illumina®

Human Epigenome-Wide Association Studies

Infinium® HumanMethylation450 BeadChip identifies differential DNA methylation in newborns exposed to maternal smoking during pregnancy.

Introduction

In recent years, the approach to understanding genetic factors in complex disease has relied largely on genome-wide association studies (GWAS). The tremendous success of GWAS has been propelled—in large part—by advances in microarray and sequencing technologies. Since 2005, high-throughput microarrays have supported the identification of ~2000 SNP associations with more than 300 complex diseases.[1] While GWAS has reliably identified thousands of SNP-disease associations, it is poorly equipped to reveal the specific, molecular mechanisms of complex diseases. Therefore, investigators have increasingly turned to epigenome-wide association studies (EWAS) to explore how methylation states play a role in the etiology of complex disease phenotypes.[2,3] Additionally, advances in DNA methylation chips—such as increasing throughput capacity and complexity of coverage—have enabled investigators to identify significantly smaller effect size associations. While early EWAS focused primarily on cancer research,[4,5] these technology advances have fueled the expansion of research into a broad range of heritable diseases such as type 2 diabetes, obesity, and autoimmunity.[6,7]

This application note describes the methods and results of an EWAS relating maternal smoking during pregnancy to differential DNA methylation patterns in infants.[8] The study used the Infinium HumanMethylation450 BeadChip, an assay which allows researchers to interrogate > 485,000 methylation sites per sample at single-nucleotide resolution. This BeadChip covers 99% of RefSeq gene regions, 96% of CpG islands, as well as CpG dinucleotides outside of islands and promoter regions. The BeadChip protocol requires 500 ng of bisulfite-treated DNA per sample and can process 12 samples per chip. The kit also includes a protocol for processing formalin-fixed paraffin-embedded (FFPE) samples, making it an ideal choice for EWAS of tumor samples.

Experimental Design

Types of EWAS

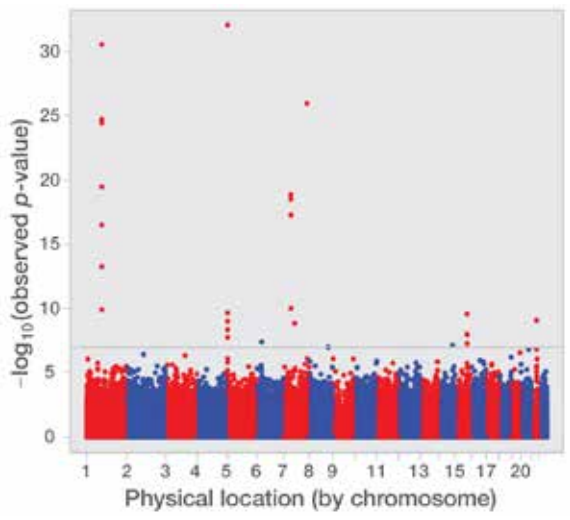
Current EWAS designs fall into one of four common categories:

- *Retrospective studies* include unrelated individuals who are recruited into case and control groups based on observed phenotypes. Retrospective studies take advantage of existing gene expression databases that can be integrated with epigenetic data; however, they cannot control for environmental or treatment-related factors that may cause epigenetic variation.
- *Disease-discordant monozygotic twin studies* can control for genetic factors associated with a disease of interest. However, it can be difficult to recruit large cohorts to provide adequate statistical power.
- *Parent-offspring pair studies* can identify epigenetic markers that are transmitted across generations. They can integrate genomic and epigenomic profiling to determine whether parental environment influences phenotypes in offspring. However, as with twin studies, it can be difficult to recruit adequately large cohorts.
- *Longitudinal cohort studies* follow individuals who are initially disease-free over many years. They minimize the effects of confounders due to differences between cases and controls. These studies are ideal for observing the progression of epigenetic changes over time. However, they are slow and expensive to conduct.

Methods

The study analyzed 1,062 parent-offspring pairs from the Norwegian Mother and Child Cohort Study (MoBa).9,10 The MoBa cohort was assembled to examine the association between maternal plasma folate and childhood asthma status at three years of age. The authors of the current study used blood cotinine as a marker for maternal smoking. A replication analysis was also performed as part of the newborn epigenetics study (NEST) in Durham, NC.11 The replication study used cord blood DNA samples from 18 newborns of mothers

Figure 1: Epigenome-Wide Association Between Maternal Cotinine and CpG Methylation



Methylation of 473,844 CpGs was measured in cord blood from the MoBa cohort. Twenty-six CpGs (10 genes) reached Bonferroni-corrected statistical significance ($p < 1.06 \times 10^{-7}$, represented by the horizontal line). Red and blue alternating colors are used to distinguish between chromosomes.

