1-1-2018

ROLE OF BREASTFEEDING DURATION, MATERNAL SMOKING DURING PREGNANCY, AND DNA-METHYLATION FOR THE RISK OF ECZEMA

Nandini Mukherjee

ROLE OF BREASTFEEDING DURATION, MATERNAL SMOKING DURING PREGNANCY, AND DNA-METHYLATION FOR THE RISK OF ECZEMA

By

Nandini Mukherjee

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

Major: Epidemiology

The University of Memphis

May 2018

ACKNOWLEDGEMENTS

I express my sincere gratitude to my mentor Prof. Wilfried Karmaus, for giving me an opportunity perform doctoral research under his supervision, and for his continued guidance and motivation. I am forever indebted to him for his hard work and effort towards bringing out the best in me and encouraging me to develop independent thinking and research skills.

I am extremely grateful to all the members of my dissertation committee Prof. John W. Holloway, Prof. Syed Hasan Arshad, Prof. Thomas R. Sutter, and Prof. Hongmei Zhang who devoted their time and effort and provided insightful comments, guidance and support during the development of the dissertation.

I would also like to thank Dr. Yu Jiang and Dr. Su Chen who have been there whenever I have sought guidance.

I am very thankful to all the participants of the Isle of Wight birth cohort, their families and the staff at The David Hide Asthma and Allergy Research Centre for their contribution to our studies.

I thank all my fellow graduate students who have been extremely helpful and supportive during the years of my study.

Finally, I express my heartfelt gratitude towards my parents and brother without whose blessings I could not have achieved this knowledge and opportunity.

ABSTRACT

Prior studies linking breastfeeding duration and maternal gestational smoking, respectively, to offspring eczema report contradictory findings. However, it is not known whether maternal gestational smoking and breastfeeding duration can jointly influence eczema and whether these exposures mediate their effect on eczema via DNA Methylation (DNA-M) of CpG (cytosine-phosphate-guanine) sites.

With data from the Isle of Wight (IOW) birth cohort, UK, we first evaluated the combined effect of maternal gestational smoking and breastfeeding duration on eczema at ages 1-to-2 (n=980), 4 (n=902), 10 (n=966), and 18 years (n=929) using Generalized estimating equations (GEE). We found a protective effect of longer duration of breastfeeding (weeks) on eczema if the mother smoked during gestation (objective (i)). The risk ratios (95% CI) for eczema related to maternal gestational smoking for 3, 9, 15, 21 weeks of breastfeeding were 0.88 (0.66, 1.20), 0.74 (0.54, 1.01), 0.62 (0.41, 0.93), and 0.51 (0.3, 0.88), respectively.

Next, we investigated whether DNA-M at age 10 can mediate the effect of breastfeeding duration on eczema at age 18 (n=276) using linear, log linear regression and structural equation modelling. The CpGs cg03605610 (between *NRC2* and *LINC01276*), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*) and cg26979504 (*HHEX*) were in the causal pathway between duration of breastfeeding and eczema mediating its protective effect (indirect effect estimate: -0.008, p-value: 0.022 (cg03605610); Estimate: -0.008, p-value: 0.049 (cg04928096); Estimate: -0.008, p-value: 0.034 (cg26375057); Estimate: -0.007, p-value: 0.021 (cg26979504)).

Finally, we evaluated whether DNA-M at age 10 associated with combined effect of gestational smoking and duration of breastfeeding can explain the risk of eczema using linear and log linear regression. We found cg07208825 (*ALMS1P*), cg12954512 (Intergenic) and

cg21601919 (FGF18) to be associated with duration of breastfeeding if the mother smoked during gestation (Estimate: -0.006, p-value: 0.046 (cg07208825); Estimate: 0.021, p-value: 0.036 (cg12954512); estimate: 0.008, p-value: 0.016 (cg21601919). These CpGs were also associated with eczema at age 18 (Risk-ratio: 14.23, p-value: 1.27E-07 (cg07208825); Risk-ratio: 0.62, p-value: 3.90E-03 (cg12954512); Risk-ratio: 0.18, p-value: 1.56E-06 (cg21601919).

Future studies are required to corroborate these findings, and understand how the breastmilk composition if the mother smoked during pregnancy influences these genes affecting eczema pathogenesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Eczema or atopic dermatitis is a chronic inflammatory skin disease associated with the impaired epidermal barrier along with dryness and itching [1, 2]. It is an early onset disease affecting over 20% of children in the developed countries, with rising prevalence in low income countries [3]. Atopic dermatitis early in life can lead to development of allergic rhinitis and asthma in adulthood [4] and is also associated with non-allergic comorbidities such as rheumatoid arthritis and lung cancer [5, 6]. Apart from reduced quality of life, impaired psychological and behavioral development, eczema also presents an economic burden on the society [7]. In a recent National Health Interview Surveys in adults (ages 18-85 years) in the United States, it has been found that those with eczema had $371 to $489 higher out-of-pocket costs per person-year [7]. Likewise in Asia Pacific countries, the direct costs of atopic dermatitis ranged from USD 199 in Thailand to USD 4,842 in Australia per patient per year [8] making it a global economic and public health burden.

Eczema is can be explained by genetic factors and environmental factors [9]. The deficiency of filaggrin [10] and the related "loss-of-function" genetic variants of the filaggrin (*FLG*) gene have been found to be associated with heritable epidermal barrier function impairment and eczema [9, 11, 12]. These mutations affect both priming and chronicity of eczema and have been widely replicated [13]. Filaggrin (filament-aggregating protein) represents an integral part of the epidermis which aids in combining the keratin filaments into dense bundles at the granular layer–stratum corneum boundary. It is vital for the development of the cornified envelope and to maintain the epidermal barrier function [11]. Filaggrin haploinsufficiency, i.e. partial or complete loss of filaggrin protein, is associated with the

improper development epidermal barrier increasing water permeability of the skin allowing a higher allergen penetration [14]. In a study by Ziyab *et al*, it has also been found that combination of *FLG* variants and allergic sensitization leads to an increased risk of eczema in subsequent years [15]. However, *FLG* loss of function only explains less than 10% of the disease [16].

Apart from the genetics, several environmental exposures such as breast feeding and smoking influence the pathogenesis of eczema [17, 18]. A review by Kramer states that exclusive breastfeeding for at least 3 months is associated with a reduction in the risk of atopic dermatitis at infancy [19]. In a study by Ito *et al* they demonstrated that longer breastfeeding is associated with an increased risk of atopic dermatitis up to the age 42 months [20]. A randomized trial comparing an intervention with prolonged and exclusive breastfeeding, reported that there was a 54% lower risk of flexural eczema at age 16 years in the intervention group compared to those who received usual care [21]. Chiu *et al* found breastfeeding for more than 6 months was associated with a significantly lower risk of developing eczema at ages 1 and 2 years [22]. In a meta-analysis by Yang *et al*, they found a no strong evidence of exclusive breastfeeding for at least 3 months to be protective effect against AD, even if the children had a positive family history [23]. Thus, these findings indicate that the association of breastfeeding and eczema is in dispute.

Maternal smoking during pregnancy also has been found to be associated with eczema [24, 25]. Kim *et al* found that maternal smoking during pregnancy increased the risk of eczema in a cross-sectional study [26]. Some studies have also reported a protective effect of maternal smoking on eczema. A large study based in Denmark showed lower odds of atopic eczema at ages 14-18 years if the mother smoked during late pregnancy [27]. Another large cohort study

found reduced odds of eczema at 5 years of age in children whose mothers smoked during pregnancy [28]. Ek *et al*, reported that maternal smoking around birth is associated with a lower odds of being diagnosed with eczema in a large study of UK biobank data in participants ranging from ages 37 to 73 years [29].

The above studies show contradictory findings regarding the association of both smoking during pregnancy and breastfeeding with eczema. If the mother smokes during pregnancy the offspring is exposed to smoke *in utero*, and additionally components of smoke may be stored in the breast fat tissues and passed on the child during breastfeeding [30, 31]. Interestingly, the composition of tobacco smoke resembles that of coal tar, a traditional remedy for eczema [32] which targets the skin barrier dysfunction. Similar to tobacco smoke, coal tar contains a lot of high-molecular-weight hydrocarbons and polycyclic aromatic hydrocarbons (PAHs) [33]. To improve the epidermal barrier function, coal tar induces a complex molecular mechanism by activating the AhR/ARNT (aryl hydrocarbon receptor/ aryl Hydrocarbon Receptor Nuclear Translocator) pathway [32]. Ligand activation of this pathway may activate EGFR signaling leading to cell proliferation [34]. Downstream components of this signaling pathway also regulate downstream the *FLG* (filaggrin), *LOR* (loricin), and *K10* gene expression [35]. Regarding breastfeeding, it is known that if the mother smokes, nicotine, which has anti-inflammatory effect [36] can be passed to the offspring via breastmilk [30]. Apart from nicotine, cytokines such as IL-1α [37] and TNF-α[38], IL-1β and IL-8 [39] concentrations are different in the breastmilk of mothers who smoked during pregnancy compared to non-smoking mothers which may have an influence on inflammatory diseases like eczema.

It is not yet known how environmental exposures like breastfeeding and gestational smoking mediate their effects. One possible mechanism by which breastfeeding may influence

eczema is by altering DNA methylation (DNA-M), an epigenetic mechanism that can change the gene expression without mutation [40]. Findings suggest that the methylation of the 5^{th} carbon of cytosine nucleotides at cytosine-phosphate-guanine (CpG) sites in the DNA is affected by environmental exposures [41] and that early life exposures can have a long-term effect on DNA-M [42]. DNA-M is an epigenetic phenomenon involving addition of methyl groups to cytosine bases in the cytosine-phosphate-guanine (CpG) sites on the DNA [43]. It can lead to a change in gene expression without causing any mutation in the DNA [44]. Once differentially methylated, the expression of the gene is altered [45] depending upon the position of the CpG on the gene, thus translating the effect of environmental exposures into a biological mechanism. A lower DNA-M of CpGs in the promoter region (regulatory region of the gene) may lead to a higher gene expression; whereas that in the body region (region between the start and stop codon) is known to modify transcriptional efficiency [44].

Interestingly, it has been found that *in-utero* smoke influence the DNA methylation (DNA-M) of several of the above genes which are induced by coal tar including *AHR, AHRR, CYP1A1* [42, 46-49], even though the mode of intake of tobacco smoke is different from that of a topical application of coal tar. The effect of maternal smoking during pregnancy has been intensively investigated in the cord blood and CpGs belonging to the biological processes such as nervous system development, phosphate-containing compound metabolism, and cell to cell communication have been identified in a genome-wide Consortium Meta-analysis across 13 cohorts [50]. Although not many studies have explored the effect of breastfeeding duration on DNA-M, in a recent meta-analysis five human studies were reported [51] linking breastfeeding duration to methylation of the chromosome region 17q12, *LEP* (LEPTIN) gene, and also influencing global methylation patterns [51].

To date the studies linking breastfeeding duration and gestational smoking, respectively to eczema have produced conflicting results. Additionally, no prior studies have evaluated the combined effect of both these early life exposures on eczema in the offspring. In the first aim of this dissertation I aim to explore the effect of gestational smoking and breastfeeding duration jointly on eczema. Although, present studies have identified CpGs in offspring associated separately with gestational smoking and with breastfeeding duration, it not known yet whether these critical early life exposures can be linked to methylation of CpGs that can explain the risk of future eczema. Thus, in the second aim of this dissertation I will explore whether DNA-M can explain the indirect effect of duration of breastfeeding on eczema. In contrast, in the third aim, I will identify CpGs that are associated with the combined effect of gestational smoking and duration of breastfeeding which in turn predict the risk of eczema.

Thus, the specific aims of this investigation are:

SA1: To assess whether the combined effect of gestational smoking and duration of breastfeeding influence eczema throughout the first 18 years of life (1-18 years) Hypothesis 1: Eczema development is influenced jointly by the early life exposures of gestational smoking and duration of breastfeeding.

SA2: To explore the role of DNA-M as a mediator of the effect of duration of breastfeeding on eczema. Hypothesis 2: DNA-M can explain the effect of duration of breastfeeding on eczema via differential methylation of specific CpG sites.

SA3: To assess whether informative CpGs associated with a combined effect of gestational smoking and duration of breastfeeding can in turn predict eczema. Hypothesis 3: Gestational smoking and duration of breastfeeding can jointly influence DNA-M of specific CpGs which can explain the risk of eczema.

To this end we will analyze data from the Isle of Wight birth cohort, U.K (F0 and F1 generation) to assess the association of duration of breastfeeding, gestational smoking with eczema. In SA1 I will investigate the joint effect of gestational smoking and duration of breastfeeding on repeated measures of eczema at different time points in life (ages 1-2 years, 4, 10, and 18). For SA2 and SA3 in order to have a forward time order with the exposures predicting DNA-M, and then DNA-M predicting eczema, I will analyze associations between gestational smoking and duration of breastfeeding on DNA-M measured at age 10 years and eczema at age 18 years.

This dissertation aims to investigate the effect of two environmental exposures (i.e., duration of breastfeeding and gestational smoking) on eczema in the offspring integrated with epigenetic factors. These specific aims will help to bridge the knowledge gap regarding how the interplay of breastfeeding duration and gestational smoking affects eczema and whether this effect is mediated via epigenetic changes of specific genes.

The results of SA1 will demonstrate how gestational smoking and duration of breastfeeding jointly predict eczema. In SA2, we will identify the CpG sites that are influenced by breastfeeding duration and predict eczema. Thereafter, we will determine if these CpGs can mediate the effect of breastfeeding on eczema. In contrast in SA3 we will identify the CpGs which are influenced by the combined effect of gestational smoking and duration of breastfeeding such that they predict the risk of eczema.

The results of SA1 will enhance our understanding regarding the combined effect of critical early life exposures of gestational smoking and duration of breastfeeding on the development of eczema. The genes of the CpGs identified in SA2 and SA3 will provide an

insight regarding the molecular mechanism by which these exposures can alter the risk of eczema contrasting the effect of duration of breastfeeding and the combined effect of both duration of breastfeeding and gestational smoking on eczema. Overall, these results will help in improved clinical management of eczema by modifying these early life exposures of gestational smoking and duration of breastfeeding. Unlike genetic factors that are fixed at birth DNA-M is influenced by environmental exposures and the identified CpGs can therefore serve as predictive markers of eczema.

Chapter 2

BREASTFEEDING DURATION MODIFIES THE EFFECT OF SMOKING DURING PREGNANCY ON ECZEMA FROM EARLY CHILDHOOD TILL ADOLESCENCE

Abstract

Background: Studies linking breastfeeding duration and gestational smoking, respectively, to offspring eczema report contradictory findings.

Objective: To investigate the interaction of gestational smoking and breastfeeding duration on the development of eczema from early childhood to adolescence.

Methods: Information regarding gestational smoking, breastfeeding duration, and eczema at age 1-to-2, 4, 10, and 18 years were obtained from the Isle of Wight (IOW) birth cohort, UK. Generalized estimating equations (GEE) for repeated prevalence were used to assess this interaction adjusting for confounders. Similar analyses were performed using persistent, incidence, and remitting eczema determined repeatedly for three transition periods of age 1/2-4, age 4-10, and age 10-18.

Results: Longer duration of breastfeeding (weeks) was associated with a lower risk of eczema if the mother smoked during pregnancy. The risk ratios for eczema related to maternal smoking and their 95% CI for 3, 9, 15, 21 weeks of breastfeeding were 0.88 (0.66, 1.20), 0.74 (0.54, 1.01), 0.62 (0.41, 0.93), and 0.51 (0.3, 0.88), respectively. Compared to no eczema in all three transition periods, the risk of having persistent eczema was lowered if the mother smoked during pregnancy with longer duration of breastfeeding.

Conclusions: Results suggest that a protective effect of breastfeeding duration against eczema if the mother smoked during pregnancy. Future studies should investigate molecular pathways influenced by both breastfeeding and cigarette smoking affecting eczema.

Background

Eczema is a chronic relapsing inflammatory skin disease characterized by a disrupted epidermal barrier [1, 2] with a lifetime prevalence ranging from 16-20% in the UK and US [26, 52], which is further increased in the developing countries [3]. Characteristically, an early onset eczema often progresses to other non- or allergic diseases, a phenomenon that is called the atopic march [53]. In addition to reducing quality of life [7, 52], eczema also poses an economic burden on the society due to high cost of treatment [7, 8].

The "loss of function" mutation of the filaggrin (*FLG*) gene is a robustly replicated genetic factor associated with eczema [12, 52]. However, the prevalence of filaggrin mutations varies with populations [10], and these mutations do not fully explain the risk of eczema [54, 55]. This indicates a possible role of early life environmental exposures in the etiology of eczema, such as maternal smoking during pregnancy and breastfeeding.

Enviromental exposures In utero occur during a critical time window of development [56] and expoisure to maternal smoking in this period influences several long-term health outcomes [57]. Kim *et al* found that exposure to maternal smoking during pregnancy increased the risk of eczema in a cross-sectional study [26]. However, other studies have found no [58], or a protective effect of maternal smoking during pregnancy for eczema in offspring [27, 28].

With respect to breastfeeding, several studies have explored its association with the incidence of eczema in offspring and again have reported contradictory findings. A systematic review by Kramer *et al* states that exclusive breastfeeding for $\geq$ 3 months is associated with a reduction in the risk of eczema at infancy [19]. Elbert *et al* demonstrated that non-exclusive breastfeeding is associated with a weak increase in the risk of eczema (adjusted odds ratio 1.11,

95%CI: 1.01, 1.23) [59]. Ito *et al* and Hong *et al* demonstrated that longer breastfeeding is associated with an increased risk of atopic dermatitis in offspring [20, 60], whereas Lee *et al* have reported null effects [61]. Turati *et al* described that early weaning is protective of eczema [62]. Our group showed that in children with filaggrin loss of function mutation, a longer duration of breastfeeding reduced the risk of eczema at age 1 or 2 years [63].

Although the above studies show contradictory findings regarding the association of both smoking during pregnancy and breastfeeding with eczema, no study has yet investigated whether combination of the two factors contribute towards the risk of eczema. We hypothesize that maternal smoking during pregnancy and duration of breastfeeding jointly effect eczema. To this end, we analyzed the interaction of duration of breastfeeding and maternal smoking during pregnancy on eczema from infancy to late adolescence (ages 1 and 2, 4, 10 and 18 years) in boys and girls using data from the Isle of Wight birth cohort.

Methods

Isle of Wight birth cohort

Isle of Wight birth cohort is a longitudinal cohort, established on the Isle of Wight, UK to prospectively study natural history of allergic conditions. From January 1989 to February 1990 1,536 children were born on the Isle of Wight (IOW), among which, 1,456 mother-child pairs were enrolled into the cohort study after exclusion of perinatal deaths, adoptions, and refusals. Children were followed up at 1, 2, 4, 10 and 18 years of age. This study represents a dynamic cohort since some children did not participate at some assessments but rejoined in the next.

The research ethics committee on the IOW (NRES Committee South Central - Hampshire B, U.K.), and in Memphis (University of Memphis Institutional Review Board, FWA00006815)

approved the study and informed written parental consent was obtained for all participants at recruitment and subsequently at each follow-up. The IOW birth cohort has been described in detail elsewhere [64-66].

Data collection

Information regarding eczema symptoms was acquired though detailed interviews and examinations at the follow-ups at 1, 2, 4, 10, and 18 years. A postal or telephone questionnaire was sent if a visit was not possible [67].

Eczema was defined according to the Hanifin and Rajka criteria [68] as chronic or chronically relapsing, itchy dermatitis lasting more than 6 weeks with characteristic morphology and distribution [69]. Since the 1st year and 2nd year follow up was conducted in a small time window, we combined them for analytical purposes.

Smoking during pregnancy of the mothers was ascertained at birth. Postnatal smoking was ascertained by asking at the follow-ups at 1and 2 years whether there is any smoking inside the house.

Regarding breastfeeding, the mother was asked at the 1- and 2-year follow-up whether the child was breastfed and the total duration of breastfeeding (in weeks). In addition, age of starting formula feeding (in weeks) was also ascertained. Duration of exclusive breastfeeding (weeks) was determined by comparing the total duration of breastfeeding with the age of initiating formula feeding or introduction of solid foods.

Socioeconomic status, maternal and paternal eczema, *FLG* loss of function genetic variants, and gender were included as confounders since they potentially associate with maternal

smoking, breastfeeding duration, and eczema. Parental occupation was reported at birth, number of children in the index child's bedroom was collected at age 4 years, and family income was ascertained at age 10 years. These three variables were clustered to produce a family social status variable used in our analyses [70]. Gender, and maternal and paternal eczema were determined at birth.

Genotyping

DNA samples were extracted from 1,150 cohort participants from blood or saliva and assessed using GoldenGate Genotyping Assays (Illumina, Inc, SanDiego, CA) on the beadXpressVeracode platform (Illumina, Inc, SanDiego, CA) per Illumina's protocol. Data were analyzed using the genotyping module of the GenomeStudio Software package (Illumina, Inc, SanDiego, CA). Individuals carrying the minor allele for at least one of the *FLG* variants R501X, 2282del4, or S3247X were classified as having filaggrin haploinsufficiency [15].

Statistical analysis

Assessing the interaction of maternal gestational smoking and duration of exclusive breastfeeding on eczema prevalence

Eczema status repeatedly measured at age 1-or-2 (year 1 and 2 combined), 4, 10, and 18 years was analyzed using the generalized estimating equation (GEE) approach [71]. Since the prevalence of eczema is higher than 10% at each time point, risk ratios [72] were estimated using the log link function. Loss of function filaggrin genetic polymorphisms, age started formula feeding, gender, socioeconomic status of offspring at age 10, maternal and paternal eczema status, and gestational age were included in the model as potential confounders. We started with a model containing all main effects and the interaction of maternal smoking during pregnancy

and duration of exclusive breastfeeding. After evaluation of whether the interaction term was significant, we assessed confounding on the joint effect of duration of exclusive breastfeeding and maternal smoking during pregnancy i.e., RR= exp (β_1×maternal smoking during pregnancy + β_2×duration of exclusive breastfeeding +β_3× (maternal smoking during pregnancy × duration of exclusive breastfeeding)).

Additionally, to further evaluate whether the effect of breastfeeding duration differs if the mother smokes during pregnancy vs. not, we also performed a stratified analysis by maternal smoking status during pregnancy using model with repeated measures of eczema.

In addition to maternal smoking during pregnancy we compared also investigated the effect of post-natal exposure to smoke within 1-or-2 years (year in which the mother breastfed) had an effect on eczema. To this end, we created a variable by combining smoking during pregnancy and postnatal smoking and assessed its interaction with duration of exclusive breastfeeding on eczema.

The effect of maternal smoking during pregnancy may linger on to the first years of life when the child is exposed to breastmilk. Hence, it is difficult to assess the independent effects of both exposure for eczema. To prepare independent variables, we calculated the residual duration of exclusive breastfeeding not explained by smoking during pregnancy by using linear regression and repeated the analysis with the residuals. We assessed the interaction effect of the residual breastfeeding duration and maternal smoking during pregnancy on eczema using log linear models.

A p-value ≤ 0.05 was used to indicate statistical significance. Statistical analyses were performed using the SAS statistical package (version 9.4; SAS Institute, Cary, NC, USA).

Assessing the interaction of maternal gestational smoking and duration of exclusive breastfeeding on patterns of persistent, remitting, or incident eczema

The logistic regression with repeated measurements of eczema considers the prevalence of eczema at the time points of age 1-or-2, 4, 10, and 18, but does not provide insights regarding the consistency of the disease, i.e., whether it was persistent, remitting, or incident eczema. To heuristically identify such patterns of eczema , we applied a configural frequency analysis (CFA) using a FORTRAN CFA program by von Eye [73]. The CFA analysis identifies the profile of the eczema status at age 1-or-2, 4, 10, and 18 (yes/no). It generates *types* (patterns that occur significantly more often than by chance) and *antitypes* (patterns that occur significantly less often than by chance) of the variable from the data. We performed CFA first by excluding any missing eczema information at all the time points (n=1006) (supplementary table A3) using first order CFA with normal approximation of the binomial test, adjusted for multiple testing using Bonferroni correction.

Next, for the transition of age 1-or-2 to 4 years we created a variable that had the value of '1' if eczema present at both time points, and '0' if eczema is not present at both time points, thus comparing persistent vs. never eczema. We created similar variables for the other two transitions from age 4 to 10 years, and 10 to 18 years, respectively. Using the repeated transitions, we evaluated the interaction of maternal smoking during pregnancy with exclusive breastfeeding duration using log linear regression with GEE approach. Similarly, we assessed the interaction with repeated measures of remitting vs. never, and incident vs. never at all three transitions.

Results

There was no significant difference in the prevalence of eczema, duration of exclusive breastfeeding, maternal smoking during pregnancy, *FLG* loss of function variants, gender, socioeconomic status, and maternal and paternal eczema status between the whole cohort (n=1,150), and the sample analyzed at ages 1-2, 4, 10, and 18 (Table 1 and Table 2).

To explore the interplay of maternal smoking during pregnancy and duration of exclusive breastfeeding we plotted the data against the proportion of eczema separately at four time points. We categorized the duration of exclusive breastfeeding into none, 1-15 weeks and more than 15 weeks (Figure 1). The plots revealed that the proportion of eczema varies between those whose mothers smoked during pregnancy compared to those with non-smoking mothers in different categories of breastfeeding.

We then analyzed the interaction effect between maternal smoking during pregnancy and duration of exclusive breastfeeding, on repeated measures of eczema at age 1-or-2, 4, 10, and 18 years using generalized estimating equations (GEE) (Table 3). The model was adjusted for socioeconomic status, maternal and paternal eczema, *FLG* loss of function genetic variants, and gender. The results indicate that the interaction of exclusive breastfeeding and maternal smoking during pregnancy is significantly associated with eczema in offspring (p-value: 0.04).

Table 1. Comparison of prevalence of eczema at ages 1-2 years, 4 years, 10 years, and 18 years between whole cohort and sample analyzed[@]. In addition, number of participants analyzed with incident, persistent, remitting eczema in the three transitions***** of age 1-or-2 to 4, age 4 to 10, and age 10 to 18 are also presented.

		Total cohort N (%)	**Sample analyzed n (%)**	***p*-value**
	Age 1 or 2 years	**N=1150**	**n=980**	
Prevalence	Yes	162 (14.09)	153 (15.61)	0.73
	No	914 (79.48)	827 (84.39)	
	Missing	74 (6.43)		
	Age 4	**N=1150**	**n=902**	
Prevalence	Yes	119 (10.35)	105 (11.64)	0.91
	No	889 (77.30)	797 (88.36)	
	Missing	142 (12.35)		
		N=957[#]	**n=902[#]**	
Transition from age 1-2 to 4 years	*Incidence*	46/ 957 (4.8)	43/ 902 (4.8)	0.99
	Persistence	66/ 957 (6.9)	62/ 902 (6.9)	
	Remission	88/957 (9.2)	86/ 902 (9.5)	
	None	757/ 957 (79.1)	711/ 902 (78.8)	
	Age 10	**N=1150**	**n=966**	
Prevalence	Yes	164 (14.26)	142 (14.70)	0.98
	No	954 (82.96)	824 (85.30)	
	Missing	32 (2.78)		
		N=991[#]	**n=888[#]**	
Transition from age 4 to 10 years	*Incidence*	80/ 991 (8.1)	69/ 888 (7.8)	0.99
	Persistence	64/ 991 (6.5)	58/ 888 (6.5)	
	Remission	50/ 991 (5.1)	44/ 888 (5.0)	
	None	797/ 991 (80.4)	717/ 888 (80.7)	
	Age 18	**N=1150**	**n=929**	
Prevalence	Yes	132 (11.48)	110 (11.84)	0.83
	No	954 (82.96)	819 (88.16)	
	Missing	64 (5.57)		
		N=1054[#]	**n=915[#]**	
Transition from age 10 to 18 years	*Incidence*	63/1054 (6.0)	54/915 (5.9)	0.99
	Persistence	63/1054 (6.0)	52/915 (5.7)	
	Remission	87/1054 (8.2)	78/ 915 (8.5)	
	None	841/1054 (79.8)	731/ 915 (79.9)	

[@]Sample analyzed is sample with no missing for any of the covariates in the model.

*****Eczema in transition periods are defined as incident (present in the second time point but not in the first), persistent (present at both time points), remitting (present at first time point but not in the second).

[#] Only participants with no missing eczema information at both time points are considered, e.g., 991 out of 1150 had eczema information at both ages 4 and 10 years. Similarly, in the analyzed sample out of 966 only 888 had eczema information at both ages 4 and 10 years.

Table 2. Distribution of potential risk factors and confounders in the birth cohort and the analytical sample.

