kalisch-Smith JI, Ved N, Sparrow DB. Environmental Risk Factors for Congenital Heart Disease. *Cold Spring Harb Perspect Biol* 2020;12(3).

6. Gelb BD, Chung WK. Complex genetics and the etiology of human congenital heart disease. *Cold Spring Harb Perspect Med* 2014;4(7):a013953.

7. LaHaye S, Corsmeier D, Basu M, et al. Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. *Circ Cardiovasc Genet* 2016;9(4):320-9.

8. Postma AV, Bezzina CR, Christoffels VM. Genetics of congenital heart disease: the contribution of the noncoding regulatory genome. *J Hum Genet* 2016;61(1):13-9.

9. Ackerman C, Locke AE, Feingold E, et al. An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am J Hum Genet* 2012;91(4):646-59.

10. Parker SE, Mai CT, Canfield MA, et al. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol* 2010;88(12):1008-16.

11. NIH. Down Syndrome. 2019. (https://ghr.nlm.nih.gov/condition/down-syndrome). (Accessed September 29, 2019 2019).

12. CDC. Down Syndrome. 2018. (https://www.cdc.gov/ncbddd/birthdefects/downsyndrome.html#ref). (Accessed 2019).

13. Mayo Clinic. Down Syndrome. 2018. (https://www.mayoclinic.org/diseases-conditions/down-syndrome/symptoms-causes/syc-20355977). (Accessed 2019).

14. Pfitzer C, Helm PC, Rosenthal LM, et al. Dynamics in prevalence of Down syndrome in children with congenital heart disease. *Eur J Pediatr* 2018;177(1):107-15.

15. Morrison ML, McMahon CJ. Congenital Heart Disease in Down Syndrome. In: InntechOpen, ed. *Advances in Research in Down Syndrome*, 2018.

16. Pierpont ME, Basson CT, Benson DW, Jr., et al. Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 2007;115(23):3015-38.

17. Jaiyesimi O, Baichoo V. Cardiovascular malformations in Omani Arab Children with Down's syndrome. *Cardiology in the Young* 2007;17(2):166.

18. Benhaourech S, Drighil A, El Hammiri A. Congenital heart disease and Down syndrome: various aspects of a confirm association. *Cardiovascular Journal of Africa* 2016;27(5):287-90.

19. Freeman SB, Bean LH, Allen EG, et al. Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genet Med* 2008;10(3):173-80.

20. Bermudez BE, Medeiros SL, Bermudez MB, et al. Down syndrome: Prevalence and distribution of congenital heart disease in Brazil. *Sao Paulo Med J* 2015;133(6):521-4.

21. de Rubens Figueroa J, del Pozzo Magaña B, Pablos Hach JL, et al. Malformaciones cardíacas en los niños con síndrome de Down. *Revista Española de Cardiologia* 2003;56(9):894-9.

22. Kim MA, Lee YS, Yee NH, et al. Prevalence of congenital heart defects associated with Down syndrome in Korea. *J Korean Med Sci* 2014;29(11):1544-9.

23. Bean LJ, Allen EG, Tinker SW, et al. Lack of maternal folic acid supplementation is associated with heart defects in Down syndrome: a report from the National Down Syndrome Project. *Birth Defects Res A Clin Mol Teratol* 2011;91(10):885-93.

24. Santoro M, Coi A, Spadoni I, et al. Sex differences for major congenital heart defects in Down Syndrome: A population based study. *Eur J Med Genet* 2018;61(9):546-50.

25. Epçaçan S, Tunçdemir P, ZerrinKarakuş E, et al. Association of Maternal Age with Type of Congenital Heart Disease in Patients with Down Syndrome: A Single-center Study. *Cardiology and Angiology: An International Journal* 2019;8(1):1-8.

26. Chang C, Purcell S. PLINK 1.90 beta. 2020. (https://www.cog-genomics.org/plink/). (Accessed 2020).

27. Clarke GM, Anderson CA, Pettersson FH, et al. Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011;6(2):121-33.

28. Patel CJ, Kerr J, Thomas DC, et al. Opportunities and Challenges for Environmental Exposure Assessment in Population-Based Studies. *Cancer Epidemiology Biomarkers & Prevention* 2017;26(9):1370.

III. Summary

This study investigated a potential gene-environment (GXE) interaction by maternal smoking, one known risk contributor to CHD, employing a phenotypically homogeneous dataset. The population used in this study offers the opportunity to conduct a genome-wide association study (GWAS) in which the risk for AVSD and other CHDs is more prevalent than in the general population. We hypothesize that this highly sensitized set of individuals, may provide the possibility to increase the statistical power to detect specific associations. Although phenotypically, this dataset offers many advantages, the sample size of this relatively rare disease needs to be increase substantially; sample size is currently the primary limitation of this study on those in the literature. This limitation is even greater when approaching GXE interactions.

Moreover, definitions and accurate measurement ranges for environmental exposures can be challenging. The considerations on particular exposures can change over time and the contexts involving them could encompass multidimensional indicators of exposure and behavior. These characteristics, pose greater difficulty in assessing their contributions to complex diseases and disorders (28).

The results from this study did not provide conclusive evidence of a strong effect between gene variants-maternal smoking interaction, most likely due to the small sample size and the crude specification of the smoking exposure variable. While field studies increase sample size and collect better data on smoking, interim studies could implement candidate gene approaches, if candidate genes are well supported; this could help minimize the multiple testing problem. Irrespective of current results, this study helped to define important next step to further investigate GXE interaction effects associated with CHD.