Table 1: Differential Methylation in Cord Blood DNA in Relation to Maternal Cotinine

			Unadjusted			Median methylation by cotinine category[d]			
Chr[a]	Gene	CpG	Coef[b]	SE[c]	*p*-Value	Undetectable	Low	Medium	High
1	*GFI1*	cg10399789	−0.07	0.01	4.08E-13	0.759	0.755	0.727	0.716
1	*GFI1*	cg09662411	−0.111	0.012	2.26E-20	0.730	0.733	0.669	0.654
1	*GFI1*	cg06338710	−0.112	0.013	1.34E-18	0.801	0.800	0.754	0.733
1	*GFI1*	cg18146737	−0.28	0.024	2.42E-30	0.877	0.875	0.771	0.738
1	*GFI1*	cg12876356	−0.182	0.016	2.29E-30	0.731	0.732	0.627	0.605
1	*GFI1*	cg18316974	−0.243	0.024	6.43E-24	0.921	0.923	0.856	0.841
1	*GFI1*	cg09935388	−0.196	0.015	1.05E-38	0.708	0.707	0.580	0.564
1	*GFI1*	cg14179389	−0.184	0.017	5.38E-28	0.242	0.246	0.154	0.158
5	*AHRR*	cg23067299	0.075	0.012	4.21E-10	0.789	0.789	0.813	0.837
5	*AHRR*	cg03991871	−0.057	0.008	2.04E-11	0.841	0.839	0.820	0.818
5	*AHRR*	cg05575921	−0.202	0.015	2.85E-39	0.883	0.874	0.829	0.784
5	*AHRR*	cg21161138	−0.045	0.007	1.52E-11	0.718	0.715	0.701	0.679
6	*HLA-DPB2*	cg11715943	−0.053	0.009	1.00E-08	0.842	0.833	0.824	0.820
7	*MYO1G*	cg19089201	0.083	0.013	3.22E-10	0.925	0.926	0.932	0.944
7	*MYO1G*	cg22132788	0.18	0.021	1.98E-18	0.932	0.935	0.951	0.966
7	*MYO1G*	cg04180046	0.073	0.008	8.76E-20	0.441	0.446	0.484	0.508
7	*MYO1G*	cg12803068	0.145	0.016	8.51E-19	0.713	0.721	0.774	0.813
7	*ENSG00000225718*	cg04598670	−0.063	0.009	1.29E-11	0.623	0.607	0.597	0.574
7	*CNTNAP2*	cg25949550	−0.075	0.007	4.15E-30	0.113	0.109	0.097	0.092
8	*EXT1*	cg03346806	−0.038	0.007	3.08E-08	0.801	0.795	0.793	0.779
14	*TTC7B*	cg18655025	−0.041	0.007	2.07E-08	0.854	0.847	0.841	0.836
15	*CYP1A1*	cg05549655	0.064	0.01	2.96E-10	0.189	0.188	0.221	0.226
15	*CYP1A1*	cg22549041	0.096	0.016	4.52E-09	0.385	0.379	0.414	0.475
15	*CYP1A1*	cg11924019	0.044	0.008	2.62E-08	0.434	0.430	0.457	0.475
15	*CYP1A1*	cg18092474	0.066	0.012	1.10E-08	0.510	0.504	0.549	0.573
21	RUNX1	cg12477880	0.159	0.026	1.02E-09	0.088	0.102	0.110	0.158

Methylation was measured in the MoBa study population. CpGs with Bonferroni-corrected statistical significance ($p < 1.06 \times 10^{-7}$) are listed, sorted by chromosome and position. [a]Chromosome. [b]Regression coefficient. [c]Standard error for regression coefficient. [d]Maternal plasma cotinine (nmol/L) measured around gestational week 18 (undetectable ≤ 0; low > 0–56.8; moderate > 56.8–388; high > 388). Values > 56.8 nmol/L indicate active smoking.

with self-reported smoking during pregnancy and 18 controls with no maternal smoking reported. Bisulfite conversion was performed for all DNA samples using commercially available kits (Zymo Research). Methylation status was measured at 485,577 CpGs in cord blood using the Infinium HumanMethylation450 BeadChip. Illumina GenomeStudio® Methylation Module v1.0 software was used to calculate the methylation level at each CpG as a beta value.12 The analysis reported the detection *P* value for each beta, representing the difference between the signal for a given probe and background (where background is calculated as the average for all negative controls).

Analysis and Results

Data Analysis

To facilitate analysis of results obtained from the HumanMethylation450 BeadChip assay, Illumina provides two sources of SNP data: the HumanMethylation450 BeadChip manifest file and a more comprehensive Supplementary SNP list that is updated every 3 months. These resources can be accessed from the MyIllumina customer portal. Information from the Supplementary SNP list can be imported directly into the methylation module of GenomeStudio analysis software.