Covariates	Total cohort (N=1150) (n%)	Sample at Age 1-2 (n=980) (n%)	*p*-value	Sample at Age 4 (n=902) (n%)	*p*-value	Sample at Age 10 (n=966) (n%)	*p*-value	Sample at Age 18 (n=924) (n%)	*p*-value
Maternal smoking			0.44		0.12		0.38		0.16
Yes	266	214		184		209		192	
No	879	766		718		757		737	
Missing	5								
***FLG* loss of function variants**			0.91		0.84		0.88		0.98
Yes	118	102		95		101		95	
No	1032	878		807		865		834	
Missing									
Maternal eczema status			0.90		0.76		0.94		0.91
Yes	143	121		109		120		118	
No	998	859		793		846		811	
Missing	9								
Paternal eczema status			0.99		0.8		0.93		0.89
Yes	80	69		61		69		67	
No	1055	911		841		897		862	
Missing	15								
Socio economic cluster			0.90		0.17		0.70		0.59
Low	161	139		108		133		125	
Mid	851	755		711		747		721	
High	98	86		83		86		83	
Gender			0.61		0.87		0.64		0.34
Male	569	474		443		468		440	
Female	581	506		459		498		489	

Table 2 (continued). Distribution of potential risk factors and confounders in the birth cohort and the analytical sample.

Covariates	Total cohort (N=1150)	Sample at Age 1-2 (n=980)	*p*-value	Sample at at Age 4 (n=902)	*p*-value	Sample at at Age 10 (n=966)	*p*-value	Sample at Age 18 (n=924)	*p*-value
				Median (minimum, maximum)					
Exclusive breastfeeding duration (weeks)	6 (0, 80)	6 (0, 80)	0.76	6 (0, 80)	0.46	6 (0, 80)	0.67	6 (0, 76)	0.77
Weeks of gestation	40 (28, 45)	40 (30,45)	0.59	40 (30, 45)	0.35	40 (30, 45)	0.63	40 (30, 44)	0.56

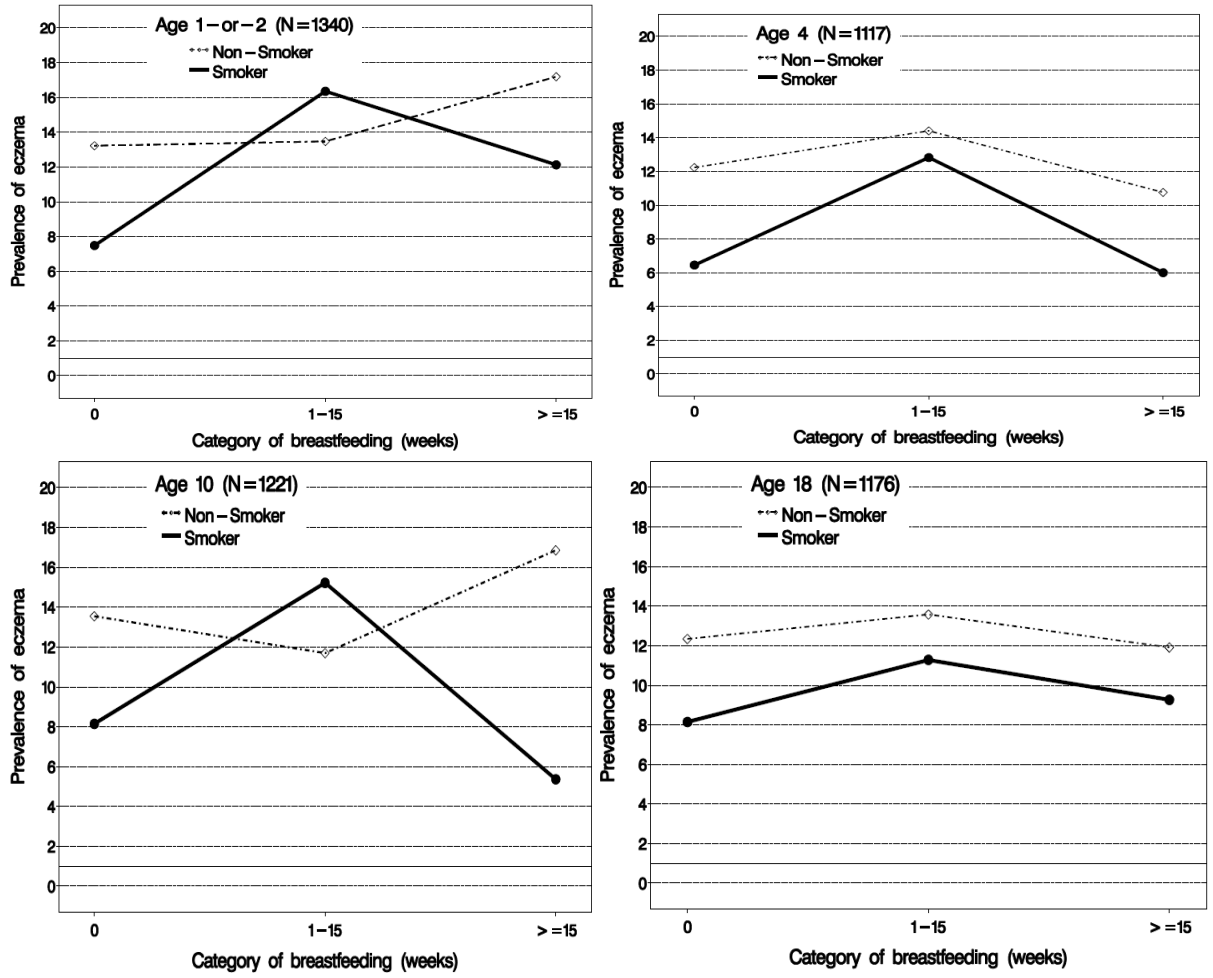


Figure 1. Comparison of the prevalence of eczema over increasing duration of breastfeeding (categorized in 15 week interval) if with or without maternal smoking during pregnancy at ages 1-2 years, 4 years, 10 years and 18 years.

Table 3. Association of maternal smoking during pregnancy and duration of exclusive breastfeeding with eczema at ages 1-or-2, 4, 10, and 18 adjusted for covariates. Transition 1,2 and 3 pertain to the transition from age 1-2 years to 4 years; 4 years to 10 years; and 10 years to 18 years

		Outcome: Eczema prevalence at all the time points **(n=3777)**			Outcome: Eczema persistence at all three transitions **(n=2331)**	
Parameter	**Levels**	**Risk ratio (95% CI)**	**p-value**		**Risk ratio (95% CI)**	**p-value**
Maternal smoking during pregnancy		0.97 (0.70, 1.36)	0.87		0.79 (0.45, 1.40)	0.42
Duration of exclusive breastfeeding		1.00 (0.99, 1.01)	0.87		0.99 (0.98, 1.01)	0.75
Maternal smoking during pregnancy × Duration of exclusive breastfeeding[#*]	0 weeks	0.97 (0.70,1.36)	0.87		0.8 (0.45, 1.42)	0.45
	3 weeks	0.88 (0.66, 1.20)	0.44		0.7 (0.4, 1.22)	0.19
	9 weeks	0.74 (0.54, 1.01)	0.06		0.54 (0.30, 0.95)	0.03
	15 weeks	0.62 (0.41, 0.93)	0.02		0.41 (0.21, 0.82)	0.01
	21 weeks	0.51 (0.30, 0.88)	0.016		0.31 (0.13, 0.74)	0.009
FLG loss of function variants		1.57 (1.16, 2.12)	0.0034		1.91 (1.19, 3.06)	0.007
Time	Year 1/ 2	Ref.	Ref.	Transition 1	Ref.	Ref.
	Year 4	0.74 (0.62, 0.89)	0.0012	Transition 2	1.02 (0.97, 1.08)	0.40
	Year 10	0.95 (0.80, 1.12)	0.54	Transition 3	0.89 (0.74, 1.07)	0.21
	Year 18	0.76 (0.63, 0.93)	0.08			
Weeks of gestation		1.1 (1.0, 1.21)	0.034		1.1 (0.93, 1.31)	0.24
Maternal eczema status		1.31 (0.98, 1.76)	0.06		1.26 (0.75, 2.09)	0.38
Paternal eczema status		1.74 (1.26, 2.41)	0.0008		2.09 (1.26, 3.46)	0.004
Gender (Female is the reference)	Male	0.94 (0.75, 1.17)	0.57		0.97 (0.67, 1.41)	0.89
Socio economic status (High socioeconomic status)	Mid	0.98 (0.67,1.42)	0.90		1.02 (0.56, 1.90)	0.93
	Low	0.92 (0.58, 1.45)	0.70		0.81 (0.36, 1.86)	0.63

Overall significance is 0.04 when the outcome is eczema prevalence at ages 1-or-2, 4, 10, and 18 years.

*Overall significance is 0.03 when the outcome is persistent eczema for all transitions from age 1-or-2 to 4 years; age 4 to 10 years; age 10 to 18 years

Maternal smoking was not associated with eczema if the mothers did not breastfeed (RR: 0.97 (95%CI: 0.70, 1.36)). However, the risk of eczema if the mother smoked was lower as duration of breast feeding increases showing a dose response relationship. Compared to no maternal smoking, if the mother smoked the risk of eczema decreases over duration of exclusive breastfeeding. The risk ratio at 3, 9, 15, 21 weeks were 0.88 (95%CI: 0.66, 1.20), 0.74 (95%CI: 0.54, 1.01), 0.62 (95%CI: 0.41, 0.93), and 0.51 (95%CI: 0.3, 0.88), respectively (Table 3, Figure 2). Among the other covariates, *FLG* loss of function variants, weeks of gestation, maternal and paternal history of eczema, were significantly associated with risk of eczema.

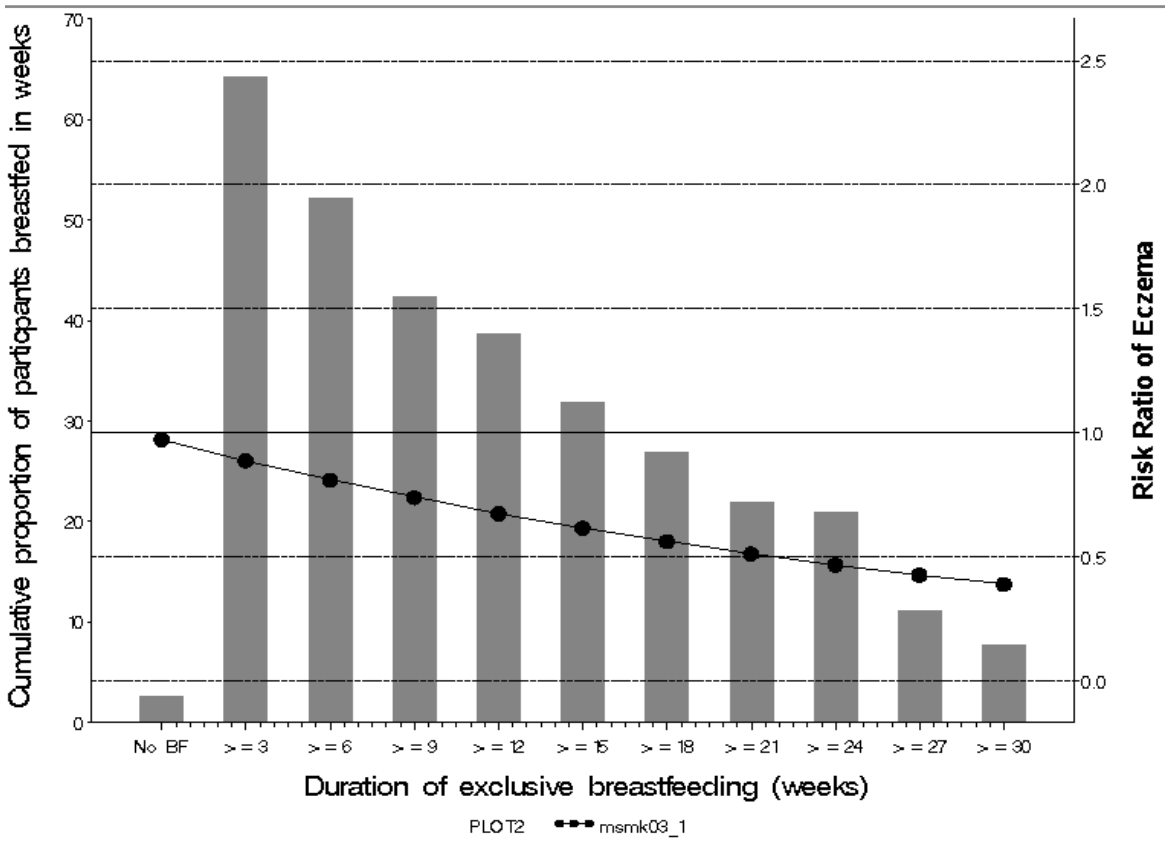


Figure 2. Risk of eczema in offspring with increasing duration of breastfeeding if the mothers smoked during pregnancy compared to non-smoking during pregnancy for repeated measurements of the prevalence eczema at ages 1-2 years, 4 years, 10 years and 18 years (n=3777)

To further inspect whether duration of exclusive breastfeeding is protective of eczema if the mother smoked during pregnancy, we stratified the analyses by maternal smoking (Supplementary table A1). In the strata where the mothers did not smoke, there was no association between duration of exclusive breastfeeding and eczema (overall RR in non-smokers: 1.00, 95%CI: 0.99, 1.01). However, in the strata with maternal smoking, the risk ratio of eczema was lower in the offspring with longer duration of breastfeeding compared to no breastfeeding. Risk ratios for exclusive breastfeeding at 3, 9, 15, 21 weeks are 0.91 (95%CI: 0.84, 0.99), 0.76 (95%CI: 0.58, 0.98), 0.63 (95%CI: 0.41, 0.93), and 0.52 (95%CI: 0.28, 0.95), respectively (Supplementary table A1).

The effect of maternal smoking during pregnancy on eczema may extend to the initial years of life, making it difficult to differentiate between the effects of breastfeeding and maternal smoking during pregnancy. Hence, both variables are not independent. In an additional step, we produced independent variables by calculating the residual exclusive breastfeeding duration that is not explained by maternal smoking status during pregnancy using linear regression. These residuals were proportional to duration of exclusive breastfeeding. Analyses using GEE method showed a significant interaction between residuals of exclusive breastfeeding duration (Supplementary table A2, overall RR: 0.98, 95% CI: 0.96, 0.997, p-value: 0.02). The risk ratios show a similar decreasing trend of the risk ratios if the mother smoked during pregnancy and duration of exclusive breastfeeding 3, 9, 15, 21 weeks 0.74 (95%CI: 0.54, 1.00), 0.61 (95%CI: 0.40, 0.93), 0.51 (95%CI: 0.29, 0.89), and 0.42 (95%CI: 0.21, 0.86), respectively (Supplementary table A2).

Identifying patterns of eczema using configurational frequency analysis (CFA)

The CFA analysis without missing data of four eczema variables (Yes/No) at ages 1-or-2, 4, 10, and 18 years, respectively yielded five *types* i.e., patterns of eczema that occur significantly more often than by chance (Supplementary table A3). One of the five *type* configurations predicted no eczema, and the other four predicted having eczema at three or all four times, indicating that eczema was present statistically significantly more frequent across the time points (persistent eczema). Hence, we compare the dynamic pattern or singular occurrence or multiple occurrence of the disease.

The results of repeated measurement analyses with GEE method in logistic regression revealed that the risk of persistent eczema is lower compared to no eczema with longer duration of exclusive breastfeeding if the mother smokes during pregnancy (Table 3). The risk of persistent eczema if the mother smokes for duration of exclusive breastfeeding at 0, 3, 9, 15, 21 weeks were 0.8 (95%CI: 0.45, 1.42), 0.7 (95%CI: 0.4, 1.22), 0.54 (95%CI: 0.30, 0.95), 0.41 (95%CI: 0.21, 0.83), and 0.31 (95%CI: 0.13, 0.74), respectively. The interaction was not significantly associated with remission vs. none, or incident vs. none eczema.

Discussion

In this longitudinal and prospective birth cohort study we demonstrated that if the mother smokes during pregnancy, a longer duration of exclusive breastfeeding is protective of eczema in offspring measured repeatedly at 1-2, 4, 10, and 18 years (p–value: 0.04). The risk ratios were 0.88 (95%CI: 0.66, 1.20), 0.74 (95%CI: 0.54, 1.01), 0.62 (95%CI: 0.41, 0.93), and 0.51 (95%CI: 0.3, 0.88), at 3, 9, 15, and 21 weeks of duration of exclusive breastfeeding, respectively. To further examine this, we stratified the analysis by maternal smoking, and found that longer

breastfeeding duration was protective of eczema only in the strata where mothers smoked during pregnancy. The risk ratios for exclusive breastfeeding at 3, 9, 15, 21 weeks were 0.91 (95%CI: 0.84, 0.99), 0.76 (95%CI: 0.58, 0.98), 0.63 (95%CI: 0.41, 0.97), and 0.52 (95%CI: 0.28, 0.95), respectively. In addition, for the three transitions of age 1-2 to 4 years, age 4 to 10 years, and age 10 to 18 years, the interaction of maternal smoking during pregnancy and exclusive breastfeeding duration was significantly associated with persistent eczema compared to never eczema.

In the past, several studies have separately assessed the effect of maternal smoking during pregnancy, and duration of breastfeeding, respectively on eczema and reported contradictory results [58, 74, 75]. Our findings indicated that combinations of these two factors influence early onset persistent eczema in the offspring and provide a potential explanation for the previous conflicting reports.

There was no significant difference in the study characteristics of the whole cohort and the study participants at ages 1-2, 4, 10, and 18 years indicating that a selection bias is unlikely. The diagnosis of eczema was made using the questionnaire based on Hanifin and Rajka's criteria [76]. The chances of misclassification are low since the participants developed rashes in typical locations (anticubital or popliteal fossae, ankles, face or neck) validating the eczema definition [76]. All questionnaires focused on eczema in the preceding 12 months to minimize the potential for recall bias that might occur if participants were asked to recall the entire period between assessments.

One of the limitations of this study could be that the maternal smoking status is self-reported. However, we have assessed the DNA methylation of a CpG (cytosine-phosphate-

guanine) site cg05575921 of the *AHRR* (Aryl-Hydrocarbon Receptor Repressor) gene, a well-established epigenetic biomarker of smoking. The offspring whose mothers smoked during pregnancy had significantly lower cord blood methylation of this CpG site at birth (estimate: -0.43 and p-value: 0.009), which is in agreement with the prior studies of prenatal smoke exposure and supports the validity of the smoking information [48]. Breastfeeding duration was determined retrospectively from the mothers at ages 1 year and 2 year follow up. It is unlikely to be affected by recall, since it has been shown that maternal recall within <=3 years provides a valid and reliable estimate of breastfeeding history [77].

It is known that the effect of maternal smoking during pregnancy on the offspring may extend to the first year of life or even longer [57]. Given that exposure to breastfeeding also takes place during the first year of life, it is important to distinguish these two effects on eczema. To take this into account in the analysis, we regressed duration of exclusive breastfeeding on maternal smoking during pregnancy and calculated the residual duration of exclusive breastfeeding. The interaction of residual exclusive breastfeeding duration and maternal smoking during pregnancy showed a similar significant protective effect on eczema with longer duration of breastfeeding.

Apart from assessing the effect of maternal smoking during pregnancy, we performed an additional analysis combining the variables of maternal smoking during pregnancy and postnatal smoking within 1-or-2 years (year in which the mother breastfed) and checked its interaction o with duration of exclusive breastfeeding on eczema. We found a similar protective effect of smoking in pregnancy and/or the first year of life on eczema if the mother breastfeeds exclusively for a longer duration (Supplementary table A4) compared to non-smoking mothers.

Regarding biological mechanism, existing studies suggest that breast milk contains proteins [78] and fatty acids [79] that may stimulate or suppress the immune response affecting eczema. Other studies propose *in utero* smoke exposure changes the T lymphocyte system linking to eczema [28]. None of the studies have yet connected both maternal smoking during pregnancy and gestational smoking and linked these to eczema. However, we found no association of maternal smoking with offspring eczema if the mother did not breastfeed (i.e., 0 weeks in Table 3) and the protective effect was more pronounced at women with longer duration of breastfeeding. This indicates that it is not *in utero* smoke exposure alone that produces a protective effect, and that a significant protective effect is only found if the smoking mother had breastfed for ≥ 9 weeks. Interestingly, tobacco smoke contains compounds such as polycyclic aromatic hydrocarbons (PAHs) similar to coal tar, a topical remedy of eczema [33]. Enzymes of cytochrome P450 family involving the *AHR* and *ARNT* genes detoxify these high molecular weight compounds, in turn, PAHs may induce filaggrin production alleviating eczema [80, 81]. If the mother smokes during pregnancy, tobacco components being lipophilic in nature, are stored in maternal breast fat tissue [82] and may be passed to the infant via breast milk [83]. Thus, longer breastfeeding duration offspring exposed to the coal tar like compounds present in cigarette smoke *in-utero* may experience a different risk of eczema if the mother breastfed the child.

Our findings suggest that interaction of maternal smoking during pregnancy and breastfeeding duration is associated with non-remitting persistent eczema at age 1-or-2, 4, 10, and 18 years suggesting a long term effect protective effect of these early life exposures towards early onset and chronic eczema. It is possible that components of cigarette smoke passed through breastmilk may influence the *AHRR*/*ARNT* system triggering filaggrin and reducing eczema.

Thus, our results suggest that mothers who do not quit smoking during pregnancy should continue to breastfeed the offspring for the longest period possible, since a longer duration of exclusive breastfeeding may be protective of eczema in these children. Future studies should explore the underlying biological mechanisms by examining the genetic, molecular, and epigenetic pathways influenced by breastfeeding and cigarette smoking in relation to eczema.

Chapter 3

DNA METHYLATION MEDIATES THE LONG TERM EFFECT OF BREASTFEEDING DURATION ON ECZEMA AT AGE 18

Abstract

Rationale: Studies suggest that longer breastfeeding duration may be protective of eczema. We hypothesize that this protective effect is mediated by DNA methylation (DNA-M).

Methods: Peripheral blood sample was collected at age 10 years (n=276). Breastfeeding duration information was obtained at age 1-2 years and eczema information ascertained at age 18 year follow up. DNA-M of CpG (cytosine-phosphate-guanine) sites was assessed using the Illumina HumanMethylation450K and 850K BeadChips. In hierarchical screening steps, we identified CpGs at age 10 years associated with breastfeeding duration at $\alpha \leq 0.05$ (linear regression) and eczema at age 18 years (log linear regression) followed by multiple testing adjustment (FDR p-value<=0.05) and confounding assessment. The significant CpGs were then tested in structural equation analyses. To influence by eczema at age 10 on DNA-M at age 10 years we included only the children who were eczema-free at age 10 years (n=276).

Results: Longer breastfeeding duration was associated with higher DNA-M at age 10 years of cg03605610 (between *NRC2* and LINC01276), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*) and cg26979504 (*HHEX*). Higher DNA-M at age 10 years of these CpGs was associated with lower risk of eczema at age 18 years. Structural equation modelling revealed significant indirect effects indicating that methylation of these CpGs at age 10 years are in the causal pathway between duration of breastfeeding and eczema at age 18 (Indirect effects: Estimate: -0.008, p-value: 0.022 for cg03605610; Estimate: -0.008, p-value: 0.049 for cg04928096; Estimate: -0.008, p-value: 0.034 for cg26375057; Estimate: -0.007, p-value: 0.021 for cg26979504).

Conclusion: DNA-M of specific genes may mediate the protective effect of longer breastfeeding duration on eczema.

Background

Eczema, an inflammatory disease accompanied with epidermal barrier dysfunction, is increasing worldwide in children and adolescents [3, 84]. It causes significant morbidity such as dryness, scaling and itching of the skin in early life and may progress to other allergic diseases later in life, a phenomenon named as 'allergic march' [4]. The strongest genetic factor associated with eczema that has been identified to date is loss of function mutations of the filaggrin (*FLG*) gene [12] found in varying proportions in ethnicities and regions (17% to 42% in Caucasian populations and 20% to 30% in Asian populations) [85-87] However, presence of *FLG* LOF mutations explain less than 10% of all eczema cases [16].

In addition to genetic factors, several environmental factors are associated with the risk of eczema, one of which is breastfeeding. In a recently conducted meta-analysis, Lodge *et al* found that exclusive breastfeeding compared to other feeding types was associated with a reduced risk of eczema below 2 years of age [88]. An ISAAC phase three study reported a protective effect of ever breastfeeding with severe eczema (OR 0.79, 95% CI 0.66–0.95) in children aged 6-7 years [89]. Focusing on delayed long-term effects of breastfeeding on eczema, Saarinen *et al* found longer duration of breastfeeding to be protective against atopic eczema in a prospective follow up study from childhood to adolescence [90]. Elbert et al showed that shorter duration of breastfeeding increased the risk of eczema [59]. Others have reported null association [61, 91] and even an opposite effect, showing early weaning to be protective for eczema [62]. Our group

demonstrated that among children with the *FLG* loss of function variants, eczema risk at age 1 or 2 years reduced with longer breastfeeding duration [63].

Although several studies have found breastfeeding to be protective for eczema, it is not yet known how breastfeeding may mediate a protective effect against eczema. One possible mechanism by which breastfeeding may influence eczema is by altering DNA methylation (DNA-M), an epigenetic mechanism that can change the gene expression without mutation [40]. Findings suggest that the methylation of the 5′ carbon of cytosine nucleotides at cytosine-phosphate-guanine (CpG) sites in the DNA is affected by environmental exposures [41] and that early life exposures can have a long-term effect on DNA-M [42]. To date there have been very few studies exploring associations between breastfeeding and DNA-M. A recent meta-analysis reported five human studies [51] showing a link between breastfeeding duration with global methylation pattern [92] and with *LEP* promoter methylation [93]. The other studies in the meta-analysis reported that breastfeeding duration modifies effects of genetic variants on the methylation of the chromosome region 17q12 [94]; and that never breastfeeding was found to be associated with higher odds of methylation (Yes/ No) of the *CDKN2A* promoter [95]. However, none of the studies have yet investigated whether the DNA-M may mediate the effect of breastfeeding on the skin to protect against eczema.

To this end, we analyzed data from the two-generational Isle of Wight birth cohort, U.K. to identify methylation sites at age 10 years in peripheral blood that may mediate the protective effect of duration of breastfeeding on eczema at age 18 years in the F1 generation. We used a hierarchical screening approach to identify informative CpGs and applied structural equation

analyses to assess the indirect effect of duration of breastfeeding on eczema via the selected CpGs.

Methods

Study population

In 1989 a population based birth cohort (n=1536) was established on the Isle of Wight, UK, to prospectively study natural history of allergic conditions (F1 generation). 1,536 F1-children were born on the Isle of Wight (IOW), from January 1989 to February 1990, among which, 1,456 F0-mother-F1-child pairs were enrolled into the cohort study after excluding adoptions, perinatal deaths, and refusals. Follow up of these F1 children was conducted at 1, 2, 4, 10 and 18 years of age [96]. The children of these F1 participants were also enrolled in the study (F2-offspring). Cord blood was collected from the F2-offspring for DNA methylation and gene expression measurements.

The Institutional Review Board of the University of Memphis and the local research ethics committee (South Central - Hampshire B Research Ethics Committee) approved the study. for All participants provided informed written parental or child's consent (at age 18 years) at recruitment and each follow-up. The IOW birth cohort has been described in detail elsewhere [64-66]. In this study we focus on the DNA-M of peripheral blood samples collected from a sample of boys and girls from Guthrie cards (collected within a week after birth), and at 10 years of age (F1 generation), and on DNA-M from cord blood samples collected at delivery of F2-offspring (F2 generation).

Phenotypes

At age 18 years follow up eczema symptom information was acquired though detailed interviews and examinations. Eczema was defined as per the the Hanifin and Rajka criteria [68] as itchy dermatitis lasting more than 6 weeks which is chronic or relapsing with characteristic distribution and morphology [69].Mother was asked at 1- and 2-year follow-ups of the F1 generation, if the child was breastfed and the total duration of breastfeeding in weeks. Maternal smoking was determined at birth. A postal or telephone questionnaire was completed if a visit was not possible [67]. In F2 generation duration of breastfeeding and eczema was assessed in F2 children at 3, 6, and 12 months of age, maternal smoking was determined at birth.

The confounding variables considered in the analyses are socioeconomic status, *FLG* loss of function genetic variants, maternal smoking, maternal and paternal eczema, birth order and gender were included as confounders. To define social status variable we used a cluster of parental occupation was reported at birth, number of children in the index child's bedroom was collected at age 4 years, and family income was ascertained at age 10 years [70].

Genotypes

For genotyping blood or saliva samples were collected from 1150 F1 generation participants and DNA was extracted from these samples. *FLG* genotypes were assessed using GoldenGate Genotyping Assays (Illumina, Inc, SanDiego, CA) on the beadXpressVeracode platform (Illumina, Inc, SanDiego, CA) per Illumina's protocol. Data were analyzed using the genotyping module of the GenomeStudio Software package (Illumina, Inc, SanDiego, CA). Individuals with minor allele for at least one of the *FLG* variants R501X, S3247X , or 2282del4 were classified as having the filaggrin haploinsufficiency [15].

DNA methylation analysis

In the F1 generation at age 10 years DNA was extracted from whole blood, from a subsample of 330 participants (F1) using a standard salting out procedure. Additionally, blood was collected on dried blood spots in 724 neonates in the F1 generation and DNA was isolated from these blood spots using a method based on the procedure described by Beyan *et al* [97]. DNA was extracted from cord blood from subsample of 193 samples in the F2 generaiton. Qubit quantitation was used to determine the DNA concentration. Bisulfite-treatment for cytosine to thymine conversion of one microgram DNA was performed using the EZ 96-DNA methylation kit (Zymo Research, CA, USA), following standard protocol by the manufacturer. Epigenome-scale DNA methylation was assessed using two arrays; the Illumina Infinium HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates >484,000 CpG sites, and the Illumina Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates > 850,000 CpGs associated with over 24,000 genes. The data preprocessing and quality control steps are described below:

F1 generation

Guthrie card methylation data

Methylation data from F1 generation were present in total 8 batches - two from 450k and rest are from EPIC array platform. The quantile normalized beta values are pre-processed and the quality control is performed using the CPACOR pipeline [98]. Since the feature numbers are varying in different batches in EPIC arrays, we pre-processed them separately and then combined with pre-processed 450k array sapmles. Only the shared probes between EPIC and 450k arrays were only chosen. Batch effect has been removed using ComBat [99] in combined

dataset. We excluded from analyses any CpGs which had probe-SNPs within ten base pairs and with minor allele frequency (MAF) greater than 0.007 (which represented about 10 subjects in expectation in the complete study cohort). Finally, we have the methylation information of 302700 number of sites in 724 participants.

Age 10 methylation data

For the methylation data of age 10 years, CPACOR pipeline was used for quality control and and pre-processing the quantile normalized beta values [98]. At age 10 years we have methylation data in total of 7 batches – two from 450k and five from EPIC array platform. We combined samples from the two above mentioned array platforms and batches. We only used the probes which were commonly present in EPIC and 450k arrays were only chosen. ComBat [99] has been used to correct for batch effect. Any CpGs with probe-SNPs within ten base pairs and with minor allele frequency (MAF) greater than 0.007 were not included in the analyses. Finally, we have 349455 CpGs from 330 participants.

F2 generation

Cord blood methylation data

The QC and pre-processing has been done using CPACOR pipeline [98]. These are beta values were generated after doing quantile normalization and background subtraction. Two different runs was performed for 130 samples, from six different batches, in 450k platform and 63 samples from EPIC platform respectively. One sample from 450k array was discarded due to maternal blood contamination. These two groups have been combined and only the shared probes between 450k and EPIC have been taken. Further the data has been batch corrected using ComBat [99]. CpG sites with probe-SNPs within ten base pairs and with minor allele frequency

(MAF) greater than 0.007 were excluded from analyses. Finally, we have 365697 of sites from 192 number of samples.

Cell Type proportions: White blood cell count for the Guthrie card methylation data (F1 generation) were generated from the three original groups (EPIC samples with 2 different chip design and 450k samples). The *estimateCellCounts()* function of Minfi package [100] with reference panel from Bakulski *et al.* 2016 [101] has been to generate cell proportion. Then the cell counts of the three groups were combined together. Similarly for 10 year old samples (F1 generation) white blood cell counts were generated in the separate platform batches (450k and EPIC) and then combined. The *estimatecellcounts()* function in minfi with the default settings using reference panel from Houseman *et al.* 2012 [102] were used for age 10 year cell count. For F2 generation cord blood samples same method mentioned for F1 generation Guthrie card samples has been used. The cell counts have been generated separately for the 450k and EPIC array samples and then later combined.

Statistical analysis

We record the methylation levels of each CpG as beta (β) value that ranges between zero and one. Beta value is the proportion of methylated (M) over methylated (M) plus unmethylated (U) probes (β=M/[c+M+U], with constant c being a constant to prevent diving by zero. For analyses purposes, M-values i.e., logit-transformed β values (M-values, approximated by log2(β / (1-β)) are preferred because β values close to 0 or 1 tend to suffer from severe heteroscedasticity [103]. In our analyses, we regressed the M-values n the cell types and calcuted

the cell type residuals of DNA-M which is free from the cell type effects. These cell type residuals were used throughout the analyses.

To have a clear predictive time order between duration of breastfeeding, DNA-M at age10 and eczema age 18 years, and to avoid DNA-M at age 10 years being influenced by eczema at age 10 years, we included only those participants were eczema free at age 10 years (n=276). Eczema at age 18 thus either presents a re-occurrence of eczema (eczema in childhood but not at age 10) or the first occurrence of eczema. The statistical analysis is divided into the following two stages: (i) Identification of informative CpGs at age 10 years that are linked to breastfeeding and eczema at age 18 years using a hierarchical screening approach (Figure 3). (ii) Structural equational modelling to assess the indirect effect of breastfeeding on eczema at age 18 via methylation at age 10 years of CpGs identified in step (i). Since DNA methylation is influenced by cell type compositions [104], the residual DNA-M after regressing out the cell type proportions were used in all steps.

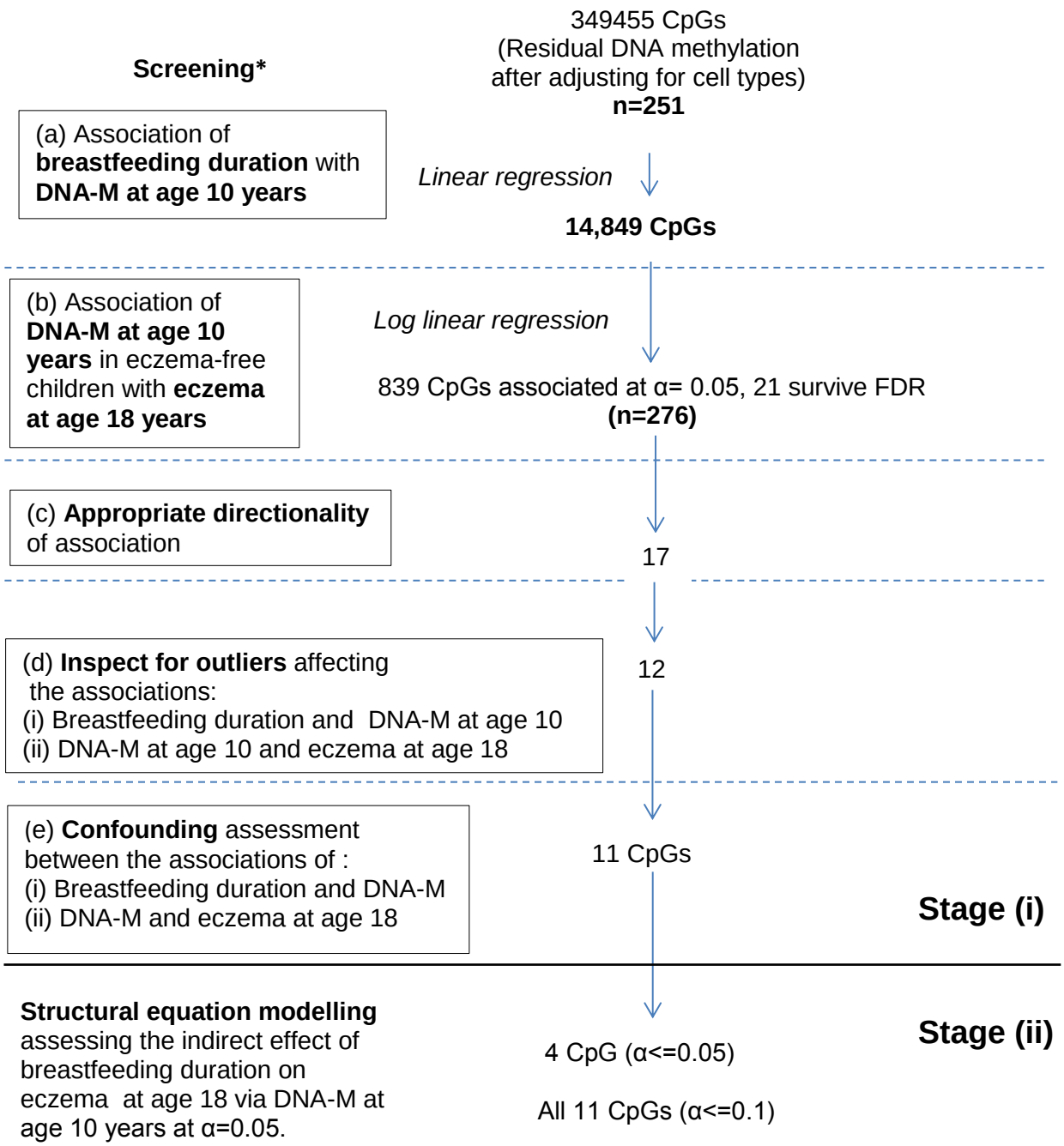


Figure 3. Flow of analyses to identify CpGs at age 10 years that potentially mediate the protective effect of duration of breastfeeding on eczema at age 18. years

The analyses are described in detail below:

(Stage i) Identification of informative CpGs at age 10 years using a hierarchical screening method in the eczema-free at age 10 years (F1 generation) (Figure 3):

Of the 276 children analyzed, 251 had complete information for duration of breastfeeding in weeks. In step (a) of the screening, 349,455 CpGs at age 10 years were tested for association with breastfeeding duration using linear regression (Figure 3). In the next step, with the statistically significant ($\alpha \leq 0.05$) CpGs from step (a), we assessed their association with eczema at age 18 years using log-linear regression adjusting for multiple testing by controlling false discovery rate (FDR) at 0.05. CpGs which survive multiple adjustment (FDR$\leq$0.05) in step (b), were passed on to step (c). In step (c), in order to identify CpGs that could potentially mediate a protective association we selected those CpGs whose methylation was positively associated with duration of breastfeeding, and negatively associated with eczema at age 18 years, or vice versa. For instance, if longer breastfeeding increases the methylation of a CpG and higher methylation of this CpG is associated with lower risk of eczema, then this CpG will be selected. Hence, to mediate protective effect, the selected CpGs should be significantly associated with breastfeeding duration and eczema in opposite directions. The CpGs that pass step (c) werev then examined for presence of outliers in their distributions, which could influence the associations (step (d)) by checking the regression fit plots, and if present outliers were removed. The CpGs that still showed a significant association with duration of breastfeeding and with eczema were was more stringently assessed with confounders in step (e). Potential variables that may confound the association between breastfeeding duration and DNA-M were added in the linear regression models separately for each of the selected CpGs. Similarly, potential confounders were added in the log-linear regression models assessing the association between DNA-M and

eczema at age 18 separately for each CpG. Several genetic and clinical confounders, such as maternal and paternal history of eczema, birth order of the child, filaggrin (*FLG*) loss of function genetic variants, socioeconomic status, and maternal smoking were included. Confounders that change the observed effect sizes in regression models by more than 10% were retained in the model. The CpGs that were still significantly associated after retaining such confounders were qualified for the stage (ii) of the analyses (Figure 3).

(*Stage ii*) *Structural equation modelling*:

To distinguish the direct, indirect, and total effect of breastfeeding on eczema via DNA methylation of CpGs selected in stage (i) we performed structural equation modelling. Separate path models were fitted for each of the CpGs including the relevant confounders using maximum likelihood estimation with robust standard errors.

To substantiate the role of the duration of breastfeeding in altering the methylation level of the above identified CpGs at 10 years of age, we utilized methylation data obtained prior to an established breastfeeding effect collected within a week of birth using heel sticks (obtained from dried blood spots on Guthrie cards) in the F1 generation. We inspected whether duration of breastfeeding can explain the variability in DNA-M at age 10 that is not explained by DNA-M obtained from Guthrie cards. To this end, using the selected CpGs from stage (ii) of the analyses, we ran a linear regression with DNA-M at age 10 years as the response and DNA-M from Guthrie cards and duration of breastfeeding as predictors (n= 172).

In addition, the selected CpGs identified in the F1 generation data were also examined in F2 generation (n=192) using cord blood methylation to assess their association with eczema in infancy prior to the initiation of breastfeeding.

The screening steps involving linear and log linear models was performed using the lm() and glm() functions in R (R package 3.1.1 version). Path analyses were implemented using Mplus Version 7, and the remaining analyses with the selected CpGs were performed using SAS, Version 9.4 (SAS Institute, Cary, NC, USA).

Results

We found no significant differences in the proportion of eczema prevalence, duration of breastfeeding, gender, maternal smoking during pregnancy, socioeconomic status, filaggrin loss of function mutation, and maternal and paternal eczema assessed during pregnancy between the whole cohort (N=1294) and those with DNA-M information at age 10 (n=276) (Table 4).

Focusing on eczema-free children at age 10 years (n=276), in step (a) of the hierarchical screening (Figure 3, Table 5), we tested the association of duration of breastfeeding with epigenome-wide DNA-M at age 10 years. Out of the 349,455 CpGs tested, 14,849 were significantly associated with duration of breastfeeding at $\alpha \leq 0.05$. In the second hierarchical step (b) the 14,543 breastfeeding associated CpGs were tested for an association with eczema at age 18. We found 839 CpGs to be associated with eczema and we corrected for FDR at this step. 21 CpGs survived multiple testing adjustments using FDR. In step (c), 17 out of the 21 were found to have an appropriate directionality; however, 5 of these were not significantly associated with breastfeeding duration and eczema after removing outliers in step (d). The remaining 12 CpGs were then tested for confounding in the next step (e). Gender, socioeconomic status, maternal smoking during pregnancy, and maternal and paternal eczema were added as potential confounders in linear regression assessing the association between breastfeeding duration and DNA-M. For the association of DNA-M and eczema at age 18, all the above confounder and

FLG loss of function and birth order were added in the log-linear model. On retaining the confounders that change the observed effect sizes in regression models by more than 10%, 11 of these 12 still remained significantly associated with both duration of breastfeeding and with eczema, respectively (Figure 3, Table 5).

Of the eleven CpGs that pass the screening, the methylation of cg03605610 (*intergenic*), cg04708216 (*intergenic*), cg04928096 (*AXIN1*), cg015854022 (*intergenic),* cg26375057 *(ZIC1),* cg26979504 *(HHEX),* cg04994084 (*intergenic*), cg24408469 (*CELSR1*), and cg27215578 (*intergenic*) were positively associated with duration of breastfeeding and negatively with eczema at age 18 (Table 5). Whereas the methylation of cg02054493 (*PTRH1*), and cg25311764 (*GCNT3*) were negatively associated with duration of breastfeeding and positively with eczema at age 18 years (Table 5).

In stage (ii), structural equation modelling was conducted with each of the eleven CpGs that were selected in stage (i) of analyses (Table 5) with participants who were eczema free at age 10 years (n=276). The indirect effect of breastfeeding duration on eczema at age 18 years mediated through DNA-M, the direct effect and total effect (combination of direct and indirect) were estimated. Each of the 11 selected CpGs, were separately analyzed using structural equational modelling to test whether the protective effect of breastfeeding is mediated via these CpGs on eczema at age 18 years. Four CpGs cg03605610 (intergenic), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*), and cg26979504 (*HHEX*), tested individually were in the causal pathway between duration of breastfeeding and eczema at age 18 years.

Table 4. Comparison of population characteristics between the whole cohort and those with DNA methylation at age 10 years.

Covariates	**Whole cohort** **N =1294**	**Sample with DNA-M at age 10 years** **n=276**	
	N (%)	n (%)	*p*-value
Maternal smoking during pregnancy			0.25
Yes	300 (23.2)	55 (19.9)	
No	989 (76.4)	219 (79.3)	
Missing	5 (0.39)	2 (0.7)	
Eczema at age 18			0.07
Yes	157 (12.1)	23 (8.3)	
No	1131 (87.4)	253 (91.6)	
Missing	6 (0.46)		
Sex			0.06
Boys	644 (49.7)	155 (56.2)	
Girls	650 (50.2)	121 (43.8)	
Socio economic status			0.97
High	102 (7.9)	24 (8.7)	
Medium	937 (72.4)	209 (75.7)	
Low	176 (13.6)	41 (14.8)	
Missing	79 (6.1)	2 (0.7)	
Filaggrin loss of function variants			0.43
Yes	107 (8.3)	22 (8.0)	
No	968 (74.8)	241 (87.3)	
Missing	219 (16.9)	13 (4.7)	
Maternal eczema during pregnancy			0.84
Yes	160 (12.4)	33 (12.0)	
No	1124 (86.9)	241 (87.3)	
Missing	10 (0.8)	2 (0.7)	
Paternal eczema during pregnancy			0.31
Yes	90 (6.9)	24 (8.7)	
No	1186 (91.6)	249 (90.2)	
Missing	18 (1.4)	3 (1.1)	
Birth Order			0.30
1	469 (36.2)	105 (38.04)	
2	362 (30)	80 (30)	
3	262 (20.2)	71 (25.7)	
Missing	201 (15.5)	20 (7.25)	
Median (p5, p95)			
Duration of breastfeeding (weeks)	8.0 (0, 40)	12.0 (0, 40)	0.06

Table 5. Adjusted association of cell type residuals of CpGs with duration of breastfeeding and eczema at age 18. Participants with eczema at age 10 years were excluded from the analyses (n=276). The associations that include confounders from the screening step are marked, and these confounders were also included in the path analyses. For the rest, no confounder changed the association by more than 10%. The indirect effect is determi ned in the stage (ii) in path analyses (duration of breastfeeding → CpG → eczema at age 18). Only significant indirect effects are shown.

		Association between duration of breastfeeding and DNA-M at age 10 (n=251)		Association between DNA-M at age 10 and eczema at age 18 (n=276)		Indirect effect of duration of breastfeeding on eczema at age 18 via DNA-M	
Ilmnid	**Gene names**	**Estimate**	**p-value**	**Risk ratio**	**pvalue**	**Estimate**	**p-value**
cg02054493	PTRH1	-0.003	0.02	6.9	0.0005*		
cg03605610		0.006	0.01	0.324	3.59E-05	-0.008	0.02
cg04708216		0.002	0.03	0.015	<.0001 $^{\alpha}$		
cg04928096	AXIN1	0.004	0.01	0.224	2.32E-06	-0.008	0.049
cg15854022		0.001	0.02	0.002	<.0001* $^{\beta}$		
cg26375057	ZIC1	0.005	0.03	0.301	3.09E-05	-0.008	0.034
cg26979504	HHEX	0.005	0.01	0.301	5.23E-05	-0.007	0.021
cg04994084		0.005	0.04	0.3927	<.0001		
cg24408469	CELSR1	0.003	0.01	0.096	0.0017$^{\beta}$		
cg25311764	GCNT3	-0.002	0.03	27.4	<.0001 $^{\alpha\,\#\,\beta}$		
cg27215578		0.005	0.03	0.4	0.0048		

Adjusted for *gender $^{\#}$ socioeconomic status $^{\alpha}$ fillagrin loss of function mutation $^{\beta}$birth order

All four CpGs were positively associated with duration of breastfeeding and negatively with eczema at age 18 years in the screening. In other words, longer duration of breastfeeding was associated with higher DNA-M of cg03605610 (Estimate: 0.006, p-value: 0.01), cg04928096 (*AXIN1*) (Estimate: 0.004, p-value: 0.01), cg26375057 (*ZIC1*) (Estimate: 0.005, p-value: 0.03) and cg26979504 (*HHEX*) (Estimate: 0.005, p-value: 0.01). Higher DNA-M was in turn, associated with lower risk of eczema for cg03605610 (Risk ratio: 0.3, p-value: 3.59E-05), cg04928096 (*AXIN1*) (Risk ratio: 0.2, p-value: 2.32E-06), cg26375057 (*ZIC1*) (Risk ratio: 0.3, p-value: 3.09E-05) and cg26979504 (*HHEX*) (Risk ratio: 0.3, p-value: 5.23E-05) (Table 5, Figure 4).

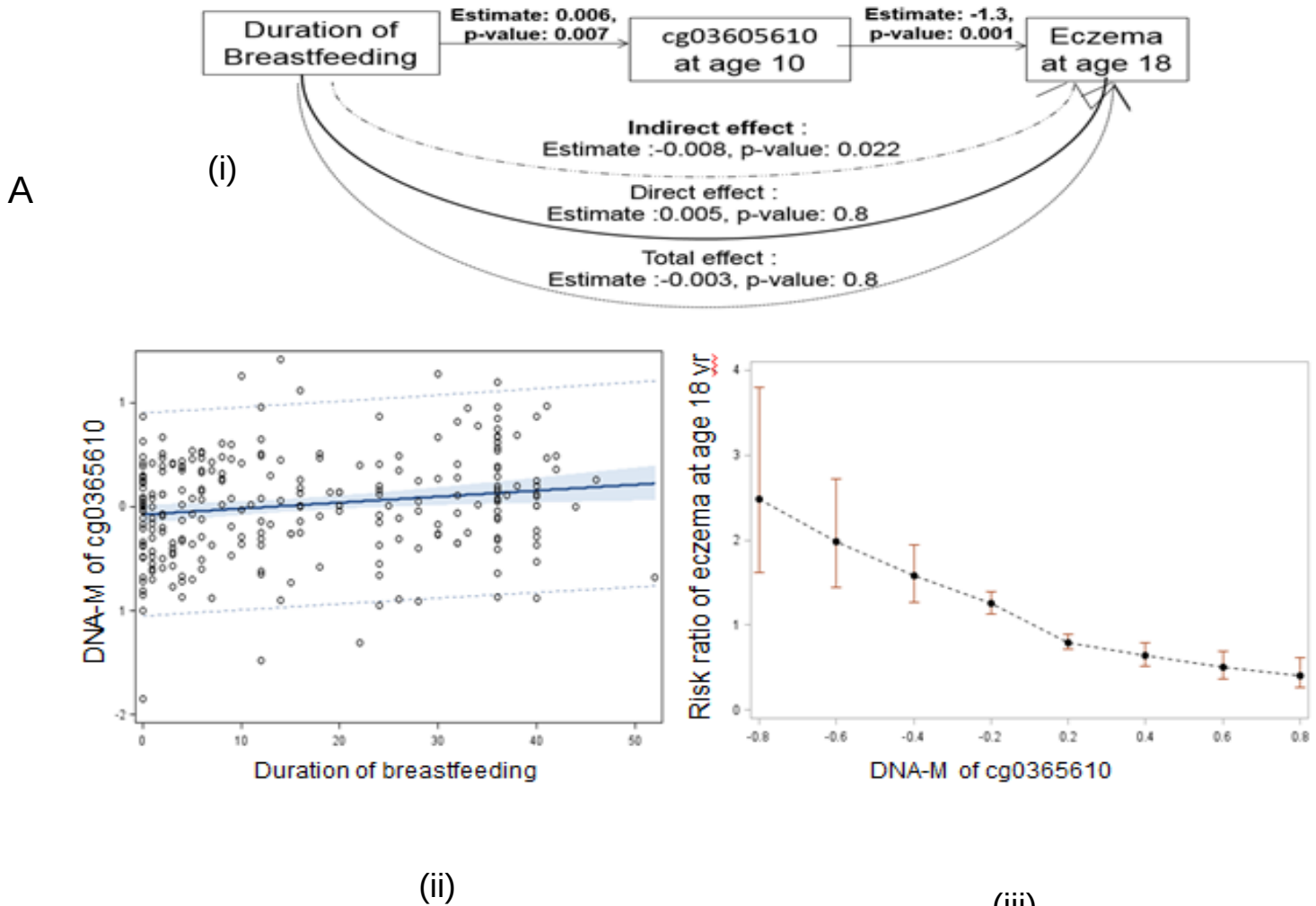


Figure 4. Association of cell type residuals of DNA-M with breastfeeding duration and eczema at age 18 years. CpGs that mediate the protective effect are shown: (A) cg03605610 (B) cg04928096 (*AXIN1*) (C) cg26375057 (*ZIC1*) (D) cg26979504 (*HHEX*). The panels are described as follows: (i) path analyses showing the indirect effect of breastfeeding duration via DNA-M of CpGs on eczema at age 18 years; (ii) Association of DNA-M at age 10 years with duration of breastfeeding; (iii) Association of DNA-M at age 10 years with eczema at age 18 years.

B

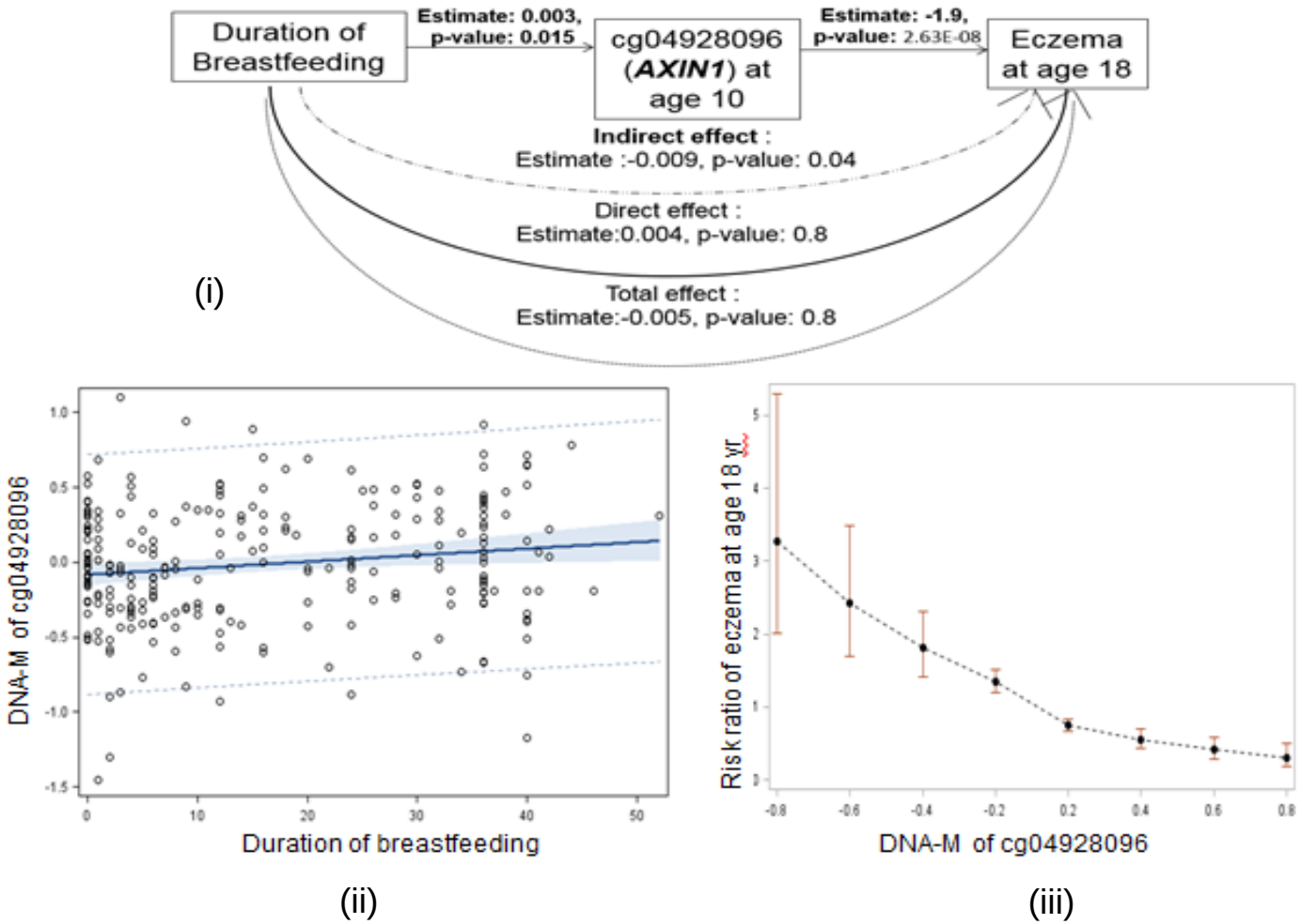


Figure 4. (Continued): Association of cell type residuals of DNA-M with breastfeeding duration and eczema at age 18 years.