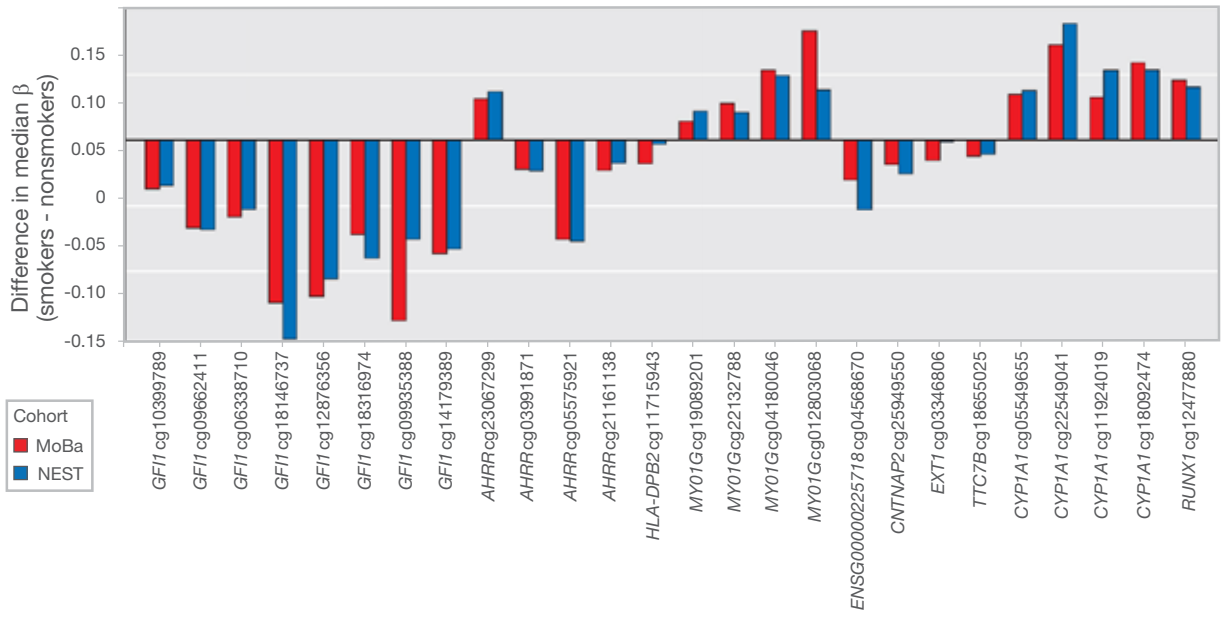
The final statistical model in this study included variables that were associated with cotinine ($p < 0.1$) and possibly related to methylation levels, such as maternal age, maternal education, and parity. The results were adjusted for childhood asthma status at three years of age, although this made little difference in the results.

All NEST samples and CpGs passed quality control. The researchers used unadjusted linear regression models to examine the association between maternal smoking during pregnancy and methylation in newborn cord blood at each of the 26 CpGs that were significantly associated with plasma cotinine ($p < 1.06 \times 10^{-7}$) in the MoBa population, and they calculated a one-sided *p* value for each CpG. After applying a Bonferroni correction for 26 tests, the level of significance was adjusted to 0.0019.

To assess the potential impact of variation according to white blood cell subtype, the researchers measured DNA methylation using the HumanMethylation450 BeadChip in 21 cord blood samples collected at the same facilities as the NEST samples that had been separated, while fresh, into mononuclear cells (MN) and polymorphonuclear cells (PM) using Lympholyte®-poly (Cedarlane Laboratories Limited, Hornby, Ontario). A paired *t*-test was used to evaluate differential methylation between PM and MN cell types for the top 26 CpGs.

The researchers employed a single CpG lookup approach to compare these results with those from other methylation studies. A CpG with a *p* value < 0.05 was considered to be statistically significant. All

Figure 2: Replication Analysis of Differential Methylation by Smoking Status



The difference in median methylation intensity (β) by smoking status among MoBa and NEST participants is shown for the top 26 CpGs identified, grouped by gene.

statistical analyses were performed using R (R Development Core Team 2010) and Bioconductor[13] packages.

Epigenome-Wide and Replication Analyses Results

The researchers observed statistically significant associations between maternal smoking during pregnancy and altered methylation levels in offspring based on maternal plasma cotinine levels and methylation in cord blood. These changes occured in 26 CpGs that mapped to 10 genes in the MoBa cohort. These CpGs met strict Bonferroni-corrected statistical significance ($p < 1.06 \times 10^{-7}$; Figure 1). In the NEST cohort, despite the small sample size (n = 18), researchers found a striking degree of replication, with estimates for 21 of the 26 identified CpGs showing *P* values < 0.05 (Table 1). Five of the CpGs met the criteria for strict Bonferroni-corrected statistical significance. Two of the CpGs were located in *CYP1A1* and one in *AHRR*, genes known to be involved in detoxification pathways for tobacco smoke. In addition, two CpGs were identified in *GFI1*, a gene that has not been previously implicated in a response to tobacco smoke. The magnitude of the differences between smokers and non-smokers in NEST, and those between women with plasma cotinine > 56.8 nmol/L versus ≤ 56.8 nmol/L in MoBa, were very similar (Figure 2).