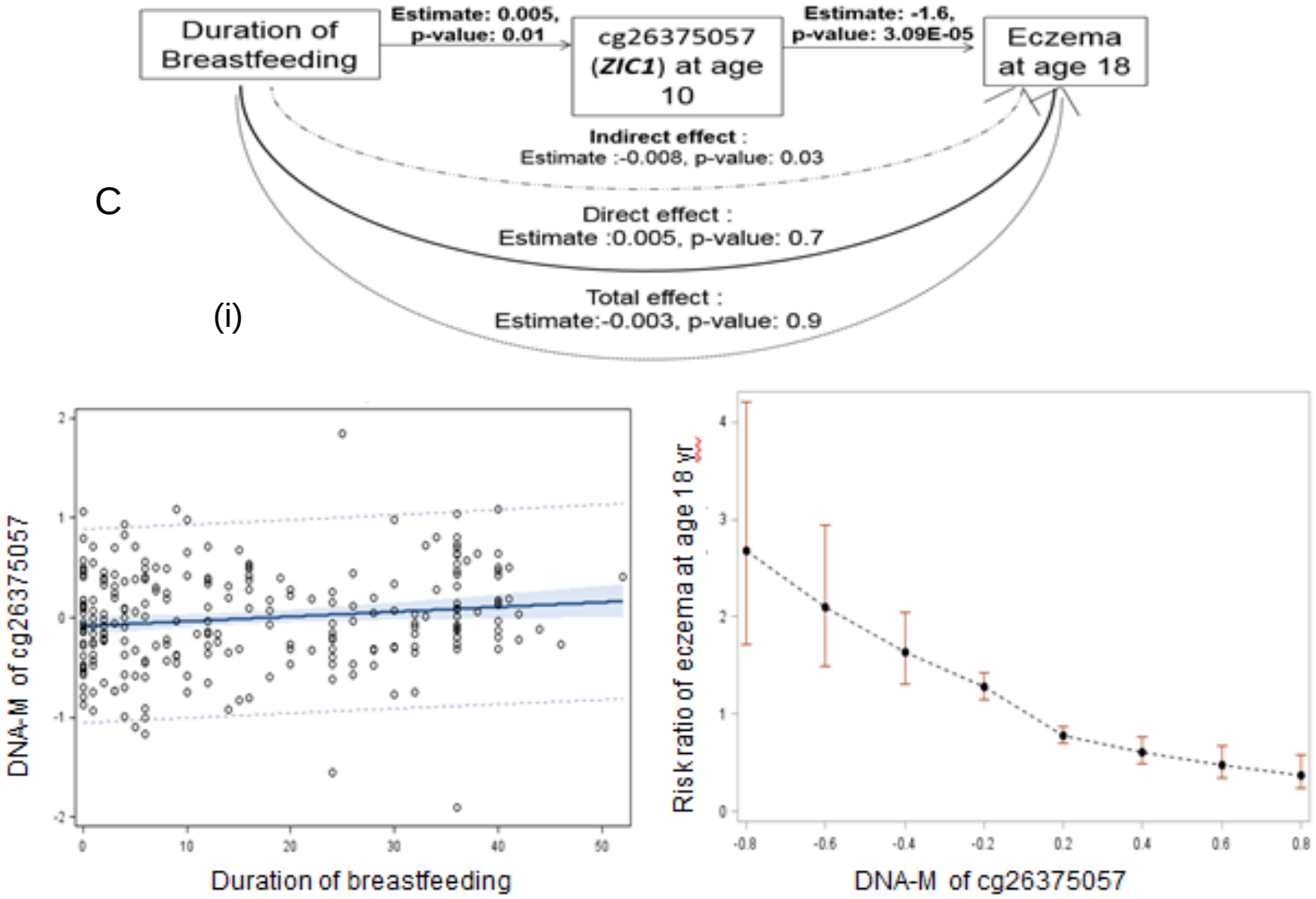


Figure 4 (Continued): Association of cell type residuals of DNA-M with breastfeeding duration and eczema at age 18 years.

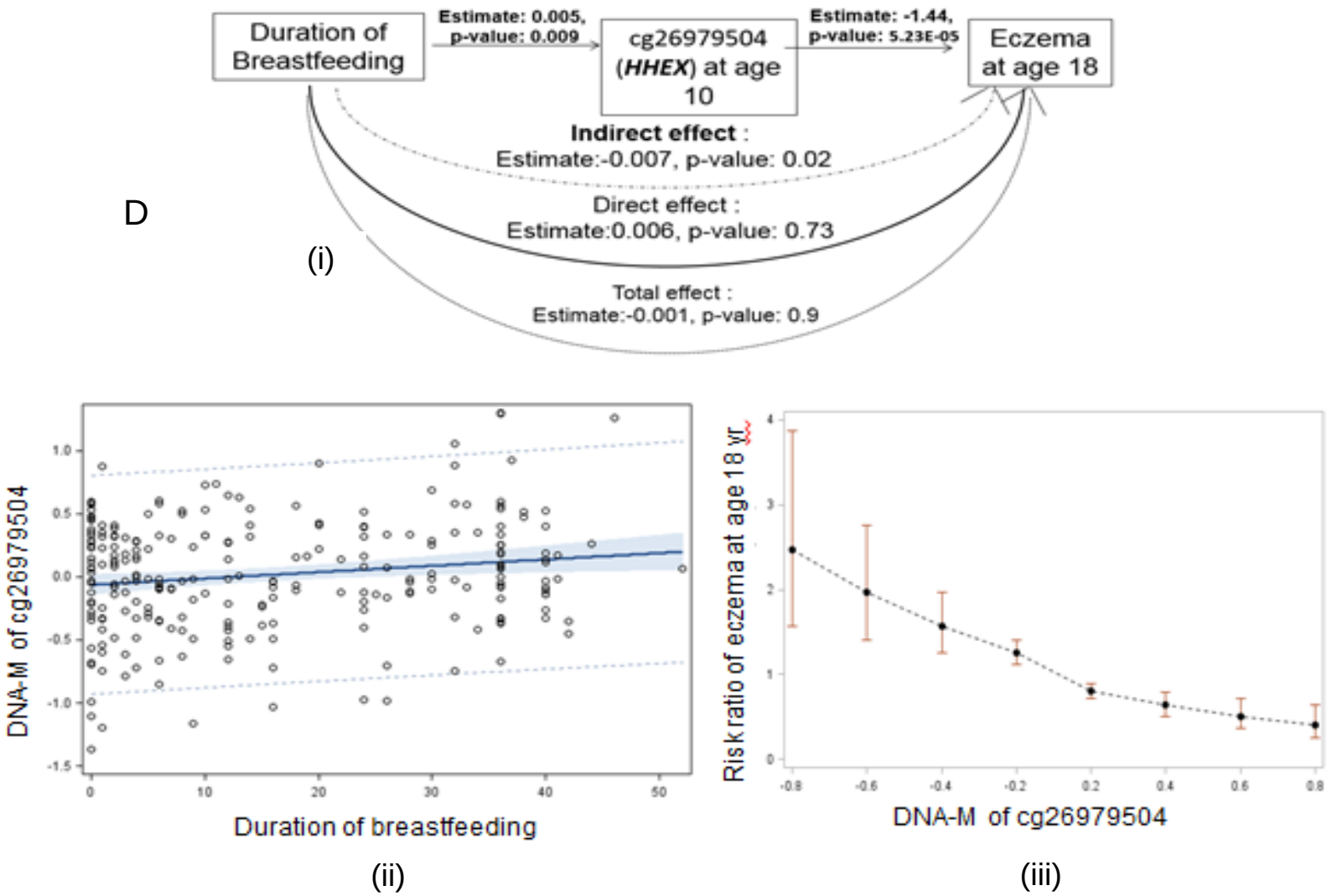


Figure 4 (Continued): Association of cell type residuals of DNA-M with breastfeeding duration and eczema at age 18 years.

In the path analyses in stage (ii), the significant indirect effect indicates that longer duration of breastfeeding showed a protective effect on eczema at age18 (Estimate: -0.008, p-value: 0.022 for cg03605610; Estimate: -0.008, p-value: 0.049 for cg04928096 (*AXIN1*); Estimate: -0.008, p-value: 0.034 for cg26375057 (*ZIC1*); Estimate: -0.007, p-value: 0.021 for cg26979504 (*HHEX*)) (Table 5, Figure 4).

The mean methylations of the four CpGs are presented in Table 6. Whereas cg03605610 is an intergenic CpG between *NRC2* (NK Cell-Activating Receptor) and Long Intergenic Non-Protein Coding RNA 1276, cg04928096 (*AXIN1*) is located on the 3'UTR. The CpG cg26375057 is located in the first exon of the *ZIC1* gene. Longer duration of breastfeeding is associated with a higher methylation of all these CpGs, which are in turn, linked to a lower risk of eczema (Figure 4).

Table 6. Distribution of beta values, M-Values, and cell type residuals of DNA-M (n=276)

Ilmnid	DNA methylation	Mean	Median	Std Dev	5th Pct	95th Pct
cg03605610	Beta value	0.10	0.10	0.03	0.06	0.17
	M value	-3.17	-3.11	0.51	-4.07	-2.31
	Cell type residual	-4.25E-16	0.04	0.50	-0.88	0.80
cg04928096	Beta value	0.90	0.90	0.03	0.85	0.93
	M value	3.18	3.20	0.42	2.49	3.80
	Cell type residual	4.70E-16	-0.01	0.40	-0.66	0.64
cg26375057	Beta value	0.08	0.08	0.03	0.05	0.13
	M value	-3.58	-3.59	0.51	-4.40	-2.72
	Cell type residual	-5.00E-16	-0.01	0.50	-0.80	0.79
cg26979504	Beta value	0.03	0.03	0.01	0.02	0.04
	M value	-5.16	-5.14	0.45	-5.92	-4.51
	Cell type residual	-7.08E-16	-2.51E-03	0.44	-0.74	0.64

To substantiate the role of breastfeeding on these CpGs at age 10 years, we utilized their methylation before initiation of breastfeeding (obtained right after delivery, or within a week from heel sticks) and after breastfeeding (obtained at age 10 years) (n=172). We ran a linear regression with age 10 DNA-M as the outcome and DNA-M obtained from Guthrie cards as the predictor. We found that duration of breastfeeding was significantly associated ($\alpha \leq 0.05$) with DNA-M of cg03605610, cg04928096 (*AXIN1*), and cg26979504 (*HHEX*) and marginally (p-value =0.11) with cg26375057 (*ZIC1*) (Table 7). This indicates that duration of breastfeeding explained the variability of DNA-M at age 10 of these CpGs that is not explained by DNA-M from Guthrie cards. The direction of association was positive for all four CpGs, thus strengthening our findings that longer duration of breastfeeding is linked to higher DNA-M at age 10 years.

Further, to ensure that it is the breastfeeding-associated methylation of these CpGs mediates the effect on eczema, we tested in the F2 generation whether cord blood methylation of these CpGs obtained prior to breast feeding initiation can explain eczema in infancy in the F2 generation (n=192). None of the other CpGs were associated with eczema at infancy (Table 7).

Table 7. Association of duration of breastfeeding with DNA-M in F1 generation and DNA-M and eczema in F2 generation. (**A**) Assessing whether duration of breastfeeding explains variability of DNA-M at age 10 (post breastfeeding) in addition to that of DNA-M prior to breastfeeding from Guthrie cards (collected at birth or within a week of birth) (**B**) Association of cord blood DNA-M
with eczema in infancy in F2 generation. Cell type residuals of DNA-M are used throughout the analyses.

		(A)		**(B)**	
		F1 generation		**F2 generation**	
		Association of DNA-M from Guthrie card (collected at birth or within a week of birth) and duration of breastfeeding with DNA-M at age 10 (n=172)		Association of cord blood methylation with eczema in infancy (n=192)	
Ilmnid	**Parameter**	**Estimate**	**p-value**	**Risk ratios**	**p-value**
cg03605610	Guthrie DNAM	0.3	<.0001	1.02	0.95
	Duration of breastfeeding	0.008	0.002		
cg04928096 (AXIN1)	Guthrie DNAM	-.04	0.66	0.48	0.06
	Duration of breastfeeding	0.006	0.007		
cg26375057 (ZIC1)	Guthrie DNAM	-.03	0.60	0.74	0.36
	Duration of breastfeeding	0.004	0.11		
cg26979504 (HHEX)	Guthrie DNAM	0.04	0.56	1.30	0.31
	Duration of breastfeeding	0.005	0.049		

Discussion

Our results demonstrate that the protective effect of longer breastfeeding duration on eczema at age 18 can be mediated via DNA methylation of four CpGs: cg03605610 (intergenic), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*), cg26979504 (*HHEX*).

Presence of eczema at age 10 years will be associated with eczema at age 18, and also affect DNA-M at age 10 years, driving the association of DNA-M at age 10 years predicting eczema at age 18 years. To rule out this possibility, we performed the analyses focusing on participants without eczema at age 10 years (n=276). To identify potential CpGs at age 10 years that mediate the protective effect of duration of breastfeeding on eczema at age 18 years we performed a screening with several hierarchical steps (stage (i)). The informative CpGs (related to breastfeeding and eczema) that passed the screening step were analyzed using structural equational modelling (stage (ii)) to test mediation. The structural equational modelling focused on these 276 participants showed that the CpGs cg03605610 (intergenic), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*), and cg26979504 (*HHEX*) were in the causal pathway mediating the protective effect of longer breastfeeding duration on eczema at age 18.

There are several limitations in this study. Firstly, the association of duration of breastfeeding with DNA-M at age 10 years for all the four identified CpGs in the pathway between breastfeeding and eczema at age 18 years is small. However, the importance of small changes in methylation of a mixture of white blood cells associated with exposures and complex human diseases are now being recognized [105, 106]. Nevertheless, to investigate these 'small effects', we tested the impact of duration of breastfeeding on methylation in two steps. First, in the F1 generation we utilized methylation levels collected prior to breastfeeding initiation (after

delivery from heel stick blood) and that collected at age 10 years (n=172). We assessed whether duration of breastfeeding can explain additional variability in age 10 DNA-M that is not explained by DNA-M collected soon after birth from Guthrie cards. Second, in the F2 generation we tested whether methylation obtained after delivery from cord blood prior to breastfeeding initiation can explain eczema in the first year of life (n=192).

In the F1 generation, we ran a linear regression separately for each of the four CpGs, with DNA-M at age 10 years as outcome, and DNA-M from Guthrie cards and duration of breastfeeding as predictors. For the CpG cg03605610, DNA-M obtained after birth from Guthrie card explained the variability of DNA-M at age 10 years. Additionally, duration of breastfeeding also explained the variability in DNA-M at age 10 years for cg03605610, such that longer duration was linked to higher methylation. For the other CpGs, DNA-M from Guthrie card was not associated with DNA-M at age 10 years. However, duration of breastfeeding explained the variability of DNA-M at age 10 of these CpGs that is not explained by DNA-M from Guthrie cards. Longer duration of breastfeeding linked to higher methylation of cg04928096 (*AXIN1*), cg26979504 (*HHEX*) (p-value $\leq$0.05), and cg26375057 (*ZIC1*) (p-value$\leq$0.1) (Table 7). These results indicate that methylation changes of these CpGs are indeed related to longer breastfeeding duration. In the F2 generation, cord blood methylation collected prior to breastfeeding initiation of none of the CpGs were associated with eczema, supporting the findings in the F1 generation regarding mediation of the breastfeeding effect via these CpGs on eczema at age 18 (Table 7).

Apart from the above, measurement error is unlikely since the Illumina Infinium HumanMethylation850 beadchip array, is known to have high reliability and reproducibility

[107]. Since information regarding duration of breastfeeding was collected at year one and two, it can be considered reliable with minimal recall bias [77]. Information regarding eczema was acquired from detailed questions during clinical visits. Eczema was defined following the Hanifin and Rajka criteria minimizing misclassification [69, 76].

In addition to the above, we also investigated the pattern of eczema at ages 1-2 years and 4 years in our sample analyzed (n=276, eczema free at age 10 years). There were six participants who had eczema in either age 1-2 or age 4 or both, with remission at age 10 and re-occurrence at age 18 years. It is likely that these participants were misclassified to be eczema free at age 10 years. To avoid any such misclassification of eczema at age 10 years, we excluded these six participants. However, the associations of CpGs with eczema and breastfeeding duration were still significant.

The main strength of the study is clear predictive time order with duration of breastfeeding preceding age 10 methylation followed by the eczema that occurs at age 18 years, ruling out the possibility of reverse causation. To exclude the influence of eczema at age 10 on the association between DNA-M at age 10 years and eczema at age 18, we have considered only eczema free participants at age 10 years (n=276). The CpGs cg03605610 (intergenic), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*), cg26979504 (*HHEX*) were in the causal pathway between breastfeeding duration and eczema at age 18. We also took into account the various cell types that influence DNA-M by using the residual methylation values calculated by regressing the DNA-M on the cell types at all time points in the analyses.

There are only a few studies that have explored the association between breastfeeding and DNA-M [51]. Obermann-Borst et al, reported a decrease in the *LEP* (leptin) gene promoter

at age 1.4 years with increasing duration of breastfeeding categories (-0.06, 95% CI: -1.19, -0.01), however, methylation was ascertained at the same time as breastfeeding, failing to establish a clear time order [93]. In a case-only-study of breast cancer patients, a higher odds ratio of *CDKN2A* promoter methylation status (Yes/No) in breast tumor tissue was found in never breastfed women (OR: 2.75, 95% CI: 1.14, 6.62) in the premenopausal group [95]. Soto-Ramirez, found an interaction of breast duration with the genetic variants on the methylation of the 17q12 region of the chromosome [94]. In animal studies, the promoters of Nyp and *Slc2a4* were differentially methylated in rats fed with breastmilk compared to formula [108, 109].

With respect to eczema, there seveal reports describing assocaiation of methylation of *FLG* [110], *TSLP* and *FCER1G* [111], and *PROZ* and *NEU1* [112] with eczema. However, many of these studies have used concurrent measurements of DNA-M and outcome, making it difficult to distinguish between DNA-M predicting to eczema and DNA-M modified as a result of eczema. We demonstrate for the first time that DNA-M at age 10 years of cg03605610 (intergenic), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*), and cg26979504 (*HHEX*) may mediate the protective effect of a longer duration of breastfeeding on eczema at age 18 years.

Regarding these four CpGs, since higher methylation in the promoter and first exon methylation downregulates to lower gene expression [44, 113]; it implies that longer breastfeeding could be associated with lower expression cg26375057 (*ZIC1*). The CpG cg26979504 (*HHEX*) is in the body region of the gene and CpG cg04928096 (*AXIN1*) is in the 3'UTR region. Although the role of methylation in the body region and 3'UTR is not completely understood [114], body region methylation is known to modify transcriptional efficiency [44], and 3'UTR methylation may also regulate gene expression [115].

The genes of these four CpGs are considered functionally relevant with regard to breastfeeding and eczema in several animal and human studies. The *AXIN1* (Axis Inhibition Protein 1) gene is a part of the Wnt signaling pathway and induces the ubiquination and degradation of cytosolic β-catenin [116]. β-catenin is involved in cell–cell adhesion by interacting with cadherin proteins and its degradation may lead to apoptosis of cells [117]. Ablation of β-catenin in epidermis has been reported to lead to barrier dysfunction indicating its role in epidermal function [118], thus corroborating our findings of higher methylation for the CpG of *AXIN1* associated with lower risk of eczema. No direct association is reported in the literature regarding breastfeeding duration and the *AXIN1* gene. However, it is known that Epidermal Growth Factor (EGF), a ligand of Epidermal Growth Factor Receptor is abundant in breastmilk [119]. Studies have demonstrated that ligand activated EGFR leads to phosphorylation of several downstream proteins leading to transfer of Axin complexed with other proteins (GSK3-APC-Axin complex)[120] to the cell membrane, thus promoting the stabilization of β-catenin[120]. EGF in breastmilk may also inactivate GSK3 (glycogen synthase kinase 3) and stabilize β-catenin [119].

HHEX (Hematopoietically Expressed Homeobox) codes for a transcription factor that is involved in several developmental and differentiation processes including that of the mammary epithelial cells [121]. Puppin *et al* have reported that higher nuclear localization of *HHEX* in lactating tissue compared to normal non-lactating human breast tissue [121]. Kothapalli *et al* found a higher expression of *HHEX i*n the cerebral cortex region of baboons who were fed higher amounts of Docosahexaenoic acid (DHA) than controls [122] indicating a possible link between fatty acids found in breast milk with *HHEX* expression. Additionally, lower expression

of *HHEX* was reported in dermal mesenchymal stem cells in psoriatic skin lesions compared to controls [123] implying its role in inflammatory skin disease.

The *ZIC1* (zinc finger in the cerebellum 1) gene produces a transcriptional activator and is involved in neurogenesis [124], and is a classic marker of brown adipose tissue (BAT) [125] which is present in breast tissue [126]. An in-vitro analysis with mice adipose cells reported that DHA synthesis takes place in the BAT cells [125]. Zic1 is important for maintaining the expression of retinoic acid (RA)-degrading enzyme cyp26a1 in the forebrain of zebrafish, with Zic1 loss of function linked to higher the levels of RA [127]. Interestingly, RAs play a crucial role in the maintenance of epithelial tissues and retinoic acid isomers are also used as drugs to treat several dermatological diseases [128] including eczema [129].

The CpG cg03605610 is an intergenic CpG on chromosome 6 lying between the genes *NRC2* (NK Cell-Activating Receptor) and Long Intergenic Non-Protein Coding RNA 1276. *NRC2*, also known NKp44 is an NK cell activating receptor [130] is implicated in psoriatic inflammation [131]. Lesional skin and peripheral blood of psoriasis patients have a high proportion of innate lymphoid cells expressing this receptor (NKp44 ++ ILC3) which also produce the IL22, a cytokine that triggers epidermal thickening in psoriatic skin [131]. Long non-coding RNAs (lncRNAs) produce transcripts over 200 nucleotides in length, which do not translate in to proteins [132]. They are found encapsulated in extracellular vesicles of breastmilk [132] playing a role in the regulation of gene expression [133] and are linked to actinic dermatosis [134] and psoriasis [135].

Although peripheral blood methylation of the detected genes is relevant with respect to eczema, it remains an open question whether peripheral blood methylation can be used as a

substitute for tissue specific genes. Interestingly, these genes are expressed not only skin, but also several other tissues including in whole blood (www.genecards.com). Additionally, several studies have found similar methylation of specific genes in blood and other organs [136, 137], reinforcing that whole blood methylation can reflect the methylation pattern of specific tissues.

To summarize, the genes of the detected methylation sites are associated with fatty acids, cell–cell adhesion, apoptosis, and retinoic acid signaling pathway, providing an insight regarding their role in eczema pathogenesis linked to breastfeeding exposure. This is the first study to demonstrate that DNA-M potentially mediates the effect of breastfeeding on eczema reinforcing the importance of breastfeeding. Additionally, our results also demonstrate a long-term effect of duration of breastfeeding on DNA-M at age 10. Methylation of these CpGs may be utilized as target based markers that predict eczema and can be modified by longer breastfeeding thus preventing the disease. Future studies are needed to replicate our findings. In addition, there is a need to understand how the breastmilk composition influences the methylation of genes. This information may aid to develop therapeutic/preventative targets for eczema.

CHAPTER 4

INTERACTION OF GESTATIONAL SMOKING AND DURATION OF BREASTFEEDING ON DNA METHYLATION LINKED TO LOWER RISK OF ECZEMA AT AGE 18 YEARS.

Abstract

Rationale: Studies regarding the association of breastfeeding duration and gestational smoking with eczema report contradictory findings. We hypothesize that combined effect of breastfeeding duration and gestational smoking is associated with DNA-M, which is in turn is then linked to eczema.

Methods: Peripheral blood samples were collected at age 10 years. Gestational smoking, breastfeeding duration information, and eczema at age 18 years were obtained during follow ups after birth at 1, 2, and at 18 years, respectively. DNA-M of CpG (cytosine-phosphate-guanine) sites was assessed using the Illumina HumanMethylation450K and 850K BeadChips. In hierarchical screening steps, we identified CpGs at age 10 years associated with interaction of breastfeeding duration and gestational smoking at at p-value of $\alpha \leq 0.05$ (linear regression) and with eczema at age 18 years using log-linear regression followed by multiple testing adjustment (FDR p-value<=0.05). To identify protective CpGs at age 10 years, we selected those that have opposite direction of association with interaction of breastfeeding duration and gestational smoking, and with eczema at age 18 years, respectively, adjusting for relevant confounders. To eliminate reverse causation by eczema at age 10 years influencing the DNA-M at age 10 years we excluded children with eczema at age 10 years (n=276).

Results: The combined effect of gestational smoking and duration of breastfeeding influenced the DNA-M at age 10 years of cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and

cg21601919 (*FGF18*) such that they predict a lower risk of eczema at age 18 years. Conclusion: DNA-M related to combined effect of gestational smoking and breastfeeding duration is linked to a lower risk of eczema, constituting a mediation between these risk factors and eczema.

Background

Eczema is a chronic inflammatory disease characterized by itching and inflammation of the skin [1]. Eczema in early life may progress to other diseases such as food allergy, allergic rhinitis, asthma, known as atopic march [53] leading to increased health care costs and loss of quality of life [7]. Eczema is considered as a Th2 mediated disease [138] and has a complex and multifactorial pathogenesis. It can be genetically inherited and genetic variants of several genes have been associated the disease [139]. The most widely replicated genetic factor is "loss of function (LOF)" mutation of the filaggrin (FLG) gene [10, 52] which is responsible for aggregating keratin filaments and maintain the barrier function of the skin [140]. However, presence of FLG LOF mutations explain less than 10% of the eczema cases [16].

Apart from the genetic factors, environmental factors such as gestational smoking and breastfeeding duration may also play a role in immune system dysregulation and epidermal barrier dysfunction leading to eczema. In a systematic review Kramer et al states that exclusive breastfeeding for at least 3 months reduces the risk of atopic dermatitis in infancy [19]. A randomized trial comparing an intervention with prolonged and exclusive breastfeeding, reported that there was a 54% lower risk of flexural eczema at age 16 years in the intervention group compared to those who received usual care [21]. Chiu et al found breastfeeding for more than 6 months was associated with a significantly lower risk of developing eczema at ages 1 and 2 years [22].

Some studies have also reported a protective effect of maternal smoking on eczema. A large study based in Denmark showed lower odds of atopic eczema at ages 14-18 years if the mother smoked during late pregnancy [27]. Another large cohort study found reduced odds of eczema at 5 years of age in children whose mothers smoked during pregnancy [28]. Ek et al, reported that maternal smoking around birth is associated with a lower odds of being diagnosed with eczema in a large study of UK biobank data in participants ranging from ages 37 to 73 years [29]. Interestingly, coal tar which has several components similar to tobacco smoke, has been used as a topical remedy of eczema since ancient times [32]. Experimental studies have shown coal tar improves barrier function by activating the activating the AhR/ARNT (Aryl hydrocarbon receptor/ Aryl Hydrocarbon Receptor Nuclear Translocator) genes [33] which are also responsible metabolizing cigarette smoke compounds [141].

Although the above studies have found breastfeeding and gestational smoking to be protective of eczema, it is not yet known how breastfeeding and gestational smoking mediate their effects. One possible mechanism by which breastfeeding may influence eczema is by altering DNA methylation (DNA-M), an epigenetic mechanism that can change the gene expression without mutation [40]. Findings suggest that the methylation of the 5′ carbon of cytosine nucleotides at cytosine-phosphate-guanine (CpG) sites in the DNA is affected by environmental exposures [41] and that early life exposures can have a long-term effect on DNA-M [42]. A recent meta-analysis reported five human studies [51] linking breastfeeding duration to methylation of the chromosome region 17q12, LEP (LEPTIN) gene, and also influencing global methylation patterns [51]. Interestingly, regarding smoking multiple studies have consistently reported methylation of the AHRR gene to be associated with maternal smoking during pregnancy [47, 49, 142]. In a genome-wide Consortium Meta-analysis across 13 cohorts,

several CpGs belonging to the biological processes such as nervous system development, phosphate-containing compound metabolism, and cell to cell communication have been identified in cord blood linked to maternal smoking during pregnancy [50].

Reports linking breastfeeding duration and gestational smoking to eczema have conflicting results. If the mother smokes during pregnancy the offspring is exposed to smoke in utero, and additionally components of smoke may be stored in the breast fat tissues and passed on the child during breastfeeding [30, 31]. Till date, there has been no study investigating the interplay of gestational smoking, breastfeeding duration, and DNA-M on eczema. We hypothesize that gestational smoking and breastfeeding duration can jointly influence DNA-M of CpGs such that they explain a lower risk of eczema. To this end, we analyzed data from the Isle of Wight cohort, U.K. to identify methylation sites at age 10 years in peripheral blood that are associated with combined effect of duration of breastfeeding and smoking during pregnancy. Further, to extend the above time-order of risk and effects, we assess whether these CpGs at age 10 years are associated with eczema at age 18 years. We used a hierarchical screening approach in participants who were eczema-free at age 10 years to identify whether the above CpGs are linked to maternal smoking, duration of breastfeeding and eczema at age 18 years.

Methods

Study population

A population based birth cohort (n=1,536) was established in 1989 on the Isle of Wight, UK, to prospectively study natural history of allergic conditions (F1 generation). From January 1989 to February 1990 1,536 F1-children were born on the Isle of Wight (IOW), among which, 1,456 F0-mother-F1-child pairs were enrolled into the cohort study after exclusion of perinatal

deaths, adoptions, and refusals. Children were followed up at 1, 2, 4, 10 and 18 years of age. [96] These F1 participants themselves became pregnant or fathered a child (F2-offspring), which was enrolled in a follow-up study up.

The local research ethics committee (South Central - Hampshire B Research Ethics Committee) and the Institutional Review Board of the University of Memphis approved the study. Informed written parental or child's consent (at age 18 years) were obtained for all participants at recruitment and each follow-up. The IOW birth cohort has been described in detail elsewhere [64-66].

Phenotypes

Information on eczema symptoms was acquired though detailed interviews and examinations at the 10 and 18 year follow up. Eczema was defined according to the Hanifin and Rajka criteria [68] as chronic or chronically relapsing, itchy dermatitis lasting more than 6 weeks with characteristic morphology and distribution [69]. Regarding breastfeeding, the mother was asked at the 1- and 2-year follow-up of the F1 generation whether the child was breastfed and the total duration of breastfeeding (in weeks). Detailed questionnaires were completed for each child at each follow-up. A postal or telephone questionnaire was sent if a visit was not possible [67].

Socioeconomic status, maternal and paternal eczema, *FLG* loss of function genetic variants, maternal smoking, and gender were included as confounders. Parental occupation was reported at birth, number of children in the index child's bedroom was collected at age 4 years, and family income was ascertained at age 10 years. These three variables were clustered to

produce a family social status variable used in our analyses [70]. Gender, maternal smoking, and maternal and paternal eczema were determined at birth.

Genotypes

FLG genetic variants were genotyped from the DNA samples were extracted from 1,150 cohort participants from blood or saliva and assessed using GoldenGate Genotyping Assays (Illumina, Inc, SanDiego, CA) on the beadXpressVeracode platform (Illumina, Inc, SanDiego, CA) per Illumina's protocol. Data were analyzed using the genotyping module of the GenomeStudio Software package (Illumina, Inc, SanDiego, CA). Individuals carrying the minor allele for at least one of the *FLG* variants R501X, 2282del4, or S3247X were classified as having filaggrin haploinsufficiency [15].

DNA methylation analysis

DNA was extracted from whole blood using a standard salting out procedure, from a subsample of 333 10–year old samples (F1). Additionally, in the F1 generation DNA was isolated from dried blood spots on Guthrie cards of 724 neonates using a method based on the procedure described by Beyan *et al* [97]. In the F2 generation, DNA was extracted from cord blood from subsample of 193 samples. DNA concentration was determined by Qubit quantitation. One microgram of DNA was bisulfite-treated for cytosine to thymine conversion using the EZ 96-DNA methylation kit (Zymo Research, CA, USA), following the manufacturer's standard protocol.

Epigenome-scale DNA methylation was assessed using the Illumina Infinium HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA, USA), which interrogates >484,000 CpG sites, and the Illumina Infinium MethylationEPIC BeadChip (Illumina, Inc., San

Diego, CA, USA), which interrogates > 850,000 CpGs associated with over 24,000 genes. The quality control and data preprocessing steps are described below:

F1 generation

Guthrie card methylation data

There are 8 total batches for the methylation data from F1 generation – two from 450k and rest are from EPIC array platform. CPACOR pipeline [98] has been used for quality control (QC) and pre-processing the quantile normalized beta values from the samples. Due to varying feature numbers in different batches in EPIC arrays, we pre-processed them separately and combined them with pre-processed samples from 450k array. These groups have been combined and only the shared probes between EPIC and 450k arrays were chosen. ComBat [99] has been used to remove the batch effect in combined dataset. Finally, we have 302700 number of sites from 724 number of samples.

Age 10 methylation data

For 10 year samples, quality control was undertaken and quantile normalized beta values were processed following the CPACOR pipeline[98] from total of 7 batches – two from 450k and five are from EPIC array platform. These samples from different array platforms and batches were combined and only the shared probes between EPIC and 450k arrays were only chosen. ComBat [99] has been used to remove the batch effect in combined dataset. Finally, we have 349455 number of sites from 330 samples.

Cell Type proportions: For Guthrie card methylation data, white blood cell counts were generated from the three original groups (EPIC samples with 2 different chip design and 450k samples), and then combined together. The *estimateCellCounts()* function from Minfi package

[100] has been used, using reference panel from Bakulski *et al.* 2016 [101], to generate cell proportion. For 10 year old samples (F1 generation) white blood cell counts were generated originally in the separate platform batches (450k and EPIC) and then combined. This method used the *estimatecellcounts()* function in minfi with the default settings using reference panel from Houseman *et al.* 2012 [102].

Statistical Analyses

Methylation levels for each CpG site were recorded as beta (β) values, which represent the proportion of methylated (M) over methylated (M) plus unmethylated (U) probes (β=M/[c+M+U], with constant c being a constant to prevent diving by zero. M-values or logit-transformed β values (M-values, approximated by log2(β / (1-β)) were calculated since β values close to 0 or 1 tend to suffer from severe heteroscedasticity [103]. CpG sites with probe-SNPs within ten base pairs and with minor allele frequency (MAF) greater than 0.007 were excluded from all analyses.

We performed a hierarchical analysis to identify CpGs linked to interaction of duration of breastfeeding and maternal smoking during pregnancy and with eczema at age 18 years (Figure 5). We excluded all participants that have eczema at age 10 years to avoid its influence on DNA-M at age 10 years and eczema at age 18 years. Thus, we have a clear predictive forward time-order with exposures (maternal smoking during pregnancy and breastfeeding duration) predicting DNA-M at age 10 years and DNA-M predicting eczema at age18 years. Since cell types can influence DNA-M, we have used the residual methylation of all CpGs after regressing the DNA-M on cell type proportions throughout in the analyses.

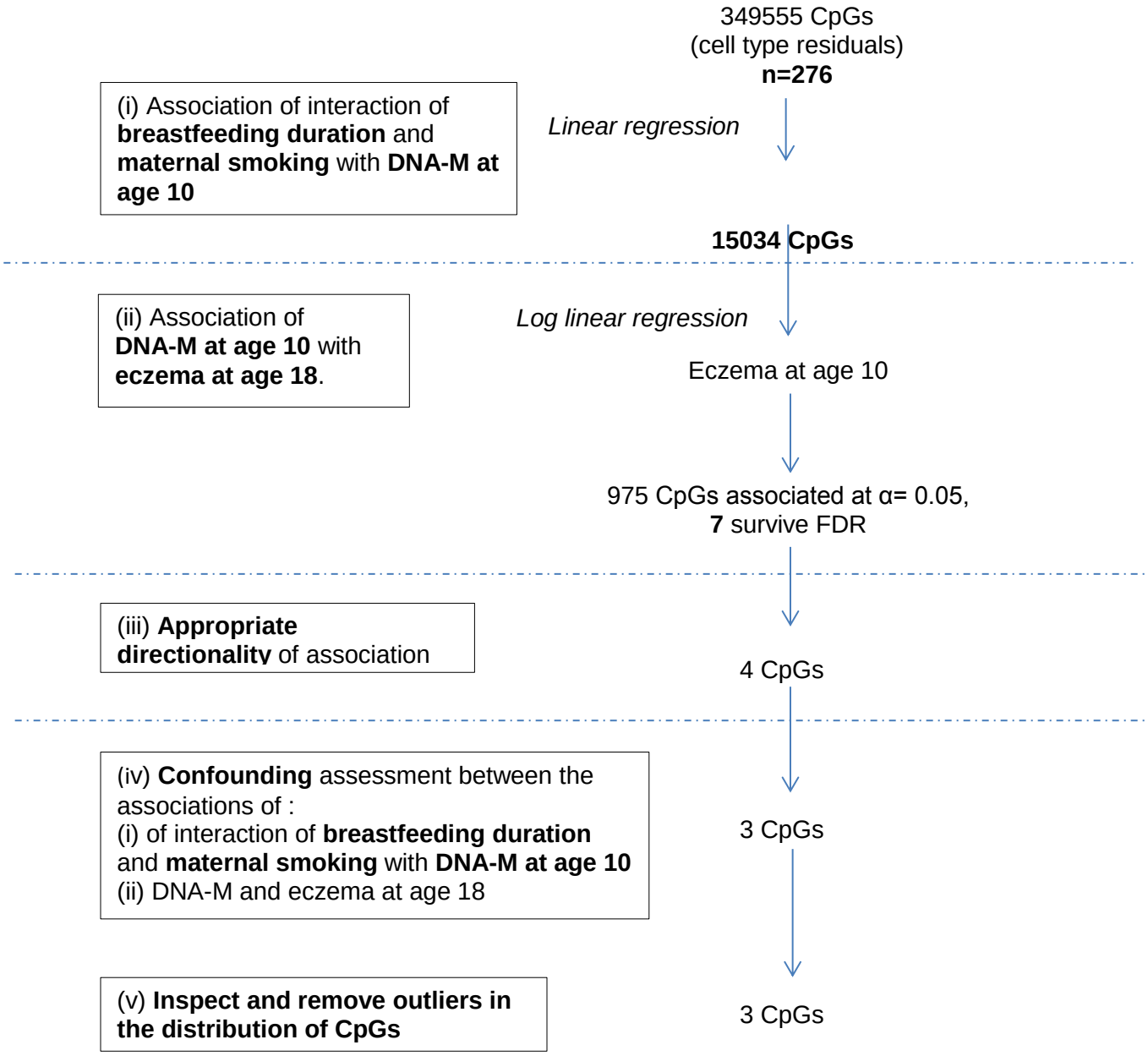


Figure 5 Flow of analyses to identify CpGs at age 10 years associated with longer duration of breastfeeding if the mother smokes during pregnancy and eczema at age 18 years.

The hierarchical steps of analyses are described in detail below (Figure 5):

In step (i) to identify CpGs associated with interaction of maternal smoking and duration of breastfeeding in eczema free children at age 10 years (n=276) we performed a genome-wide screening with 349,455 CpGs at age 10 years as outcome and with main effects and their interaction of gestational smoking and duration of breastfeeding as predictors. We selected the CpGs at age 10 years that had a statistically significant ($\alpha \leq 0.05$) interaction term, i.e., the CpGs show a significant difference in methylation per week increase of duration of breastfeeding if the mother smoked during pregnancy compared to if she did not smoke. In the next step (ii), we tested the selected CpGs for association with eczema at age 18 years using a log-linear model adjusting for multiple testing (FDR p-value ≤ 0.05). The CpGs that survive multiple testing adjustments (false discovery (FDR) p-value ≤ 0.05) in step (ii) were assessed for appropriate directionality in step (iii). We selected CpG that show an opposite direction of associations with interaction of gestational smoking and duration of breastfeeding on the one hand and eczema on the other hand. In other words, a CpG can be protective under two situations:

(a) CpGs whose methylation at age 10 years is increased with increasing duration of breastfeeding if the mother smoked during pregnancy, and this higher methylation is associated with a reduced risk of eczema at age 18 years; or

(b) CpGs whose methylation at age 10 years is reduced with increasing duration of breastfeeding if the mother smoked during pregnancy, and this lower methylation is associated with a reduced risk of eczema at age 18 years.

In step (iii) we select the CpGs with direction of associations as (a) or (b) above. The associations of these selected CpGs were more stringently assessed by adding confounders in step (iv). Potential variables that may confound the association between interaction of gestational

smoking and breastfeeding duration with DNA-M were added in the linear regression models separately for each of the selected CpGs. Similarly, potential confounders were added in the log-linear regression models assessing the association between DNA-M and eczema at age 18 separately for each CpG. Several genetic and clinical confounders, such as maternal and paternal history of eczema, birth order of the child, filaggrin (*FLG*) loss of function genetic variants, socioeconomic status, and maternal smoking were checked. Confounders that change the observed effect sizes in regression models by more than 10% were retained in the model. The CpGs that were still significantly associated after retaining such confounders were then examined for presence of outliers in their distributions, which could influence the associations (step (v)) by checking the regression fit plots. The CpGs that remain significantly associated with the interaction term and the outcome were finally selected.

Additionally, we assessed the DNA-M of the selected CpGs at birth/within a week of birth (from Guthrie cards, n=724). When the blood is collected on Guthrie cards, the offspring has been exposed to the smoke *in utero* and may have had very few days of breastfeeding. We investigated whether Guthrie DNA-M are associated with gestational smoking and whether methylation from Guthrie cards can predict eczema at age 18 years prior to breastfeeding effects.

We investigated whether the combined effect of both gestational smoking and duration of breastfeeding contributes to the variation in the DNA-M at age10 years other than the variation explained by DNA-M from Guthrie cards. To this end, using methylation data at both time points in 172 participants we ran a linear regression with DNA-M at age 10 as outcome and the interaction of gestational smoking and duration of breastfeeding and DNA-M from Guthrie cards as predictors.

Gestational smoking and duration of breastfeeding are interdependent variables, because women who smoke during gestation initiate breastfeeding later [143] and wean earlier [144]. For this reason, we regressed duration of breastfeeding on gestational smoking and calculated the residual duration of breastfeeding. Thereafter, we checked whether the interaction of gestational smoking and residual duration of breastfeeding is associated with the selected CpG sites.

Results

There were no significant differences in the proportion of eczema prevalence, duration of breastfeeding, gender, maternal smoking during pregnancy, socioeconomic status, filaggrin loss of function mutation, and maternal and paternal eczema assessed during pregnancy between the whole cohort (N=1294) and those with DNA-M information at age 10 (n=276) (Table 8). Since eczema at age 10 years may concurrently influence DNA-M at age 10 years and also be linked to future eczema at age 18 years, we included only the participants, who were eczema-free at age 10 years (n=276). In the first step (i), 349,555 whole genome CpGs at age 10 years (Figure 5) were tested in a linear regression model with main effects of maternal smoking and duration of breastfeeding and their interaction term as predictors. 15,034 CpGs were significantly associated with the interaction term, which were then tested for association with eczema at age 18 in step (ii). In this step, we found 975 CpGs to be associated with eczema, of which seven survive multiple testing. In the next step (iii), we selected protective CpGs showing opposite association with the risk factors and eczema. Four CpGs fulfilled the above mentioned directionality conditions (Figure 5). These four CpGs were further assessed with confounders in step (iv) where gender, socioeconomic status, maternal smoking during pregnancy, and maternal and paternal eczema were added as potential confounders in linear regression assessing the

association between interaction of maternal smoking and breastfeeding duration and DNA-M. For the association of DNA-M and eczema at age 18, all the above confounders and *FLG* loss of function and birth order were added in the log-linear model. We retained the confounders that change the observed effect sizes in regression models by more than 10%. The CpGs that remained significant after retain such confounders were checked for presence of outliers which were excluded if present in step (v). On excluding the outliers, one the CpGs was no longer significantly associated with interaction of maternal smoking and duration of breastfeeding and was no longer assessed. The three CpGs: cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and cg21601919 (*FGF18*) still remained significantly associated with both interaction of maternal smoking during pregnancy and duration of breastfeeding and with eczema, respectively (Table 9 and 10).

Table 8. Comparison of population characteristics between the whole cohort and those with DNA methylation at age 10 years.

Covariates	**Whole cohort** **N =1294**	**Sample with DNA-M at age 10 years** **n=276**	
	N (%)	n (%)	*p*-val
Maternal smoking during pregnancy			0.25
Yes	300 (23.2)	55 (19.9)	
No	989 (76.4)	219 (79.3)	
Missing	5 (0.39)	2 (0.7)	
Eczema at age 18			0.07
Yes	157 (12.1)	23 (8.3)	
No	1131 (87.4)	253 (91.6)	
Missing	6 (0.46)		
Sex			0.06
Boys	644 (49.7)	155 (56.2)	
Girls	650 (50.2)	121 (43.8)	
Socio economic status			0.97
High	102 (7.9)	24 (8.7)	
Medium	937 (72.4)	209 (75.7)	
Low	176 (13.6)	41 (14.8)	
Missing	79 (6.1)	2 (0.7)	
Filaggrin loss of function variants			0.43
Yes	107 (8.3)	22 (8.0)	
No	968 (74.8)	241 (87.3)	
Missing	219 (16.9)	13 (4.7)	
Maternal eczema during pregnancy			0.84
Yes	160 (12.4)	33 (12.0)	
No	1124 (86.9)	241 (87.3)	
Missing	10 (0.8)	2 (0.7)	
Paternal eczema during pregnancy			0.31
Yes	90 (6.9)	24 (8.7)	
No	1186 (91.6)	249 (90.2)	
Missing	18 (1.4)	3 (1.1)	
Birth Order			0.30
1	469 (36.2)	105 (38.04)	
2	362 (30)	80 (30)	
3	262 (20.2)	71 (25.7)	
Missing	201 (15.5)	20 (7.25)	
Median (p5, p95)			
Duration of breastfeeding (weeks)	8.0 (0, 40)	12.0 (0, 40)	0.06

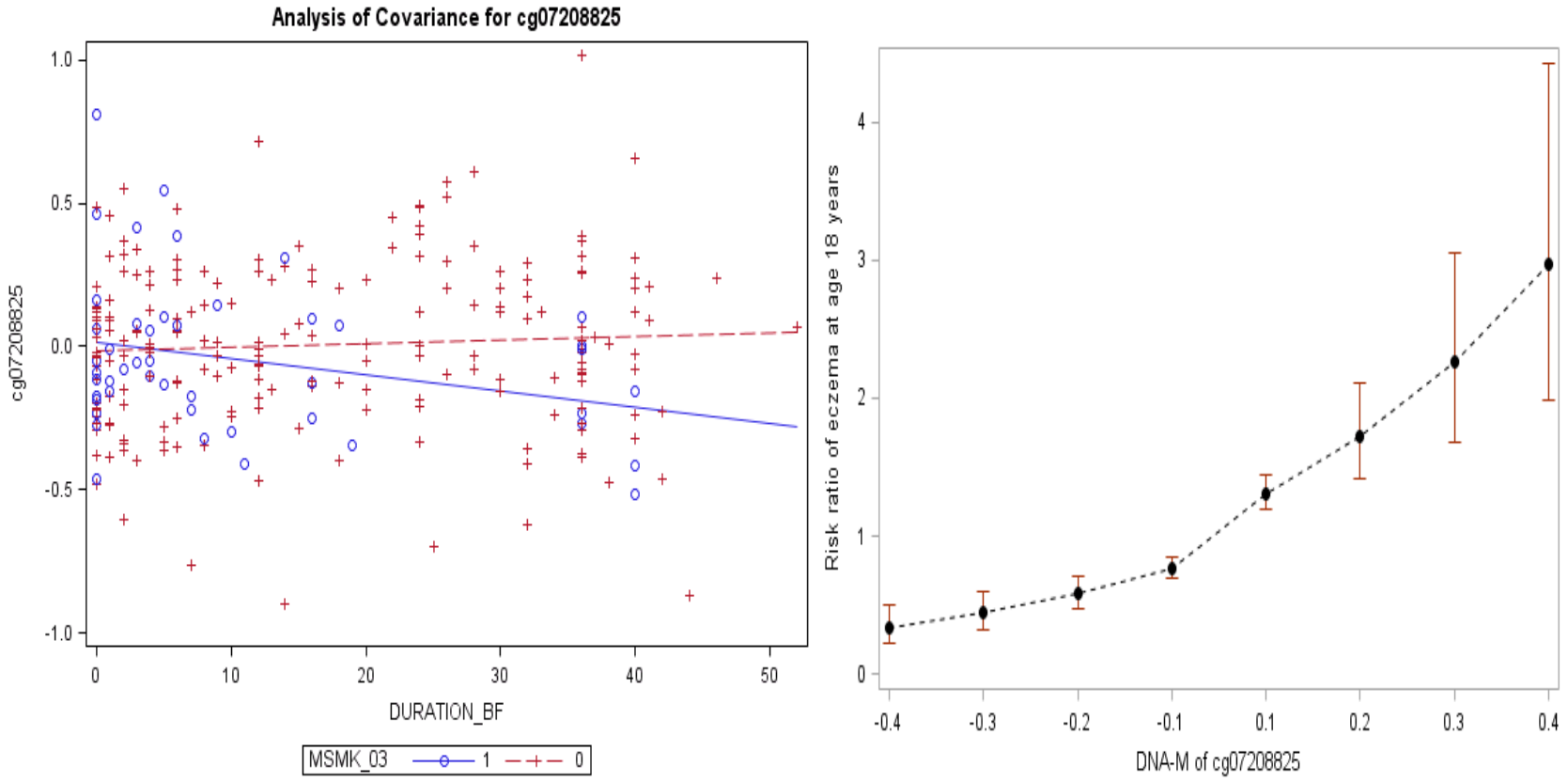
Table 9. Association of maternal smoking, duration of breastfeeding and their interaction with methylation at age 10 years (n=276) of cg07208825 (*ALMS1P*), cg21601919 (*FGF18*), and cg12954512 (Intergenic). Cell type residuals of methylation are used in the analyses.

		cg07208825 (*ALMS1P*)		**cg21601919** (*FGF18*)		**cg12954512** (Intergenic)	
Parameter	**Levels**	**Estimate**	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**
Maternal smoking during pregnancy		0.031	0.62	-0.117	0.13	-0.16	0.41
Duration of breastfeeding (weeks)		0.001	0.35	-0.002	0.19	-0.001	0.75
Maternal smoking during pregnancy × Duration of breastfeeding (weeks)	Overall	-0.007	0.046	0.01	0.02	0.022	0.04
	0 weeks	0.03	0.62	-0.11	0.13	-0.16	0.4
	4 weeks	0.003	0.95	-0.07	0.25	-0.07	0.67
	12 weeks	-0.05	0.26	0.007	0.9	0.11	0.46
	24 weeks	-0.13	0.03	0.13	0.09	0.38	0.05
	36 weeks	-0.21	0.02	0.25	0.03	0.65	0.03

Table 10. Association of methylation at age 10 years (n=276) of cg07208825 (*ALMS1P*), cg21601919 (*FGF18*), and cg12954512 (Intergenic) with eczema at age 18. Cell type residuals of methylation are used in the analyses.

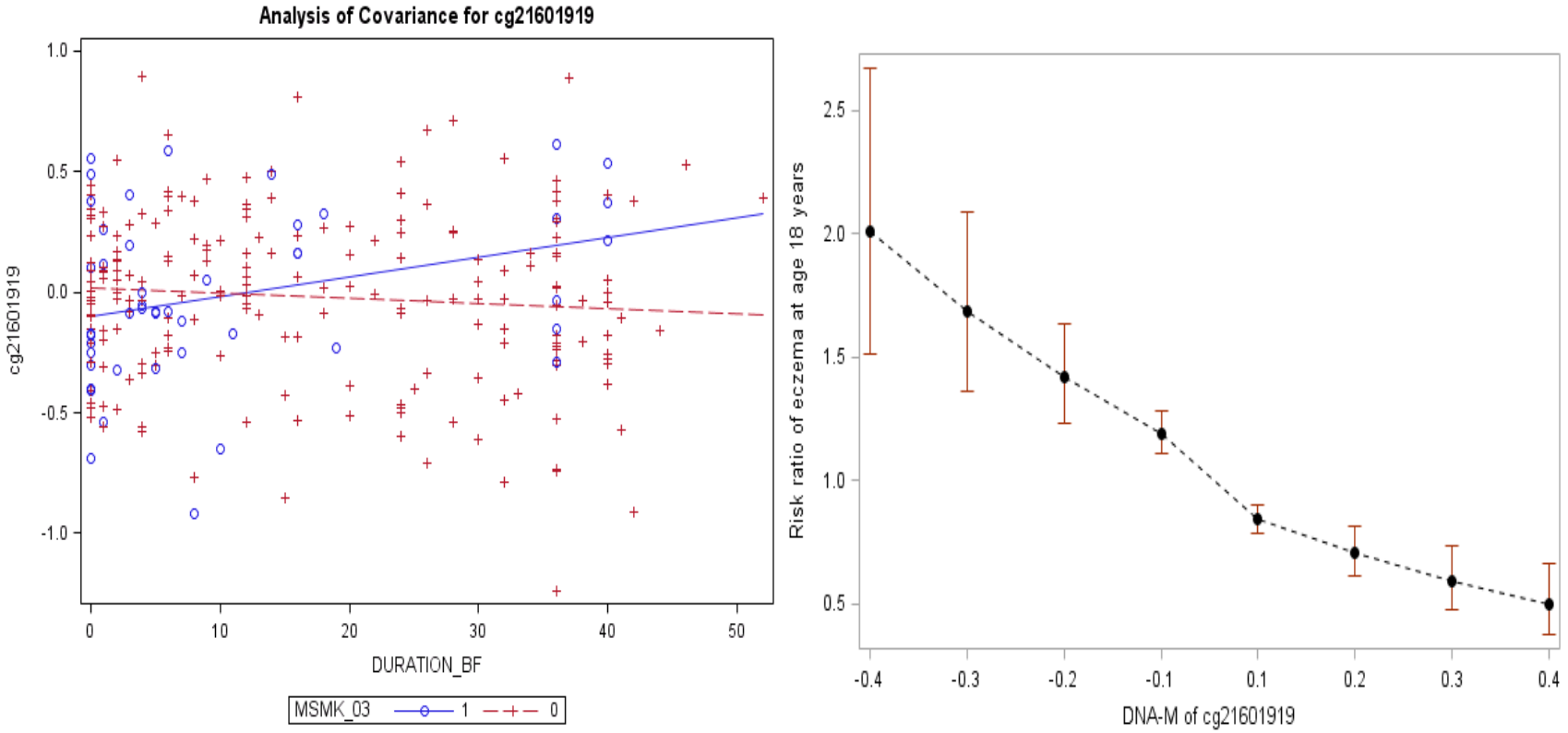
Ilmnid (Gene name)	Levels of DNA-M	Risk ratio	95% CI (Lower)	95% CI (Upper)	p-value
cg07208825 (*ALMS1P*)		14.9	5.5	40.4	1.3E-07
	-0.4	0.3	0.23	0.50	9.42E-08
	-0.3	0.4	0.33	0.60	9.42E-08
	-0.2	0.5	0.48	0.71	9.42E-08
	-0.1	0.8	0.69	0.84	9.42E-08
	0.1	1.3	1.19	1.45	9.42E-08
	0.2	1.7	1.41	2.10	9.42E-08
	0.3	2.3	1.68	3.05	9.42E-08
	0.4	3	1.99	4.43	9.42E-08
cg21601919 (*FGF18*)		0.20	0.09	0.4	1.56E-06
	-0.4	2.0	1.51	2.67	1.52E-06
	-0.3	1.7	1.36	2.09	1.52E-06
	-0.2	1.4	1.23	1.63	1.52E-06
	-0.1	1.2	1.11	1.28	1.52E-06
	0.1	0.8	0.78	0.90	1.52E-06
	0.2	0.7	0.61	0.81	1.52E-06
	0.3	0.6	0.48	0.73	1.52E-06
	0.4	0.5	0.37	0.66	1.52E-06
cg12954512 (*Intergenic*)		0.7	0.45	0.86	3.90E-03
	-2.0	2.6	1.36	4.91	3.60E-03
	-1.5	2.0	1.26	3.30	3.60E-03
	-1.0	1.6	1.17	2.22	3.60E-03
	-0.5	1.3	1.08	1.49	3.60E-03
	0.1	1.0	0.92	0.98	3.60E-03
	0.5	0.8	0.67	0.93	3.60E-03
	1.0	0.6	0.45	0.86	3.60E-03
	1.5	0.5	0.30	0.79	3.60E-03

The methylation of cg07208825 (*ALMS1P*) decreased with increasing duration of breastfeeding if the mother smoked during gestation (Estimate: -0.006, p-value: 0.046 for cg07208825 (*ALMS1P)*). In turn, higher methylation of this CpG was associated with increased risk of eczema at age 18 years (parameter estimate: 14.23, p-value: 1.27E-07 for cg07208825 (*ALMS1P)*) (Tables 9 and 10, Figure 6).



(i) **(A)** (ii)

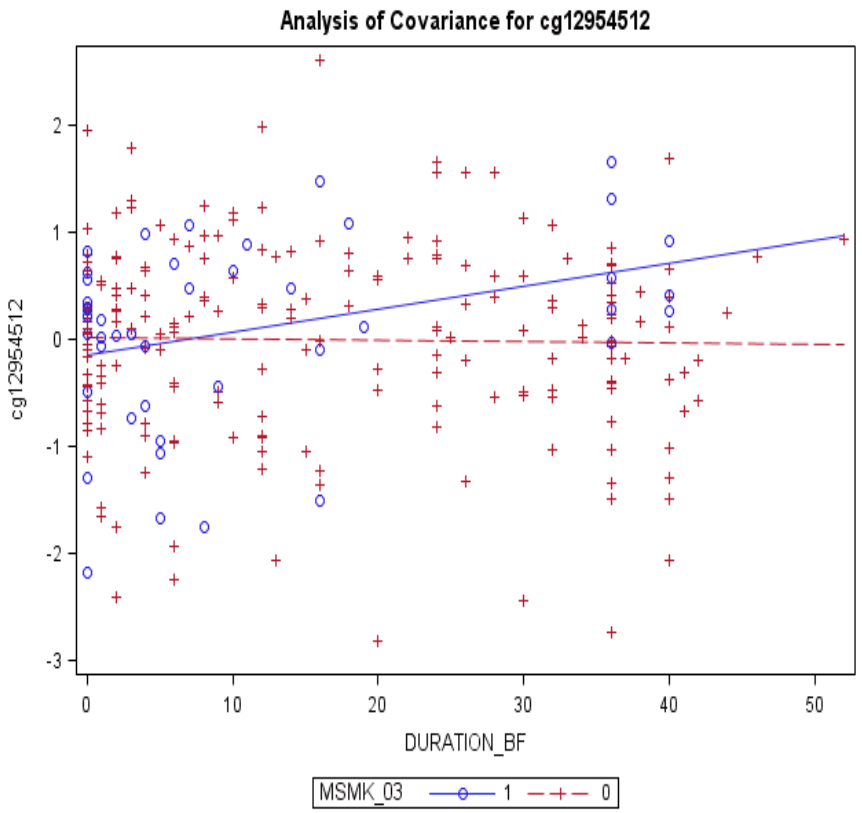
Figure 6. Association of cell type residuals of DNA-M with interaction of maternal smoking and breastfeeding duration, and with eczema at age 18 years (A) cg07208825 (*ALMS1P*) (B) cg21601919 (*FGF18*) (C) cg12954512 (Intergenic). The panels are described as follows: (i) Association of DNA-M at age 10 years with duration of breastfeeding; (ii) Association of DNA-M at age 10 years with eczema at age 18 years in participants with maternal smoke exposure during pregnancy.



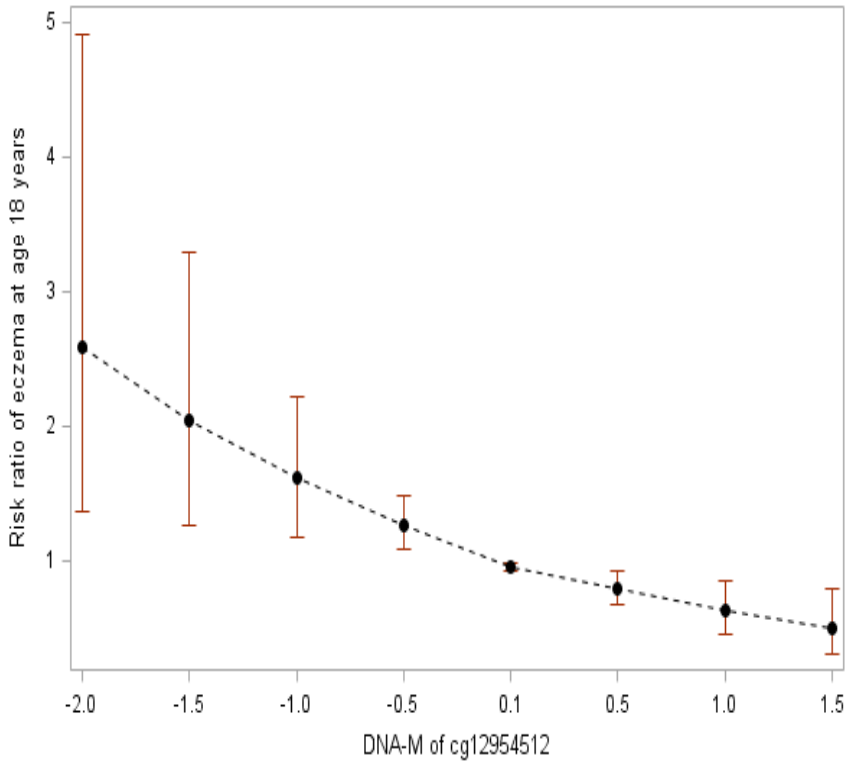
(iii) (iv)

(B)

Figure 6 (continued): Association of cell type residuals of DNA-M with interaction of maternal smoking and breastfeeding duration, and with eczema at age 18 years (A) cg07208825 (*ALMS1P*) (B) cg21601919 (*FGF18*) (C) cg12954512 (Intergenic).



(v)



(C)

(vi)

Figure 6 (continued): Association of cell type residuals of DNA-M with interaction of maternal smoking and breastfeeding duration, and with eczema at age 18 years (A) cg07208825 (*ALMS1P*) (B) cg21601919 (*FGF18*) (C) cg12954512 (Intergenic).

For the other two CpGs, the methylation of cg12954512 (*Intergenic*) and cg21601919 (*FGF18*) increased with increasing duration of breastfeeding if the mother smoked during gestation (Estimate: 0.021, p-value: 0.036 for cg12954512 (*Intergenic*) and parameter estimate: 0.008, p-value: 0.016 for cg21601919 (*FGF18*)). In turn, higher methylation of these CpGs was associated with lower risk of eczema at age 18 years (Risk ratio: 0.62, p-value: 3.90E-03 for cg12954512 (*Intergenic*) and Risk ratio: 0.18, p-value: 1.56E-06 for cg21601919 (*FGF18*)) (Table 9 and 4.3, Figure 6).

The CpG cg07208825 (*ALMS1P*) is located on chromosome 2 on the 200 bp upstream from transcription start site (promoter region), cg21601919 (*FGF18*) is located in the body region of chromosome 5 whereas the CpG cg12954512 is located on chromosome 7 between the genes *WDR60* and *LINC00689*.

Then it was determined whether exposure to gestational smoking alone affected the DNA-M collected from Guthrie cards (prior to establishing a breastfeeding effect) such that they predict a lower risk of eczema at age 18 years. DNA-M of the selected CpGs measured in Guthrie cards (n=724) were not associated with maternal smoking during pregnancy, and did not predict eczema at age 18 years (Data not shown). This shows that these CpGs at birth were not differentially methylated by gestational smoking alone, and could not explain future eczema at age 18 years, without the influence of breastfeeding.

Next, DNA-M at birth (from Guthrie cards) and DNA-M at age 10 years in participants with methylation information at both time points (n=172) were compared. Linear regression were used for each of the three selected CpGs with DNA-M at age 10 as outcome and the interaction of gestational smoking and duration of breastfeeding and controlling for DNA-M

from Guthrie cards as predictors. For the CpG cg21601919 (*FGF18*), DNA-M at birth was not associated with DNA-M at age 10 years. However, interaction of gestational smoking and duration of breastfeeding remained significantly associated with DNA-M at age 10 years (Estimate: 0.01, p-value: 0.045) indicating that their combined effect explains the variation in DNA-M at age 10 years that is not explained by DNA-M at birth (or within a week of birth) of cg21601919 (*FGF18*). Additionally, the direction of association is still the same as we found in our screening step with DNA-M at age 10 years. In other words, longer duration of breastfeeding is associated with an increase in the DNA-M at age 10 years if the mother smoked during pregnancy even after adjusting for DNA-M at birth/within a week of birth of cg21601919 (*FGF18*).

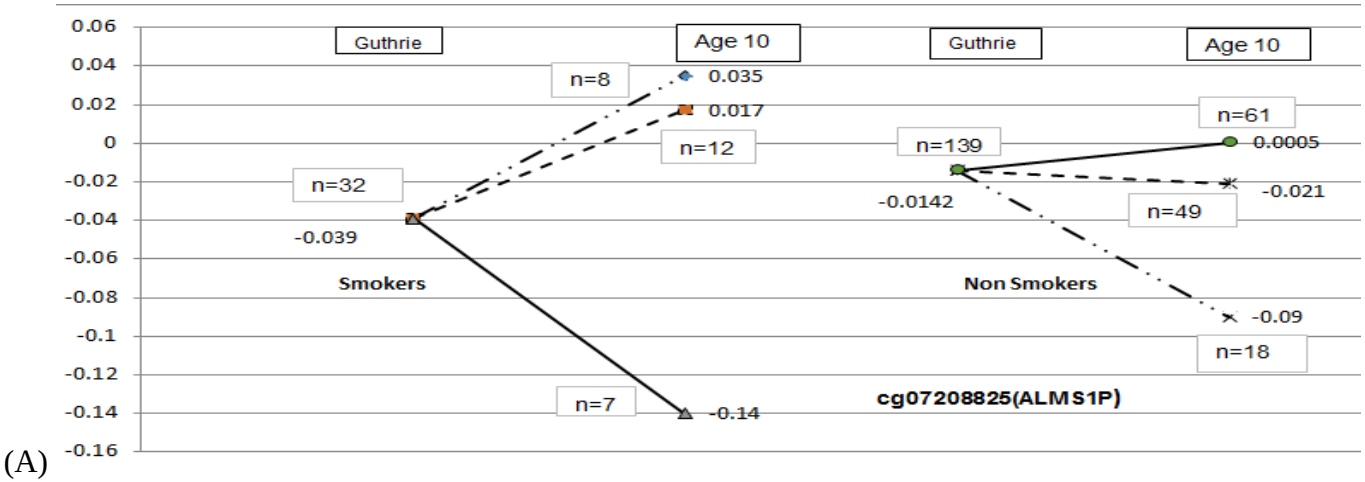
Although for the other two CpGs the interaction term was no longer significant after adjusting for DNA-M at birth/within a week of birth, the direction of the interaction was still in agreement as we found at age 10 years (parameter estimate: -.006, p-value: 0.13 for cg07208825 (*ALMS1P*); parameter estimate: 0.019, p-value: 0.3 for cg12954512 (*Intergenic*)) (Table 11).

Table 11. Association of maternal smoking during pregnancy and duration of breastfeeding with DNA-M at age 10 (post breastfeeding) adjusted for DNA-M from Guthrie cards (collected at birth or within a week of birth)

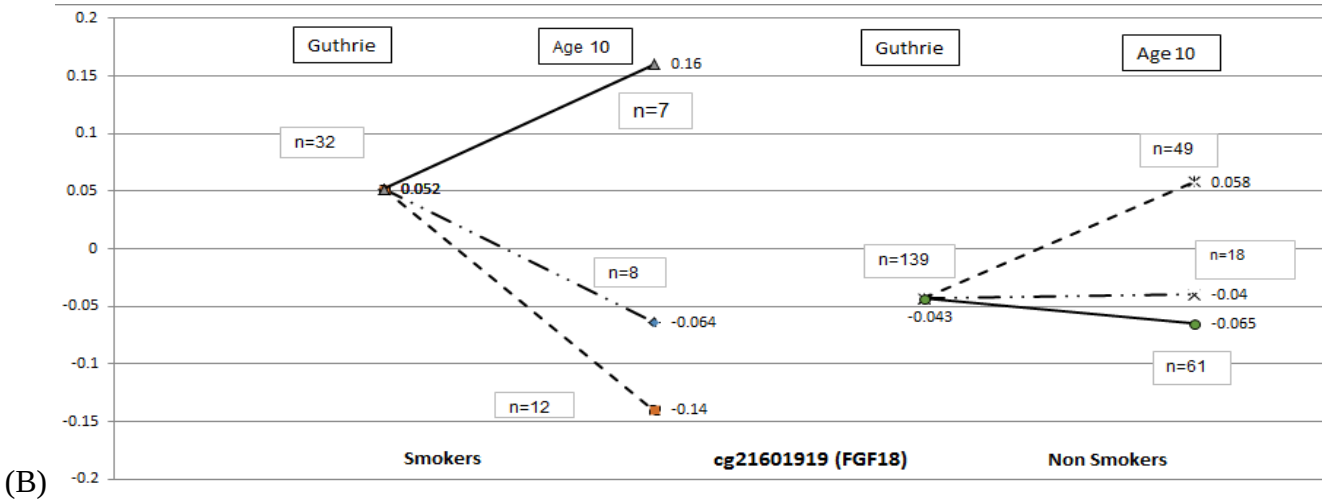
Ilmnid	Parameter	Levels	Estimate	p-value
cg07208825 (*ALMS1P*)	DNAM from Guthrie cards		0.31	0.0003
	Maternal smoking during pregnancy	Yes	0.08	0.23
		No (Ref.)		
	Duration of breastfeeding (weeks)		0.0012	0.45
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)		-0.007	0.13
cg21601919 (*FGF18*)	DNAM from Guthrie cards		-0.11	0.14
	Maternal smoking during pregnancy	Yes	-0.13	0.13
		No (Ref.)		
	Duration of breastfeeding (weeks)		-0.002	0.26
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)		0.011	0.045
cg12954512 (*Intergenic*)	DNAM from Guthrie cards		0.39	<.0001
	Maternal smoking during pregnancy	Yes	-0.68	0.02
		No (Ref.)		
	Duration of breastfeeding (weeks)		-0.0005	0.93
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)		0.02	0.3

The pattern of DNA-M in participants with methylation information at both time points (n=172) is shown in Figure 7. The development of the mean DNA-M in participants of different breastfeeding groups (not breastfed, breastfed for <=3 months, and breastfed for >3 months) were plotted from birth to age 10 years, comparing children with and without maternal smoke exposure during gestation. Figure 7 (a) shows that if mothers smoked during gestation (left graph) the methylation of cg07208825 (*ALMS1P*) is lower at age 10 years compared to that at birth in those who breastfeed for > 3 months and lower DNA-M of this CpG is protective for eczema (Table 10, Figure 6a). However, if the mothers did not smoke during gestation (right graph), the DNA-M is higher at age 10 years than at birth in those who breastfeed for > 3 months. Similarly, the methylation of cg21601919 (*FGF18*) and cg12954512 (*Intergenic*) also show contrasting patterns for longer periods of breastfeeding if the mother smoked during gestation vs. if she did not.

Regarding the two exposures variables, gestational smoking is known to reduce duration of breastfeeding and even delay the initiation of breastfeeding [143, 144]. Hence, duration of breastfeeding may be affected by gestational smoking and thus cannot be an independent risk factor used in the interaction term of gestational smoking and breastfeeding duration. We considered this by 'extracting' the part of the duration of breastfeeding that is due to gestational smoking. To this end duration of breastfeeding was regressed on gestational smoking and the residual duration of breastfeeding were calculated. Taking the interaction of gestational smoking and residual duration of breastfeeding into account, the associations were was still significantly associated in the same direction and similar effect sizes with DNA-M at age 10 years of all three selected CpGs (Table 12).



(A)



(B)

Figure 7. Pattern of DNA-M from Guthrie cards and age 10 years in participants who did not breastfeed, breastfeed for less than three months, and more than three months between smokers and non-smokers (A) cg07208825 (*ALMS1P*) (B) cg21601919 (*FGF18*) (C) cg12954512 (Intergenic).

(C)

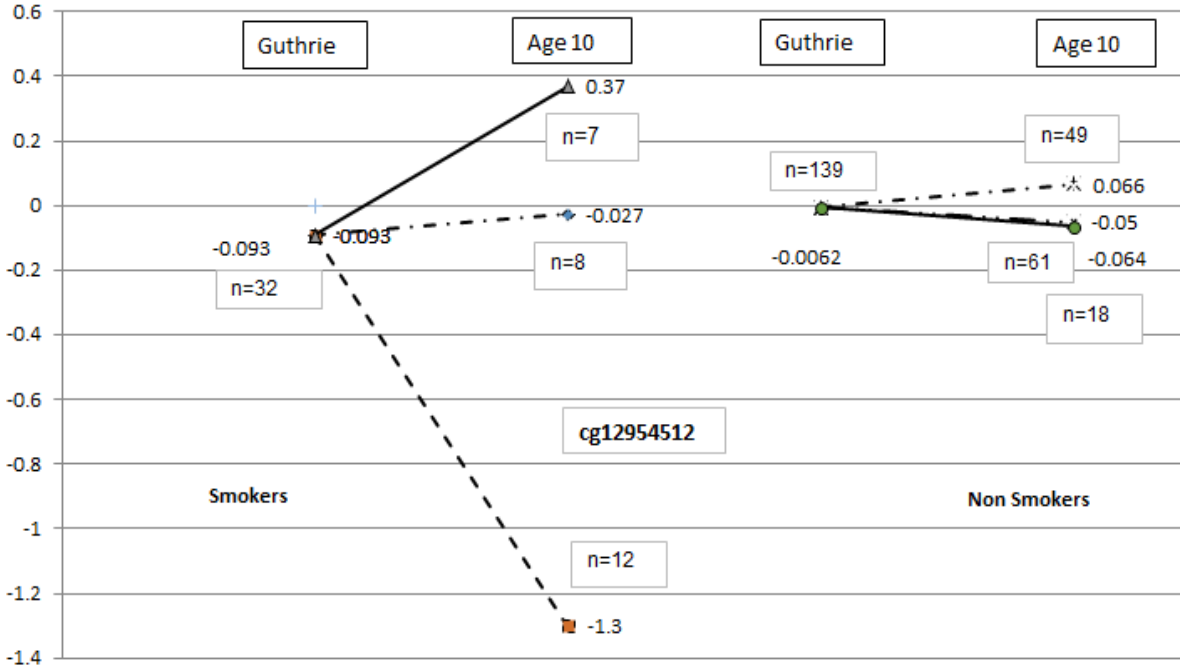


Figure 7 (Continued). Pattern of DNA-M from Guthrie cards and age 10 years in participants who did not breastfeed, breastfeed for less than three months, and more than three months between smokers and non-smokers (A) cg07208825 (*ALMS1P*) (B) cg21601919 (*FGF18*) (C) cg12954512 (Intergenic).

Table 12. Association of gestational smoking and residual duration of breastfeeding with DNA-M at age 10 (n=276). The residual duration of breastfeeding is calculated by regressing the duration of breastfeeding on gestational smoking.

Ilmnid	Parameter	Levels	Estimate	p-value
cg07208825 (*ALMS1P*)	Maternal smoking during pregnancy	Yes	-0.05	0.26
		No (Ref.)		
	Duration of breastfeeding (weeks)		0.0012	0.4
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)	Yes	-0.006	0.045
cg21601919 (*FGF18*)	Maternal smoking during pregnancy	Yes	0.007	0.9
		No (Ref.)		
	Duration of breastfeeding (weeks)		-0.022	0.2
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)	Yes	0.01	0.02
cg12954512 (*Intergenic*)	Maternal smoking during pregnancy	Yes	-0.13	0.5
		No (Ref.)		
	Duration of breastfeeding (weeks)		-0.001	0.8
	Maternal smoking during pregnancy × Duration of breastfeeding (weeks)	Yes	0.03	0.03

Discussion

Our results suggest that combined effects of gestational smoking and duration of breastfeeding influence the DNA-M at age 10 years of cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and cg21601919 (*FGF18*), which in turn predict a lower risk of eczema at age 18 years.

To identify CpGs at age 10 years, which methylation are linked to duration of breastfeeding and maternal smoking during pregnancy and with eczema at age 18 (n=276), we performed a hierarchical screening. We excluded any participants who had eczema at age 10 years to avoid reverse causation of eczema at age 10 years influencing DNA-M at age 10 years and its association with eczema at age 18 years.

We first checked the association of interaction of maternal smoking during pregnancy and duration of breastfeeding with DNA-M of genome wide CpGs at age 10 years. The CpGs significantly associated with the interaction term were tested for association with eczema at age 18 years and corrected for multiple testing using FDR. The CpGs which survived multiple testing were assessed further for appropriate directionality. For instance, we focused on CpGs whose methylation at age 10 years increase with increasing duration of breastfeeding if the mother smoked during pregnancy, and higher methylation was associated with a reduced risk of eczema at age 18 years, or vice versa. We assessed the associations for confounding and removed any outliers in the methylation that could influence the regressions. Finally, the three CpGs: cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and cg21601919 (*FGF18*) still remained significantly associated with both interaction of maternal smoking during pregnancy and duration of breastfeeding and with eczema, respectively (Table 9 and 10).

Additionally, the selected CpGs measured at birth/within a week of birth from Guthrie cards (n=724) were not associated with gestational smoking, and did not predict eczema at age 18 years (data not shown). This suggests that gestational smoking alone is not sufficient to bring about a change in methylation, and methylation at this time point, not yet influenced by breastfeeding, could not explain eczema at age 18 years.

There are several limitations of this study. Firstly, we used a nominal p-value ($\alpha <= 0.05$) to select CpGs at age 10 years associated with interaction of gestational smoking and duration of breastfeeding without adjusting for multiple testing. Although there is a possibility of identifying false positives CpGs, the purpose of this step was to screen for informative CpGs linked to gestational smoking and duration of breastfeeding to reduce number of variables to be tested for association with eczema. However, in the next step we corrected for multiple testing and select only those CpGs which have an FDR p-value <=0.05 for the association between DNA-M at age 10 years and eczema at age 18 years. Further filtering of the CpGs took place in the next step based on the criteria of directionality, confounding, and outliers so that starting with the 349455 CpGs only four are finally selected.

Secondly, the effect size of association of combined effect of gestational smoking and duration of breastfeeding and eczema is small. However, we investigated this small effect size by comparing the DNA-M at birth/ within a week of birth from Guthrie cards (collected post exposure to gestational smoking but prior to breastfeeding) with that at age 10 years (collected post gestational smoking and breastfeeding) in a sample of participants with methylation at both time points (n=172). We investigated whether interaction of maternal smoking and duration of breastfeeding can explain the variation in the DNA-M at age 10 years after adjusting for DNA-M

measured at birth/within a week of birth from Guthrie cards. We found that the interaction term was significantly associated with DNA-M at age 10 years for cg21601919 (*FGF18*) whereas DNA-M from Guthrie cards was not significantly associated with DNA-M at age 10 years (Estimate: -0.11, p-value: 0.14). The DNA-M from Guthrie cards explained the variation in DNA-M at age 10 years for cg07208825 (*ALMS1P*) (parameter estimate: 0.31, p-value: 0.0003) and cg12954512 (*Intergenic*) (Parameter estimate: 0.39, p-value: <0.0001. Although the p-value of the association of interaction term with DNA-M at age 10 is only marginally significant for cg07208825 (*ALMS1P*) and non-significant cg12954512 (*Intergenic*), the direction of the effect of longer duration of breastfeeding if the mother smokes during pregnancy is predictive of lower risk of eczema at age 18 for all three CpGs (Table 11). These findings indicate that the combined effect of gestational smoking and duration of breastfeeding is required influence the DNA-M at age 10 years.

We could not take into account passive smoke exposure after birth, which remains a limitation. However, none of the participants started smoking cigarettes themselves at age 10 years when the DNA-M information was assessed. Additionally, none of the selected CpGs were associated with current smoking at age 18 years, indicating no confounding of the association between DNA-M and eczema at age 18 years.

We additionally examined the pattern of eczema at age 1 and 2, 4, 10 and 18 in the sample analyzed with DNA-M at age 10 years (n=276, eczema free at age 10 years). There were six participants who had eczema in either age 1-2 or age 4 or both, with remission at age 10 and re-occurrence at age 18 years. It is possible that these participants were misclassified to be eczema free at age 10 years or that these children still carry a non-identifiable risk. To avoid any

such misclassification of eczema at age 10 years, in a side-analysis we excluded these six participants. Nevertheless, the associations of CpGs with interaction of maternal smoking and breastfeeding duration and with eczema at age 18 years remained statistically significant.

One of the limitations of this study is that the maternal smoking status is self-reported. However, we have assessed the DNA methylation of *AHRR*, a CpG (cytosine-phosphate-guanine) site, cg05575921, of the *AHRR* gene, a well-established epigenetic biomarker of smoking. The offspring whose mothers smoked during pregnancy have significantly lower methylation of this CpG site from Guthrie cards (n=724, parameter estimate: -0.66 and p-value <0.0001), which is in agreement with the prior studies of prenatal smoke exposure and supports the validity of the smoking information [48].

Regarding DNA-M , a major measurement error is unlikely since the Illumina Infinium HumanMethylation850 beadchip array, is known to be high reliable and high reproducible [107]. Since information regarding duration of breastfeeding was collected at year one and two, it can be considered reliable reducing minimizing the recall bias [77]. Information regarding eczema was acquired from detailed questions during clinical visits. Eczema was defined as 'ever eczema' and having had an 'itchy rash during the previous 12 months' and was defined following the Hanifin and Rajka criteria [69]. The chances of misclassification is minor since the participants developed rashes developed on typical locations (anticubital or popliteal fossae, ankles, face or neck) [76].

The major strength of this study is the forward time-order with maternal smoking and breastfeeding predicting DNA-M at age 10 years of cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and cg21601919 (*FGF18*), which in turn predict a lower risk of eczema 18 years.

Although a link between maternal smoking during pregnancy and DNA-M has been investigated in various studies [47, 49, 50, 142], there are very few studies on breastfeeding duration and DNA-M in offspring [51]. Associations of eczema and DNA-M of *FLG* [110], *TSLP* and *FCER1G* [111], and *PROZ* and *NEU1* [112] has been reported. However, in most of these studies DNA-M was measured concurrently with the eczema, making it difficult to distinguish on whether DNA-M results as cause or consequence of the disease-status. No other study has yet investigated the combined effect of breastfeeding duration and gestational smoking on DNA-M, and whether the altered DNA-M can predict eczema.

The genes of the above-identified CpGs are functionally relevant with regard to epidermal differentiation. The CpG cg21601919 is located in the body region of the gene *FGF18* (*Fibroblast Growth Factor 18).* Although the role of methylation in the body region its role is not completely understood [114], it is known to modify transcriptional efficiency [44]. *FGF18* is expressed in adipose tissue, neurons, and epidermal keratinocytes [145, 146] and is a downstream target for Wnt/β-catenin signaling pathway crucial for hair and skin development [147]. In mice, *FGF18* was reported to be highly expressed in hair follicles, which are intricately innervated [148] and mediate neuroimmunoendocrine communication [149]. FGF18 remodels the hair follicles in telogen (resting) stage to anagen (hair growth) stage [145]. Interestingly, anagen phase is accompanied with improved wound healing and increased keratinocyte proliferation [150]. During remodeling of hair follicle both sensory and adrenergic skin innervation and in mast cell–nerve count undergoes several changes [151, 152]. Interestingly, interactions between mast cells, nerves and neuropeptides are implicated in atopic dermatitis [153]. Mast cell derived histamine reportedly induces expression of *FGF18* in normal human dermal fibroblasts [154]. With regard to smoke exposure, methylation of another *FGF18* CpG

site cg16654732 located within 200bp of the TSS s reportedly negatively associated with higher smoking (pack years) in Italian lung adenocarcinoma (LUAD) cases compared to normal lung tissue [155]. Additionally, both *FGF18 and AHR* (an intracellular receptor for dioxins in cigarette smoke) were reported to be differentially expressed in neuroblastoma tumors and were enriched in the same functional category of cell proliferation and differentiation [156].

The CpG cg12954512 is located on chromosome 7 between the genes *WDR60* and *LINC00689*. Interestingly, the protein WDR60 is a dynein intermediary chain that is essential for the assembly of cilia [157, 158]. Cilia sense external stimuli and direct these signals influencing cell fate decisions such as growth or differentiation [159]. Interestingly, experimental ablation of cilia in the dermal cells causes an arrest of hair follicle development [160], whereas cilia in the epidermal cells are involved in in epidermal stress responses, and normal keratinocyte differentiation [161]. Additionally, cutaneous nerve fibers are present surrounding keratinocytes and cilia [162], which also plays a role in maintaining interneuronal connectivity [163]. *LINC00689* is a long non-coding RNAs (lncRNAs) produce transcripts over 200 nucleotides in length, which do not translate in to proteins [132]. lncRNAs are found encapsulated in extracellular vesicles of breastmilk [132] play a role in regulation of gene expression [133] and are linked to actinic dermatosis [134] and psoriasis [135].

The CpG cg07208825 belongs to promoter region (TSS200) of *ALMS1P*, a pseudo gene situated next to its parental gene *ALMS1* (Centrosome and Basal Body Associated Protein) on chromosome 2. Lower methylation of pseudogenes promoters are linked to higher expression of parental gene expression [164, 165]. We find that the methylation of cg07208825 (*ALMS1P*) is lower with longer duration of breastfeeding if the mother smokes during pregnancy, which

indicates a possibility of higher expression of the parental gene *ALMS1*. Interestingly, ALMS1 plays a role assembly and function of cilia [116] and interacts with actin cytoskeleton proteins important for cell-to-cell adhesion [116].

Although peripheral blood methylation of the detected genes is relevant with respect to eczema, it remains an open question whether peripheral blood methylation can be used as a substitute for tissue specific genes. Interestingly, these genes are expressed not only skin, but also several other tissues including in whole blood (www.genecards.com). Additionally, several studies have found similar methylation of specific genes in blood and other organs [136, 137], reinforcing that whole blood methylation can reflect the methylation pattern of specific tissues.

It is known that gestational smoking may lead the mothers to initiate breastfeeding later [143] and stop breastfeeding earlier [144]. Although these two variables are interdependent, we calculated the residual duration of breastfeeding by regressing the duration of breastfeeding on gestational smoking. We find interaction of gestational smoking and residual duration of breastfeeding was still associated with DNA-M at age 10 years of all three selected CpGs (Table 12) in the same direction. This implies that part of duration of breastfeeding that is not explained by gestational smoking is still associated with DNA-M at age 10 years and this association varies between children with mothers who smoked during pregnancy compared to children with non-smoking mothers.

This is the first study investigating interplay of maternal smoking and duration of breastfeeding on DNA methylation of genes which related to eczema. Our results suggest that longer duration of breastfeeding combined with gestational smoking is associated with

methylation of genes related to hair follicle and cilia formation, which in turn are linked with lower risk of eczema.

Topical application of tobacco leaves and coal tar that has polycyclic aromatic hydrocarbons (PAHs) similar to tobacco smoke, have been used as ancient remedies of eczema [32, 166, 167]. Although nicotine and other substances from tobacco smoke may pass through placenta and reach the developing offspring[168], we do not find gestational smoking to be associated with DNA-M of these selected CpGs at birth (within a week of birth) indicating in-utero exposure is not sufficient to induce a change in DNA-M.

Our results show a combined effect of both gestational smoking and breastfeeding duration, with increasing duration affecting the DNA-M, and DNA-M in turn linked to a lower risk of eczema. Nicotine, however, is also stored in the fat particles in the breasts and high amounts and is detected in breastmilk if the mother smokes during breastfeeding [30, 31]. Interestingly, nicotine suppresses inflammation in obese mice through cholinergic anti-inflammatory pathway acting via the α7-nicotinic acetylcholine receptors (α7nAChR) [36]. These receptors are also present epidermal hair follicles and dermal fibroblasts in mice and scalp epidermis in humans [169]. Not only nicotine, in smoking mothers several hormones and cytokines such as IL-1α are also altered in breast milk of smoking women [37]. Thus, future research is warranted regarding how breastmilk composition in those who smoked during pregnancy impacts DNA-M in the offspring. Our results suggest that in particular mothers who do not quit smoking during pregnancy should still be encouraged to breastfeed their children.

CHAPTER 5

CONCLUSION

In this dissertation I have investigated the interplay of gestational smoking and duration of breastfeeding on eczema in offspring. In Specific Aim 1 (SA1), my hypothesis is that gestational smoking and duration of breastfeeding jointly influence the risk of eczema. I analyzed the repeated measurements of eczema at ages 1-2 years, 4, 10, and 18 years as outcome using the generalized estimating equation (GEE) approach. I obtained a significant interaction term of gestational smoking and duration of exclusive breastfeeding and the combined effect estimated as follows: RR= exp (β_1×maternal smoking + β_2×duration of exclusive breastfeeding +β_3× (maternal smoking× duration of exclusive breastfeeding)). I assessed confounding using the 10% rule. I found longer duration of breastfeeding (weeks) to be associated with a lower risk of eczema if the mother smoked during pregnancy. The risk ratios for eczema related to maternal smoking and their 95% CI for 3, 9, 15, 21 weeks of breastfeeding were 0.88 (0.66, 1.20), 0.74 (0.54, 1.01), 0.62 (0.41, 0.93), and 0.51 (0.3, 0.88), respectively. Upon stratification, only when the mother smoked during gestation, there is a protective effect of longer duration of breastfeeding on eczema. Using configural frequency analysis (CFA), I found that persistent eczema was statistically significantly more frequent across all the time points of 1-2 years, 4, 10, and 18 years. Additionally the risk of persistent eczema is lower compared to no eczema with longer duration of exclusive breastfeeding if the mother smoked during pregnancy.

Specific Aim 2 (SA2) was to explore whether duration of breastfeeding is linked to DNA methylation (DNA-M) which in turn explain a differential risk of eczema at age 18 years. My hypothesis is that longer duration of breastfeeding may alter methylation of CpGs such that they

are associated with lower risk of eczema at age 18 years. To have an appropriate time order, I investigated the effect of breastfeeding duration on DNA-M at age 10 years and eczema at age 18 years. I included in analysis only children who were eczema-free at age 10 years. I first identified informative CpGs with associated with duration of breastfeeding using linear regression ($\alpha <= 0.05$), followed by testing the association of these CpGs with eczema at age 18 years adjusting for FDR. Only those informative CpGs that were also associated with eczema at age 18 years at a false discovery rate adjusted p-value <=0.05 were further investigated. Since I am interested in protective effects of breastfeeding duration on eczema, I selected the CpGs that show opposite associations i.e., CpGs which are:

(i) whose methylation at age 10 years increase with longer duration of breastfeeding, and higher methylation was associated with a reduced risk of eczema at age 18 years; or

(ii) whose methylation at age 10 years reduce with longer duration of breastfeeding, and lower methylation was associated with a reduced risk of eczema at age 18 years.

In the next step, I assessed confounding for both the associations (breastfeeding and DNA-M, and DNA-M with eczema at age 18 years), followed by path analyses to assess whether there is any indirect effect via these CpGs. Four CpGs passed the applications of these criterial In the path analyses, The methylation of each of the four CpGs, namely, cg03605610 (between *NRC2* and LINC01276), cg04928096 (*AXIN1*), cg26375057 (*ZIC1*) and cg26979504 (*HHEX*) at age 10 years was in the causal pathway between duration of breastfeeding reducing the risk of eczema at age 18 years (Figure 4). The genes of the detected methylation sites are associated with fatty acids, cell–cell adhesion, apoptosis, and retinoic acid signaling pathway and provide an insight regarding their role in eczema pathogenesis linked to breastfeeding exposure.

In Specific Aim 3 (SA3), I assessed the effect of gestational smoking, breastfeeding duration, and DNA-M on eczema. The protective combined effect of gestational smoking and breastfeeding duration that I found in Specific Aim 1 directed me to investigate whether DNA-M can explain this protective effect. My hypothesis is that the combined effect of gestational smoking and breastfeeding duration is associated with DNA-M of CpG sites that can explain a lower risk of eczema at age18 years. The same time order as SA2 was applied, starting with DNA-methylation of eczema-free children. Then it was tested, whether the interaction of gestational smoking and breastfeeding duration predicts DNA-M at age 10 years and the DNA-M at age 10 years predicts eczema at age 18 years. I first identify informative CpGs associated with interaction of gestational smoking and duration of breastfeeding using linear regression ($\alpha <= 0.05$), and testing these informative CpGs with eczema at age 18 years using log-linear regression followed by FDR adjustment. Among these CpGs I focused on the following DNA-M markers:

a) CpGs whose methylation at age 10 years is increased with increasing duration of breastfeeding if the mother smoked during pregnancy, and this higher methylation is associated with a reduced risk of eczema at age 18 years; or

(b) CpGs whose methylation at age 10 years is reduced with increasing duration of breastfeeding if the mother smoked during pregnancy, and this lower methylation is associated with a reduced risk of eczema at age 18 years.

In the next step I assessed confounding for both the associations (interaction of breastfeeding duration and gestational smoking on DNA-M, and DNA-M with eczema at age 18 years).

I identified three CpG [cg07208825 (*ALMS1P*), cg12954512 (Intergenic), and cg21601919 (*FGF18*)] which remained significantly associated with duration of breastfeeding and gestational smoking, and with eczema at age 18 years after including relevant confounders (Figure 6). These CpGs are linked to hair follicle and cilia formation are play a role in keratinocyte differentiation.

Limitations and strengths:

1. *Interdependence of exposures*: Although it is known that gestational smoking may negatively affect breastfeeding initiation, and even lead to early weaning [143, 144]., in SA1 I find a lower risk of eczema in those participants who were exposed to gestational smoking and were breastfed for a longer period of time. This association remained statistically significant even after adjusting for relevant confounders. Additionally, we regressed duration of exclusive breastfeeding on maternal smoking during pregnancy and calculated the residual duration of exclusive breastfeeding. The interaction of residual exclusive breastfeeding duration and gestational smoking showed a similar significant protective effect on eczema with longer duration of breastfeeding.

2. Analytical strategy: In SA1 we model the repeated measurements of eczema status repeatedly measured at age 1-or-2 (year 1 and 2 combined), 4, 10, and 18 year using the generalized estimating equation (GEE) approach with the interaction of the exposures: gestational smoking and duration of breastfeeding as predictors along with relevant confounders.

Although in this setting, in children the development of eczema may have been initiated at age 1-2 years, at the same time when they were being breastfed. This indicates that there remains a possibility of reverse causation at this time point. However, we took this into account in SA2 and

SA3, in which a clear predictive time order was applied with duration of breastfeeding and gestational smoking predicting DNA-M at age 10 years and DNA-M at age 10 years predicting eczema at age 18 years. Additionally, we included only those participants who are eczema-free at age 10 years to assure that there is no influence of eczema at age 10 years on DNA-M at age 10 years.

Since cell types are known to influence DNA-M, I have used cell type residuals of DNA-M throughout the analyses of SA2 and SA3.

I performed a linear regression and tested the association of 349455 CpGs at age 10 years with breastfeeding duration (SA2) and the interaction of gestational smoking and duration of breastfeeding (SA3) using a nominal p-value ($\alpha <= 0.05$) without adjusting for multiple testing. Although this indicates a possibility of having false positives, in this first step I wanted to screen for informative CpGs so that I have a reduced number of variables to be tested for association with eczema. I performed a log linear regression to test the association of the selected informative CpGs with eczema at age 18 years and then adjust for FDR. Among the CpGs that survive FDR, I finally select the ones that follow the directionality criteria of opposite associations with exposures and eczema at age 18.

3. *Small effect sizes* In SA1, although the overall effect size of the interaction term of gestational smoking and duration of breastfeeding was small, it was clearly reduced the risk of eczema at longer duration of breastfeeding showing that a more pronounced effect is seen at longer durations.

In SA2 and SA3, the effect size of the associations of duration of breastfeeding (SA2) and its combined effect of gestational smoking (SA3) had a small effect on DNA-M at age 10 years. To additionally test, whether these small effects indeed show changes of DNA-M related

to breastfeeding, I compared DNA-M at birth measured in Guthrie cards and compared it with DNA-M at age10 years in a small sample (n=172). I demonstrated that breastfeeding duration (SA2)/ interaction of gestational smoking and duration of breastfeeding (SA3) can explain the variation in the DNA-M at age 10 years after adjusting for DNA-M from Guthrie cards indicating that they have an effect on DNA-M at age10 years.

4. The peripheral blood methylation of the CpGs that we detected are in SA2 and SA3 are relevant with respect to eczema. Although an open question remains on whether peripheral blood methylation can be used as a proxy for tissue specific genes. To this end, I checked the expression of these genes in an online data base (www.genecards.com) where I found these genes to be expressed not only skin, but several other tissues including whole blood. Additionally, several studies have found similar methylation of specific genes in blood and other organs [136, 137], reinforcing that whole blood methylation can reflect the methylation pattern of specific tissues.

In conclusion, this dissertation aids in the understanding of effect of two early life exposures, duration of breastfeeding and maternal gestational smoking with offspring eczema and the role of DNA methylation as an intermediate factor between these exposures and the outcome. The current studies linking maternal gestational smoking with offspring eczema have demonstrated contradictory associations. Similarly, studies linking duration of breastfeeding with eczema have also yielded conflicting results. Chapter one of this dissertation is the first study to demonstrate that a combined effect of duration of breastfeeding and maternal smoking is protective of persistent eczema in the offspring at ages 1-2, 4, 10, and 18 years, which offers a potential explanation regarding the conflicting results in the literature.

Whereas chapter one provides and epidemiological association between duration of breastfeeding, maternal smoking and eczema, chapters two and three provide an insight regarding how these exposures biologically mediate their effect via altered DNA methylation on offspring eczema. No existing study has investigated whether differentially methylated sites associated breastfeeding duration, or with the combined effect of breastfeeding, duration and maternal gestational smoking can explain the risk of eczema. Chapters two and three focusses on identification of such differentially methylated sites at age 10 years that may impart a different risk of eczema at age 18 years in the offspring. The second chapter demonstrates that longer duration of breastfeeding alone also mediate a protective effect on eczema at age18 years via DNA-M, whereas the third chapter shows that duration of breastfeeding and gestational smoking may jointly influence DNA-M that in turn explain a lower risk of eczema at age18 years. To have an appropriate predictive time order, SA2 and SA3 focus on eczema at age 18 years, where a delayed effect of the exposures are shown on eczema at age 18 years. These same CpGs may not be able to explain a protective effect at earlier ages since incident eczema at age 18 may have a different pathogenesis than that at earlier ages. Replication of these results would be useful to validate the findings, especially with methylation of the target skin tissue of eczema.

The work detected CpGs belonging to seven new genes whose DNA methylation seems to be involved in the development of eczema at age 18 years which are described as follows:

I. The CpG cg04928096 is located on the 3'UTR of the *AXIN1* gene. *AXIN1* (Axis Inhibition Protein 1) gene is a part of the Wnt signaling pathway and induces the ubiquination and degradation of cytosolic β-catenin [170]. β-catenin is involved in cell–cell adhesion by interacting with cadherin proteins and its degradation may lead to

apoptosis of cells [117]. Ablation of β-catenin in epidermis has been reported to lead to barrier dysfunction indicating its role in epidermal function [118], thus corroborating our findings of higher methylation for the CpG of *AXIN1* associated with lower risk of eczema. No direct association is reported in the literature regarding breastfeeding duration and the *AXIN1* gene. However, it is known that Epidermal Growth Factor (EGF), a ligand of Epidermal Growth Factor Receptor is abundant in breastmilk [119]. Studies have demonstrated that ligand activated EGFR leads to phosphorylation of several downstream proteins leading to transfer of Axin complexed with other proteins (GSK3-APC-Axin complex)[120] to the cell membrane, thus promoting the stabilization of β-catenin[120]. EGF in breastmilk may also inactivate GSK3 (glycogen synthase kinase 3) and stabilize β-catenin [119].

II. The CpG cg26979504 is in the body region of the gene *HHEX* (Hematopoietically Expressed Homeobox). This gene codes for a transcription factor that is involved in several developmental and differentiation processes including that of the mammary epithelial cells [121]. Puppin *et al* have reported that higher nuclear localization of *HHEX* in lactating tissue compared to normal non-lactating human breast tissue [121]. Kothapalli *et al* found a higher expression of *HHEX i*n the cerebral cortex region of baboons who were fed higher amounts of Docosahexaenoic acid (DHA) than controls [122] indicating a possible link between fatty acids found in breast milk with *HHEX* expression. Additionally, lower expression of *HHEX* was reported in dermal mesenchymal stem cells in psoriatic skin lesions compared to controls [123] implying its role in inflammatory skin disease.

III. The CpG cg26375057 is located in the first exon of the *ZIC1* gene. The ZIC1 (zinc finger in the cerebellum 1) gene produces a transcriptional activator and is involved in neurogenesis [124], and is a classic marker of brown adipose tissue (BAT) [125] which is present in breast tissue [126]. An in-vitro analysis with mice adipose cells reported that DHA synthesis takes place in the BAT cells [125]. Zic1 is important for maintaining the expression of retinoic acid (RA)-degrading enzyme cyp26a1 in the forebrain of zebrafish, with Zic1 loss of function linked to higher the levels of RA [127]. Interestingly, RAs play a crucial role in the maintenance of epithelial tissues and retinoic acid isomers are also used as drugs to treat several dermatological diseases [128] including eczema [129].

IV. The CpG cg03605610 is an intergenic CpG on chromosome 6 lying between the genes *NRC2* (NK Cell-Activating Receptor) and Long Intergenic Non-Protein Coding RNA 1276. *NRC2*, also known NKp44 is an NK cell activating receptor [130] is implicated in psoriatic inflammation [131]. Lesional skin and peripheral blood of psoriasis patients have a high proportion of innate lymphoid cells expressing this receptor (NKp44 ++ ILC3) which also produce the IL22, a cytokine that triggers epidermal thickening in psoriatic skin [131]. Long non-coding RNAs (lncRNAs) produce transcripts over 200 nucleotides in length, which do not translate in to proteins [132]. They are found encapsulated in extracellular vesicles of breastmilk [132] playing a role in the regulation of gene expression [133] and are linked to actinic dermatosis [134] and psoriasis [135].

V. The CpG cg21601919 is located in the body region of the gene *FGF18* (*Fibroblast Growth Factor 18)*. Although the role of methylation in the body region its role is not completely understood [114], it is known to modify transcriptional efficiency [44]. *FGF18* is expressed in adipose tissue, neurons, and epidermal keratinocytes [145, 146]

and is a downstream target for Wnt/β-catenin signaling pathway crucial for hair and skin development [147]. In mice, *FGF18* was reported to be highly expressed in hair follicles, which are intricately innervated [148] and mediate neuroimmunoendocrine communication [149]. FGF18 remodels the hair follicles in telogen (resting) stage to anagen (hair growth) stage [145]. Interestingly, anagen phase is accompanied with improved wound healing and increased keratinocyte proliferation [150]. During remodeling of hair follicle both sensory and adrenergic skin innervation and in mast cell–nerve count undergoes several changes [151, 152]. Interestingly, interactions between mast cells, nerves and neuropeptides are implicated in atopic dermatitis [153]. Mast cell derived histamine reportedly induces expression of *FGF18* in normal human dermal fibroblasts [154]. With regard to smoke exposure, methylation of another *FGF18* CpG site cg16654732 located within 200bp of the TSS s reportedly negatively associated with higher smoking (pack years) in Italian lung adenocarcinoma (LUAD) cases compared to normal lung tissue [155]. Additionally, both *FGF18 and AHR* (an intracellular receptor for dioxins in cigarette smoke) were reported to be differentially expressed in neuroblastoma tumors and were enriched in the same functional category of cell proliferation and differentiation [156].

VI. The CpG cg12954512 is located on chromosome 7 between the genes *WDR60* and *LINC00689*. Interestingly, the protein WDR60 is a dynein intermediary chain that is essential for the assembly of cilia [157, 158]. Cilia sense external stimuli and direct these signals influencing cell fate decisions such as growth or differentiation [159]. Interestingly, experimental ablation of cilia in the dermal cells causes an arrest of hair follicle development [160], whereas cilia in the epidermal cells are involved in in

epidermal stress responses, and normal keratinocyte differentiation [161]. Additionally, cutaneous nerve fibers are present surrounding keratinocytes and cilia [162], which also plays a role in maintaining interneuronal connectivity [163]. *LINC00689* is a long non-coding RNAs (lncRNAs) produce transcripts over 200 nucleotides in length, which do not translate in to proteins [132]. lncRNAs are found encapsulated in extracellular vesicles of breastmilk [132] play a role in regulation of gene expression [133] and are linked to actinic dermatosis [134] and psoriasis [135].

VII. The CpG cg07208825 belongs to promoter region (TSS200) of *ALMS1P,* a pseudo gene situated next to its parental gene *ALMS1* (Centrosome and Basal Body Associated Protein) on chromosome 2. Lower methylation of pseudogenes promoters are linked to higher expression of parental gene expression [164, 165]. We find that the methylation of cg07208825 (*ALMS1P*) is lower with longer duration of breastfeeding if the mother smokes during pregnancy, which indicates a possibility of higher expression of the parental gene *ALMS1*. Interestingly, ALMS1 plays a role assembly and function of cilia [116] and interacts with actin cytoskeleton proteins important for cell-to-cell adhesion [116].

Additional studies are warranted to further test the role of these genes in relation to gestational smoking, breastfeeding duration and eczema in in other cohorts.

References:

1. Sohn, A., et al., *Eczema.* Mt Sinai J Med, 2011. **78**(5): p. 730-9.
2. Boguniewicz, M. and D.Y. Leung, *Atopic dermatitis: a disease of altered skin barrier and immune dysregulation.* Immunol Rev, 2011. **242**(1): p. 233-46.
3. Flohr, C. and J. Mann, *New insights into the epidemiology of childhood atopic dermatitis.* Allergy, 2014. **69**(1): p. 3-16.
4. Bantz, S.K., Z. Zhu, and T. Zheng, *The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma.* J Clin Cell Immunol, 2014. **5**(2).
5. Schmitt, J., et al., *Atopic dermatitis is associated with an increased risk for rheumatoid arthritis and inflammatory bowel disease, and a decreased risk for type 1 diabetes.* J Allergy Clin Immunol, 2016. **137**(1): p. 130-6.
6. Meng, X., et al., *Eczema as the first manifestation of a lung cancer.* J Cancer Res Ther, 2015. **11**(3): p. 659.
7. Silverberg, J.I., *Health Care Utilization, Patient Costs, and Access to Care in US Adults With Eczema: A Population-Based Study.* JAMA Dermatol, 2015. **151**(7): p. 743-52.
8. Lee, B.W. and P.R. Detzel, *Treatment of childhood atopic dermatitis and economic burden of illness in Asia Pacific countries.* Ann Nutr Metab, 2015. **66 Suppl 1**: p. 18-24.
9. Bisgaard, H., et al., *Gene-environment interaction in the onset of eczema in infancy: filaggrin loss-of-function mutations enhanced by neonatal cat exposure.* PLoS Med, 2008. **5**(6): p. e131.
10. Kezic, S. and I. Jakasa, *Filaggrin and Skin Barrier Function.* Curr Probl Dermatol, 2016. **49**: p. 1-7.

11. Weidinger, S., et al., *Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations.* J Allergy Clin Immunol, 2006. **118**(1): p. 214-9.

12. Rodriguez, E., et al., *Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease.* J Allergy Clin Immunol, 2009. **123**(6): p. 1361-70 e7.

13. O'Regan, G.M., et al., *Filaggrin in atopic dermatitis.* J Allergy Clin Immunol, 2008. **122**(4): p. 689-93.

14. Brown, S.J. and W.H. McLean, *One remarkable molecule: filaggrin.* J Invest Dermatol, 2012. **132**(3 Pt 2): p. 751-62.

15. Ziyab, A.H., et al., *Interplay of filaggrin loss-of-function variants, allergic sensitization, and eczema in a longitudinal study covering infancy to 18 years of age.* PLoS One, 2012. **7**(3): p. e32721.

16. Bussmann, C., S. Weidinger, and N. Novak, *Genetics of atopic dermatitis.* J Dtsch Dermatol Ges, 2011. **9**(9): p. 670-6.

17. Schafer, T., et al., *Maternal smoking during pregnancy and lactation increases the risk for atopic eczema in the offspring.* J Am Acad Dermatol, 1997. **36**(4): p. 550-6.

18. Sorensen, J.A., et al., *Tobacco smoking and hand eczema - is there an association?* Contact Dermatitis, 2015. **73**(6): p. 326-35.

19. Kramer, M.S., *Breastfeeding and allergy: the evidence.* Ann Nutr Metab, 2011. **59 Suppl 1**: p. 20-6.

20. Ito, J. and T. Fujiwara, *Breastfeeding and risk of atopic dermatitis up to the age 42 months: a birth cohort study in Japan.* Ann Epidemiol, 2014. **24**(4): p. 267-72.

21. Flohr, C., et al., *Effect of an Intervention to Promote Breastfeeding on Asthma, Lung Function, and Atopic Eczema at Age 16 Years: Follow-up of the PROBIT Randomized Trial.* JAMA Pediatr, 2017: p. e174064.

22. Chiu, C.Y., et al., *Exclusive or Partial Breastfeeding for 6 Months Is Associated With Reduced Milk Sensitization and Risk of Eczema in Early Childhood: The PATCH Birth Cohort Study.* Medicine (Baltimore), 2016. **95**(15): p. e3391.

23. Yang, Y.W., C.L. Tsai, and C.Y. Lu, *Exclusive breastfeeding and incident atopic dermatitis in childhood: a systematic review and meta-analysis of prospective cohort studies.* Br J Dermatol, 2009. **161**(2): p. 373-83.

24. Thyssen, J.P., et al., *The effect of tobacco smoking and alcohol consumption on the prevalence of self-reported hand eczema: a cross-sectional population-based study.* Br J Dermatol, 2010. **162**(3): p. 619-26.

25. Adeyemi, A.S., et al., *The Prevalence, Risk Factors and Changes in Symptoms of Self Reported Asthma, Rhinitis and Eczema Among Pregnant Women in Ogbomoso, Nigeria.* J Clin Diagn Res, 2015. **9**(9): p. OC01-7.

26. Kim, H.B., et al., *Lifetime prevalence of childhood eczema and the effect of indoor environmental factors: Analysis in Hispanic and non-Hispanic white children.* Allergy Asthma Proc, 2016. **37**(1): p. 64-71.

27. Magnusson, L.L., et al., *Wheezing, asthma, hayfever, and atopic eczema in childhood following exposure to tobacco smoke in fetal life.* Clin Exp Allergy, 2005. **35**(12): p. 1550-6.

28. Taylor-Robinson, D.C., et al., *Do early-life exposures explain why more advantaged children get eczema? Findings from the U.K. Millennium Cohort Study.* Br J Dermatol, 2016. **174**(3): p. 569-78.

29. Ek, W.E., et al., *Breast-feeding and risk of asthma, hay fever, and eczema.* J Allergy Clin Immunol, 2017.

30. Amir, L.H., *Maternal smoking and reduced duration of breastfeeding: a review of possible mechanisms.* Early Hum Dev, 2001. **64**(1): p. 45-67.

31. Primo, C.C., et al., *Effects of maternal nicotine on breastfeeding infants.* Rev Paul Pediatr, 2013. **31**(3): p. 392-7.

32. McLean, W.H. and A.D. Irvine, *Old King coal - molecular mechanisms underlying an ancient treatment for atopic eczema.* J Clin Invest, 2013. **123**(2): p. 551-3.

33. van den Bogaard, E.H., et al., *Coal tar induces AHR-dependent skin barrier repair in atopic dermatitis.* J Clin Invest, 2013. **123**(2): p. 917-27.

34. Xie, G., Z. Peng, and J.P. Raufman, *Src-mediated aryl hydrocarbon and epidermal growth factor receptor cross talk stimulates colon cancer cell proliferation.* Am J Physiol Gastrointest Liver Physiol, 2012. **302**(9): p. G1006-15.

35. Robertson, E.D., et al., *ARNT controls the expression of epidermal differentiation genes through HDAC- and EGFR-dependent pathways.* J Cell Sci, 2012. **125**(Pt 14): p. 3320-32.

36. Wang, X., et al., *Activation of the cholinergic antiinflammatory pathway ameliorates obesity-induced inflammation and insulin resistance.* Endocrinology, 2011. **152**(3): p. 836-46.

37. Zanardo, V., et al., *Effect of maternal smoking on breast milk interleukin-1alpha, beta-endorphin, and leptin concentrations and leptin concentrations.* Environ Health Perspect, 2005. **113**(10): p. 1410-3.

38. Ermis, B., et al., *Influence of smoking on human milk tumor necrosis factor-alpha, interleukin-1beta, and soluble vascular cell adhesion molecule-1 levels at postpartum seventh day.* Pediatr Int, 2009. **51**(6): p. 821-4.

39. Etem Piskin, I., et al., *Effect of maternal smoking on colostrum and breast milk cytokines.* Eur Cytokine Netw, 2012. **23**(4): p. 187-90.

40. Robertson, K.D., *DNA methylation and human disease.* Nat Rev Genet, 2005. **6**(8): p. 597-610.

41. Lockett, G.A., et al., *Epigenomics and allergic disease.* Epigenomics, 2013. **5**(6): p. 685-99.

42. Richmond, R.C., et al., *Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC).* Hum Mol Genet, 2015. **24**(8): p. 2201-17.

43. Liu, L., Y. Li, and T.O. Tollefsbol, *Gene-environment interactions and epigenetic basis of human diseases.* Curr Issues Mol Biol, 2008. **10**(1-2): p. 25-36.

44. Shenker, N. and J.M. Flanagan, *Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research.* Br J Cancer, 2012. **106**(2): p. 248-53.

45. Karmaus, W., et al., *Epigenetic mechanisms and models in the origins of asthma.* Curr Opin Allergy Clin Immunol, 2013. **13**(1): p. 63-9.

46. Gao, X., et al., *DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies.* Clin Epigenetics, 2015. **7**: p. 113.

47. Markunas, C.A., et al., *Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy.* Environ Health Perspect, 2014. **122**(10): p. 1147-53.

48. Philibert, R.A., et al., *Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking.* Clin Epigenetics, 2013. **5**(1): p. 19.

49. Joubert, B.R., et al., *450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy.* Environ Health Perspect, 2012. **120**(10): p. 1425-31.

50. Joubert, B.R., et al., *DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis.* Am J Hum Genet, 2016. **98**(4): p. 680-96.

51. Hartwig, F.P., et al., *Breastfeeding effects on DNA methylation in the offspring: A systematic literature review.* PLoS One, 2017. **12**(3): p. e0173070.

52. Palmer, C.N., et al., *Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis.* Nat Genet, 2006. **38**(4): p. 441-6.

53. Spergel, J.M. and A.S. Paller, *Atopic dermatitis and the atopic march.* J Allergy Clin Immunol, 2003. **112**(6 Suppl): p. S118-27.

54. Molin, S., et al., *Filaggrin mutations may confer susceptibility to chronic hand eczema characterized by combined allergic and irritant contact dermatitis.* Br J Dermatol, 2009. **161**(4): p. 801-7.

55. Rupnik, H., M. Rijavec, and P. Korosec, *Filaggrin loss-of-function mutations are not associated with atopic dermatitis that develops in late childhood or adulthood.* Br J Dermatol, 2015. **172**(2): p. 455-61.

56. Barker, D.J. and P.M. Clark, *Fetal undernutrition and disease in later life.* Rev Reprod, 1997. **2**(2): p. 105-12.

57. Knopik, V.S., et al., *The epigenetics of maternal cigarette smoking during pregnancy and effects on child development.* Dev Psychopathol, 2012. **24**(4): p. 1377-90.

58. Kantor, R., et al., *Association of atopic dermatitis with smoking: A systematic review and meta-analysis.* J Am Acad Dermatol, 2016.

59. Elbert, N.J., et al., *Duration and exclusiveness of breastfeeding and risk of childhood atopic diseases.* Allergy, 2017.

60. Hong, S., et al., *Effect of prolonged breast-feeding on risk of atopic dermatitis in early childhood.* Allergy Asthma Proc, 2014. **35**(1): p. 66-70.

61. Lee, K.S., et al., *Does Breast-feeding Relate to Development of Atopic Dermatitis in Young Korean Children?: Based on the Fourth and Fifth Korea National Health and Nutrition Examination Survey 2007-2012.* Allergy Asthma Immunol Res, 2017. **9**(4): p. 307-313.

62. Turati, F., et al., *Early weaning is beneficial to prevent atopic dermatitis occurrence in young children.* Allergy, 2016. **71**(6): p. 878-88.

63. Ziyab, A.H., et al., *Filaggrin gene loss-of-function variants modify the effect of breast-feeding on eczema risk in early childhood.* Allergy, 2016. **71**(9): p. 1371-3.

64. Arshad, S.H. and D.W. Hide, *Effect of environmental factors on the development of allergic disorders in infancy.* J Allergy Clin Immunol, 1992. **90**(2): p. 235-41.

65. Arshad, S.H., M. Stevens, and D.W. Hide, *The effect of genetic and environmental factors on the prevalence of allergic disorders at the age of two years.* Clin Exp Allergy, 1993. **23**(6): p. 504-11.

66. Scott, M., et al., *Influence of atopy and asthma on exhaled nitric oxide in an unselected birth cohort study.* Thorax, 2010. **65**(3): p. 258-62.

67. Soto-Ramirez, N., et al., *Epidemiologic methods of assessing asthma and wheezing episodes in longitudinal studies: measures of change and stability.* J Epidemiol, 2013. **23**(6): p. 399-410.

68. Hanifin, J., Rajka G. , *Diagnostic features of atopic eczema.* . Acta Derm Venereol (Stockh), 1980. **92**: p. 44–7.

69. Arshad, S.H., et al., *Polymorphisms in the interleukin 13 and GATA binding protein 3 genes and the development of eczema during childhood.* Br J Dermatol, 2008. **158**(6): p. 1315-22.

70. Ogbuanu, I.U., et al., *Effect of breastfeeding duration on lung function at age 10 years: a prospective birth cohort study.* Thorax, 2009. **64**(1): p. 62-6.

71. Zeger, S.L. and K.Y. Liang, *Longitudinal data analysis for discrete and continuous outcomes.* Biometrics, 1986. **42**(1): p. 121-30.

72. Zhang, J. and K.F. Yu, *What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes.* JAMA, 1998. **280**(19): p. 1690-1.

73. von Eye, A., *Configural frequency analysis, Version 2000. A Program for 32 Bit windows operating system. Methods of psychological research online.* Methods of Psychological Research, 2001. **6**(2).

74. Kim, J.H., *Role of Breast-feeding in the Development of Atopic Dermatitis in Early Childhood.* Allergy Asthma Immunol Res, 2017. **9**(4): p. 285-287.

75. Tanaka, K., et al., *Pre- and Postnatal Smoking Exposure and Risk of Atopic Eczema in Young Japanese Children: A Prospective Prebirth Cohort Study.* Nicotine Tob Res, 2017. **19**(7): p. 804-809.

76. Ziyab, A.H., et al., *Trends in eczema in the first 18 years of life: results from the Isle of Wight 1989 birth cohort study.* Clin Exp Allergy, 2010. **40**(12): p. 1776-84.

77. Li, R., K.S. Scanlon, and M.K. Serdula, *The validity and reliability of maternal recall of breastfeeding practice.* Nutr Rev, 2005. **63**(4): p. 103-10.

78. Bergmann, R.L., et al., *Socioeconomic status is a risk factor for allergy in parents but not in their children.* Clin Exp Allergy, 2000. **30**(12): p. 1740-5.

79. Laitinen, K., et al., *Breast milk fatty acids may link innate and adaptive immune regulation: analysis of soluble CD14, prostaglandin E2, and fatty acids.* Pediatr Res, 2006. **59**(5): p. 723-7.

80. Du, L., S.M. Hoffman, and D.S. Keeney, *Epidermal CYP2 family cytochromes P450.* Toxicol Appl Pharmacol, 2004. **195**(3): p. 278-87.

81. Furue, M., et al., *Role of AhR/ARNT system in skin homeostasis.* Arch Dermatol Res, 2014. **306**(9): p. 769-79.

82. Terry, P.D. and T.E. Rohan, *Cigarette smoking and the risk of breast cancer in women: a review of the literature.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(10 Pt 1): p. 953-71.

83. Madhavan, N.D. and K.A. Naidu, *Polycyclic aromatic hydrocarbons in placenta, maternal blood, umbilical cord blood and milk of Indian women.* Hum Exp Toxicol, 1995. **14**(6): p. 503-6.

84. Deckers, I.A., et al., *Investigating international time trends in the incidence and prevalence of atopic eczema 1990-2010: a systematic review of epidemiological studies.* PLoS One, 2012. **7**(7): p. e39803.

85. Irvine, A.D., W.H. McLean, and D.Y. Leung, *Filaggrin mutations associated with skin and allergic diseases.* N Engl J Med, 2011. **365**(14): p. 1315-27.

86. Chen, H., et al., *Wide spectrum of filaggrin-null mutations in atopic dermatitis highlights differences between Singaporean Chinese and European populations.* Br J Dermatol, 2011. **165**(1): p. 106-14.

87. Zhang, H., et al., *Mutations in the filaggrin gene in Han Chinese patients with atopic dermatitis.* Allergy, 2011. **66**(3): p. 420-7.

88. Lodge, C.J., et al., *Breastfeeding and asthma and allergies: a systematic review and meta-analysis.* Acta Paediatr, 2015. **104**(467): p. 38-53.

89. Bjorksten, B., et al., *Global analysis of breast feeding and risk of symptoms of asthma, rhinoconjunctivitis and eczema in 6-7 year old children: ISAAC Phase Three.* Allergol Immunopathol (Madr), 2011. **39**(6): p. 318-25.

90. Saarinen, U.M. and M. Kajosaari, *Breastfeeding as prophylaxis against atopic disease: prospective follow-up study until 17 years old.* Lancet, 1995. **346**(8982): p. 1065-9.

91. Jelding-Dannemand, E., A.M. Malby Schoos, and H. Bisgaard, *Breast-feeding does not protect against allergic sensitization in early childhood and allergy-associated disease at age 7 years.* J Allergy Clin Immunol, 2015. **136**(5): p. 1302-8 e1-13.

92. Rossnerova, A., et al., *Factors affecting the 27K DNA methylation pattern in asthmatic and healthy children from locations with various environments.* Mutat Res, 2013. **741-742**: p. 18-26.

93. Obermann-Borst, S.A., et al., *Duration of breastfeeding and gender are associated with methylation of the LEPTIN gene in very young children.* Pediatr Res, 2013. **74**(3): p. 344-9.

94. Soto-Ramirez, N., et al., *The interaction of breastfeeding, DNA methylation, and genetic variants in chromosome 17q12 and the risk of asthma in girls at age 18 years.* American Journal of Respiratory and Critical Care Medicine, 2013. **American Thoracic Society 2013 International Conference Philadelphia, USA**: p. A:3517.

95. Tao, M.H., et al., *Exposures in early life: associations with DNA promoter methylation in breast tumors.* J Dev Orig Health Dis, 2013. **4**(2): p. 182-90.

96. Soto-Ramirez, N., et al., *The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years.* Clin Epigenetics, 2013. **5**(1): p. 1.

97. Beyan, H., et al., *Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans.* Genome Res, 2012. **22**(11): p. 2138-45.

98. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies.* Genome Biol, 2015. **16**: p. 37.

99. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.

100. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.

101. Bakulski, K.M., et al., *DNA methylation of cord blood cell types: Applications for mixed cell birth studies.* Epigenetics, 2016. **11**(5): p. 354-62.

102. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.

103. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.

104. Adalsteinsson, B.T., et al., *Heterogeneity in white blood cells has potential to confound DNA methylation measurements.* PLoS One, 2012. **7**(10): p. e46705.

105. Leenen, F.A., C.P. Muller, and J.D. Turner, *DNA methylation: conducting the orchestra from exposure to phenotype?* Clin Epigenetics, 2016. **8**: p. 92.

106. Breton, C.V., et al., *Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group.* Environ Health Perspect, 2017. **125**(4): p. 511-526.

107. Pidsley, R., et al., *Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.* Genome Biol, 2016. **17**(1): p. 208.

108. Mahmood, S., et al., *Epigenetic changes in hypothalamic appetite regulatory genes may underlie the developmental programming for obesity in rat neonates subjected to a high-carbohydrate dietary modification.* J Dev Orig Health Dis, 2013. **4**(6): p. 479-90.

109. Raychaudhuri, N., et al., *Postnatal exposure to a high-carbohydrate diet interferes epigenetically with thyroid hormone receptor induction of the adult male rat skeletal muscle glucose transporter isoform 4 expression.* J Nutr Biochem, 2014. **25**(10): p. 1066-76.

110. Ziyab, A.H., et al., *DNA methylation of the filaggrin gene adds to the risk of eczema associated with loss-of-function variants.* J Eur Acad Dermatol Venereol, 2013. **27**(3): p. e420-3.

111. Bin, L. and D.Y. Leung, *Genetic and epigenetic studies of atopic dermatitis.* Allergy Asthma Clin Immunol, 2016. **12**: p. 52.

112. Quraishi, B.M., et al., *Identifying CpG sites associated with eczema via random forest screening of epigenome-scale DNA methylation.* Clin Epigenetics, 2015. **7**: p. 68.

113. Brenet, F., et al., *DNA methylation of the first exon is tightly linked to transcriptional silencing.* PLoS One, 2011. **6**(1): p. e14524.

114. Jjingo, D., et al., *On the presence and role of human gene-body DNA methylation.* Oncotarget, 2012. **3**(4): p. 462-74.

115. Maussion, G., et al., *Functional DNA methylation in a transcript specific 3'UTR region of TrkB associates with suicide.* Epigenetics, 2014. **9**(8): p. 1061-70.

116. Collin, G.B., et al., *The Alstrom syndrome protein, ALMS1, interacts with alpha-actinin and components of the endosome recycling pathway.* PLoS One, 2012. **7**(5): p. e37925.

117. Jaiswal, A.S., et al., *Beta-catenin-mediated transactivation and cell-cell adhesion pathways are important in curcumin (diferuylmethane)-induced growth arrest and apoptosis in colon cancer cells.* Oncogene, 2002. **21**(55): p. 8414-27.

118. Ray, S., H.P. Foote, and T. Lechler, *beta-Catenin protects the epidermis from mechanical stresses.* J Cell Biol, 2013. **202**(1): p. 45-52.

119. Good, M., et al., *Breast milk protects against the development of necrotizing enterocolitis through inhibition of Toll-like receptor 4 in the intestinal epithelium via activation of the epidermal growth factor receptor.* Mucosal Immunol, 2015. **8**(5): p. 1166-79.

120. Zhang, X., et al., *Epidermal growth factor receptor (EGFR) signaling regulates epiphyseal cartilage development through beta-catenin-dependent and -independent pathways.* J Biol Chem, 2013. **288**(45): p. 32229-40.

121. Puppin, C., et al., *HEX expression and localization in normal mammary gland and breast carcinoma.* BMC Cancer, 2006. **6**: p. 192.

122. Kothapalli, K.S., et al., *Differential cerebral cortex transcriptomes of baboon neonates consuming moderate and high docosahexaenoic acid formulas.* PLoS One, 2007. **2**(4): p. e370.

123. Chang, W.J., et al., *LITAF, HHEX, and DUSP1 expression in mesenchymal stem cells from patients with psoriasis.* Genet Mol Res, 2015. **14**(4): p. 15793-801.

124. Aruga, J., *The role of Zic genes in neural development.* Mol Cell Neurosci, 2004. **26**(2): p. 205-21.

125. Qin, X., et al., *Brown but not white adipose cells synthesize omega-3 docosahexaenoic acid in culture.* Prostaglandins Leukot Essent Fatty Acids, 2016. **104**: p. 19-24.

126. Wang, F., et al., *Mammary fat of breast cancer: gene expression profiling and functional characterization.* PLoS One, 2014. **9**(10): p. e109742.

127. Maurus, D. and W.A. Harris, *Zic-associated holoprosencephaly: zebrafish Zic1 controls midline formation and forebrain patterning by regulating Nodal, Hedgehog, and retinoic acid signaling.* Genes Dev, 2009. **23**(12): p. 1461-73.

128. Nelson, C.H., B.R. Buttrick, and N. Isoherranen, *Therapeutic potential of the inhibition of the retinoic acid hydroxylases CYP26A1 and CYP26B1 by xenobiotics.* Curr Top Med Chem, 2013. **13**(12): p. 1402-28.

129. Bollag, W. and F. Ott, *Successful treatment of chronic hand eczema with oral 9-cis-retinoic acid.* Dermatology, 1999. **199**(4): p. 308-12.

130. Deniz, G., W. van de Veen, and M. Akdis, *Natural killer cells in patients with allergic diseases.* J Allergy Clin Immunol, 2013. **132**(3): p. 527-35.

131. Villanova, F., et al., *Characterization of innate lymphoid cells in human skin and blood demonstrates increase of NKp44+ ILC3 in psoriasis.* J Invest Dermatol, 2014. **134**(4): p. 984-991.

132. Karlsson, O., et al., *Detection of long non-coding RNAs in human breastmilk extracellular vesicles: Implications for early child development.* Epigenetics, 2016: p. 0.

133. Fatica, A. and I. Bozzoni, *Long non-coding RNAs: new players in cell differentiation and development.* Nat Rev Genet, 2014. **15**(1): p. 7-21.

134. Dongyun, L., et al., *Genome-Wide Analysis of mRNA and Long Noncoding RNA Profiles in Chronic Actinic Dermatitis.* BioMed Research International, 2017. **2017**: p. 15.

135. Ahn, R., et al., *Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs.* BMC Genomics, 2016. **17**(1): p. 841.

136. Bysani, M., et al., *Epigenetic alterations in blood mirror age-associated DNA methylation and gene expression changes in human liver.* Epigenomics, 2017. **9**(2): p. 105-122.

137. Booij, L., et al., *Looking beyond the DNA sequence: the relevance of DNA methylation processes for the stress-diathesis model of depression.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1615): p. 20120251.

138. Guttman-Yassky, E. and J.G. Krueger, *Atopic dermatitis and psoriasis: two different immune diseases or one spectrum?* Curr Opin Immunol, 2017. **48**: p. 68-73.

139. Guttman-Yassky, E., et al., *Atopic dermatitis: pathogenesis.* Semin Cutan Med Surg, 2017. **36**(3): p. 100-103.

140. Candi, E., R. Schmidt, and G. Melino, *The cornified envelope: a model of cell death in the skin.* Nat Rev Mol Cell Biol, 2005. **6**(4): p. 328-40.

141. Lee, K.W. and Z. Pausova, *Cigarette smoking and DNA methylation.* Front Genet, 2013. **4**: p. 132.

142. Kupers, L.K., et al., *DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring.* Int J Epidemiol, 2015. **44**(4): p. 1224-37.

143. Leung, G.M., L.M. Ho, and T.H. Lam, *Maternal, paternal and environmental tobacco smoking and breast feeding.* Paediatr Perinat Epidemiol, 2002. **16**(3): p. 236-45.

144. Horta, B.L., M.S. Kramer, and R.W. Platt, *Maternal smoking and the risk of early weaning: a meta-analysis.* Am J Public Health, 2001. **91**(2): p. 304-7.

145. Kawano, M., et al., *Comprehensive analysis of FGF and FGFR expression in skin: FGF18 is highly expressed in hair follicles and capable of inducing anagen from telogen stage hair follicles.* J Invest Dermatol, 2005. **124**(5): p. 877-85.

146. Hasegawa, H., et al., *Laminar patterning in the developing neocortex by temporally coordinated fibroblast growth factor signaling.* J Neurosci, 2004. **24**(40): p. 8711-9.

147. Shimokawa, T., et al., *Involvement of the FGF18 gene in colorectal carcinogenesis, as a novel downstream target of the beta-catenin/T-cell factor complex.* Cancer Res, 2003. **63**(19): p. 6116-20.

148. Paus, R. and K. Foitzik, *In search of the "hair cycle clock": a guided tour.* Differentiation, 2004. **72**(9-10): p. 489-511.

149. Paus, R., T.C. Theoharides, and P.C. Arck, *Neuroimmunoendocrine circuitry of the 'brain-skin connection'.* Trends Immunol, 2006. **27**(1): p. 32-9.

150. Ansell, D.M., et al., *Exploring the "hair growth-wound healing connection": anagen phase promotes wound re-epithelialization.* J Invest Dermatol, 2011. **131**(2): p. 518-28.

151. Arck, P.C., et al., *Stress inhibits hair growth in mice by induction of premature catagen development and deleterious perifollicular inflammatory events via neuropeptide substance P-dependent pathways.* Am J Pathol, 2003. **162**(3): p. 803-14.

152. Peters, E.M., et al., *Hair-cycle-associated remodeling of the peptidergic innervation of murine skin, and hair growth modulation by neuropeptides.* J Invest Dermatol, 2001. **116**(2): p. 236-45.

153. Jarvikallio, A., I.T. Harvima, and A. Naukkarinen, *Mast cells, nerves and neuropeptides in atopic dermatitis and nummular eczema.* Arch Dermatol Res, 2003. **295**(1): p. 2-7.

154. Murota, H., et al., *Emedastine difumarate inhibits histamine-induced collagen synthesis in dermal fibroblasts.* J Investig Allergol Clin Immunol, 2008. **18**(4): p. 245-52.

155. Freeman, J.R., et al., *Epigenome-wide association study of smoking and DNA methylation in non-small cell lung neoplasms.* Oncotarget, 2016. **7**(43): p. 69579-69591.

156. Brans, R., et al., *Association between tobacco smoking and prognosis of occupational hand eczema: a prospective cohort study.* Br J Dermatol, 2014. **171**(5): p. 1108-15.

157. Patel-King, R.S., et al., *WD60/FAP163 is a dynein intermediate chain required for retrograde intraflagellar transport in cilia.* Mol Biol Cell, 2013. **24**(17): p. 2668-77.

158. Gholkar, A.A., et al., *Tctex1d2 associates with short-rib polydactyly syndrome proteins and is required for ciliogenesis.* Cell Cycle, 2015. **14**(7): p. 1116-25.

159. Waters, A.M. and P.L. Beales, *Ciliopathies: an expanding disease spectrum.* Pediatr Nephrol, 2011. **26**(7): p. 1039-56.

160. Lehman, J.M., et al., *An essential role for dermal primary cilia in hair follicle morphogenesis.* J Invest Dermatol, 2009. **129**(2): p. 438-48.

161. Croyle, M.J., et al., *Role of epidermal primary cilia in the homeostasis of skin and hair follicles.* Development, 2011. **138**(9): p. 1675-85.

162. Reilly, D.M., et al., *The epidermal nerve fibre network: characterization of nerve fibres in human skin by confocal microscopy and assessment of racial variations.* Br J Dermatol, 1997. **137**(2): p. 163-70.

163. Guo, J., et al., *Primary Cilia Signaling Shapes the Development of Interneuronal Connectivity.* Dev Cell, 2017. **42**(3): p. 286-300 e4.

164. An, Y., K.L. Furber, and S. Ji, *Pseudogenes regulate parental gene expression via ceRNA network.* J Cell Mol Med, 2017. **21**(1): p. 185-192.

165. Yu, G., et al., *Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression.* Mol Cancer Ther, 2014. **13**(12): p. 3086-97.

166. Dilip., J. and S. Binorkar, *Traditional Medicinal Usage of Tobacco – A Review.* Spatula DD, 2012. **2**.

167. Davies, P., et al., *Acute nicotine poisoning associated with a traditional remedy for eczema.* Arch Dis Child, 2001. **85**(6): p. 500-2.

168. Hellstrom-Lindahl, E. and A. Nordberg, *Smoking during pregnancy: a way to transfer the addiction to the next generation?* Respiration, 2002. **69**(4): p. 289-93.

169. Xanthoulea, S., et al., *Nicotine effect on inflammatory and growth factor responses in murine cutaneous wound healing.* Int Immunopharmacol, 2013. **17**(4): p. 1155-64.

170. Li, V.S., et al., *Wnt signaling through inhibition of beta-catenin degradation in an intact Axin1 complex.* Cell, 2012. **149**(6): p. 1245-56.

Appendix

Supplementary Table A1: Association of exclusive breastfeeding duration with prevalence of eczema (measured repeatedly at ages 1-or-2, 4, 10, and 18 years) stratified by maternal smoking during pregnancy.

Parameter	Levels	Maternal smoking during pregnancy			
		NO (n=3011)		**YES** (n=839)	
		Risk ratio (95% CI)	**p-value**	**Risk ratio (95% CI)**	**p-value**
Duration of exclusive breastfeeding		1.00 (0.99, 1.01)	0.75	0.97 (0.94, 0.99)	0.04
	0 weeks	-	-	-	-
	3 weeks	-	-	0.91 (0.84, 0.99)	0.04
	9 weeks	-	-	0.76 (0.58, 0.98)	0.04
	15 weeks	-	-	0.63 (0.41, 0.97)	0.04
	21 weeks	-	-	0.52 (0.29, 0.96)	0.04
FLG loss of function variants		1.5 (1.06, 2.06)	0.02	1.99 (1.02, 3.90)	0.04
Time	Year 1/ 2	Ref.	Ref.	Ref.	Ref.
	Year 4	0.76 (0.63, 0.92)	0.006	0.66 (0.4, 1.08)	0.10
	Year 10	0.94 (0.78, 1.13)	0.51	0.98 (0.64, 1.48)	0.90
	Year 18	0.74 (0.60, 0.92)	0.007	0.86 (0.54, 1.39)	0.55
Weeks of gestation		1.1 (1.0, 1.23)	0.09	1.12 (0.96, 1.33)	0.15
Maternal eczema status		1.3 (0.97, 1.85)	0.07	1.17 (0.61, 2.24)	0.62
Paternal eczema status		1.87 (1.32, 2.63)	0.0004	1.24 (0.54, 2.86)	0.61
Gender	Male	0.89 (0.70, 1.14)	0.36	1.06 (0.64, 1.76)	0.81

Socio economic status not added in the model since it did not have enough variability in the smoking='Yes' group

Supplementary Table A2: Assessing interaction of maternal smoking during pregnancy with residuals of duration of exclusive breastfeeding on eczema at ages 1-or-2, 4, 10, and 18 adjusted for covariates (n=3777).

Parameter	Levels	Risk ratio (95% CI)	p-value
Maternal smoking during pregnancy		0.8 (0.60, 1.08)	0.15
Duration of exclusive breastfeeding		1.0(0.99, 1.01)	0.86
Maternal smoking during pregnancy × Residuals of duration of exclusive breastfeeding[#*]	0 weeks	0.81 (0.60, 1.08)	0.15
	3 weeks	0.74 (0.54, 1.0)	0.057
	9 weeks	0.61 (0.40, 0.93)	0.02
	15 weeks	0.51 (0.29, 0.88)	0.017
	21 weeks	0.42 (0.21, 0.86)	0.018
FLG loss of function variants		1.57 (1.16, 2.13)	0.003
Time	Year 1/ 2	Ref.	Ref.
	Year 4	0.74 (0.62, 0.88)	0.0012
	Year 10	0.95 (0.80, 1.12)	0.54
	Year 18	0.76 (0.63, 0.93)	0.008
Weeks of gestation		1.10 (1.00, 1.21)	0.03
Maternal eczema status		1.31 (0.98, 1.76)	0.06
Paternal eczema status		1.74 (1.26, 2.41)	0.0008
Gender (Female is the reference)	Male	0.93 (0.75, 1.17)	0.57
Socio economic status (High socioeconomic status is the reference)	Mid	1.09 (0.69, 1.72)	0.70
	Low	1.06 (0.77, 1.46)	0.68

Supplemental Table A3: Configural frequency analysis with complete data of eczema (no missing at any time point) at age 1-or-2, 4, 10, and 18 (n=1006)

Age 1-2	**Age 4**	**Age 10**	**Age 18**	**Observed Frequency (Fo)**	**Expected Frequency (Fe)**	**p-value**	**Classification into *Types* and *Antitypes****
-	-	-	-	701	573.202	0	Eczema free
-	-	-	+	42	74.709	4.19E-05	Antitype: Late onset
-	-	+	-	38	92.665	0	Antitype: at age 10 only
-	-	+	+	19	12.078	0.02254	
-	+	-	-	22	71.09	0	Antitype: At age 4 only
-	+	-	+	5	9.266	0.079588	
-	+	+	-	16	11.493	0.09057	
-	+	+	+	3	1.498	0.109682	
+	-	-	-	67	108.407	1.28E-05	Antitype: At age 1&2 only
+	-	-	+	7	14.129	0.028058	
+	-	+	-	14	17.525	0.197789	
+	-	+	+	7	2.284	0.000893	Type: Persistent eczema
+	+	-	-	16	13.445	0.241482	
+	+	-	+	6	1.752	0.00066	Type: Persistent eczema
+	+	+	-	16	2.174	0	Type Persistent eczema
+	+	+	+	27	0.283	0	Type: Persistent eczema

*Antitypes: patterns that occur significantly less often than by chance

Supplementary Table A4: Assessing interaction of maternal smoking during pregnancy and during lactation, with duration of exclusive breastfeeding on eczema at ages 1-or-2, 4, 10, and 18 adjusted for covariates (n=3780)

Parameter	Levels	Risk ratio (95% CI)	p-value
Maternal smoking during pregnancy and/or lactation		0.99 (0.73, 1.33)	0.95
Duration of exclusive breastfeeding		1.0 (0.99, 1.01)	0.46
Maternal smoking during pregnancy and/or lactation × Duration of exclusive breastfeeding[#*]	0 weeks	0.99 (0.74, 1.33)	0.95
	3 weeks	0.92 (0.71, 1.2)	0.55
	9 weeks	0.8 (0.63, 1.02)	0.07
	15 weeks	0.69 (0.52, 0.92)	0.01
	21 weeks	0.60 (0.42, 0.87)	0.007
FLG loss of function variants		1.58 (1.17, 2.13)	0.0025
Time	Year 1/ 2	Ref.	Ref.
	Year 4	0.74 (0.62, 89)	0.0012
	Year 10	0.95 (0.80, 1.12)	0.54
	Year 18	0.77 (0.63, 0.93)	0.007
Weeks of gestation		1.11 (1.01, 1.22)	0.02
Maternal eczema status		1.30 (0.98, 1.75)	0.07
Paternal eczema status		1.7 (1.23, 2.37)	0.0014
Gender (Female is the reference)	Male	0.94 (0.76, 1.17)	0.62
Socio economic status (High socioeconomic status is the reference)	Mid	1.0 (0.68, 1.45)	0.99
	Low	0.94 (0.59, 1.49)	0.99

Institutional Review Board
Office of Sponsored Programs
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

Aug 1, 2017

PI Name: Nandini Mukherjee
Co-Investigators:
Advisor and/or Co-PI: Wilfried Karmaus
Submission Type: Initial
Title: Role of breastfeeding, smoking, and DNA methylation in the risk of eczema
IRB ID **:** #PRO-FY2017-594

Expedited Approval: Jul 28, 2017
Expiration: Jul 28, 2018

Approval of this project is given with the following obligations:

1. This IRB approval has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.

2. When the project is finished or terminated, a completion form must be submitted.

3. No change may be made in the approved protocol without prior board approval.

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair
The University of Memphis.