In this study, maternal smoking also displayed a dose-dependent association with lower methylation of *AHRR* CpGs and higher methylation of *CYP1A1* CpGs in newborn cord blood. These opposite effects are biologically relevant, because the two genes have opposing functions in the AhR pathway.[14] Notably a previous study—which also used the HumanMethylation450 BeadChip—showed lower methylation levels at the same CpGs in adult smokers.[15] Thus, the data demonstrate that functionally important methylation changes in some adult smokers may already be present at birth due to maternal smoking during pregnancy.

The authors report that the HumanMethylation450 BeadChip offers greatly improved coverage over the previous HumanMethylation27 BeadChip: none of the 26 CpGs identified in this study were present on the HumanMethylation27 BeadChip.

Future Directions

Early EWAS focused primarily on highly pronounced epigenetic modifications associated with cancer. As this study demonstrates, EWAS can also offer significant insight into lower effect size epigenetic modifications occurring *in utero*. The increasing availability of CpG methylation information and associated high-resolution methylation maps for various tissue types will enable EWAS for many diseases. The efforts of the International Human Epigenome Consortium (IHEC) will play a key role in this regard.

Integration of EWAS and GWAS data will also enable a powerful view into the underlying causes of complex diseases. The

HumanMethylation450 BeadChip covers 99% of RefSeq genes, 20,000 of which are represented in Illumina's HumanHT-12 v4 Expression BeadChip. A study analyzing SNPs, gene expression, and DNA methylation in 77 HapMap cell lines identified SNPs that affect both gene expression and DNA methylation.[16] Further studies pairing GWAS with EWAS will provide valuable insight into the role of genetic variation in complex diseases.

Conclusions

The information derived from EWAS, facilitated by technologies such as Illumina Infinium HumanMethylation450 BeadChip, provides a fresh look at genetic variation that cannot be obtained from GWAS alone. Recent advances in methylation arrays are enabling the study of epigenetics in a growing range of complex diseases such as diabetes, autoimmunity, autism, obesity, and more.[17] Further integration of EWAS with GWAS—as well as the increased availability of biobanks for well-designed, longitudinal cohort studies—will greatly contribute to the understanding of disease-associated genetic variations.

Learn More

To learn more about epigenetics and array-based methylation analysis, go to www.illumina.com/applications/epigenetics.ilmn

To learn more about the Infinium HumanMethylation450 BeadChip, go to www.illumina.com/products/methylation_450_beadchip_kits.ilmn

References

1. Manolio TA. (2013) Bringing genome-wide association findings into clinical use. Nat Rev Genet 2013 14(8):549-58.
2. Feinberg AP and Irizarry RA. (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. PNAS 107: 1757–1764.
3. Petronis A. (2010) Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 465: 721–72.
4. Kullis M and Esteller M. (2010) DNA methylation and cancer. Adv Genet 70: 27–56.
5. Lechner M, Boshoff C, and Beck S. (2010) Cancer epigenome. Adv Genet 70: 247–276.
6. Bach JF. (2002) The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med 347: 911–920.
7. Bell CG, Finer S, Lindgren CM, Wilson GA, Raykan VK, et al. (2010) Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PLoS ONE 5: e14040.
8. Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, et al. (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 120(10): 1425–31.
9. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, et al. (2006) Cohort profile: the Norwegian mother and child cohort study (MoBa). Int J Epidemiol 35(5): 1146–1150.
10. Ronningen KS, Paltiel L, Meltzer HM, Nordhagen R, Lie KK, et al. (2006) The biobank of the Norwegian mother and child cohort study: a resource for the next 100 years. Eur J Epidemiol 21 (8): 619–625.
11. Hoyo C, Murtha AP, Schildkraut JM, Forman MR, Calingaert B, et al. (2011) Folic acid supplementation before and during pregnancy in the newborn epigenetics study (NEST). BMC Public Health 11(1): 46.
12. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288-295.
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5(10): R80.
14. Kawajiri K and Fujii-Kuriyama Y. (2007) Cytochrome P450 gene regulation and physiological functions mediated by the aryl hydrocarbon receptor. Arch Biochem Biophys 464: 207–212.
15. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, et al. (2012) Coordinated changes in AHRR methylation in lymphoblasts and pulmnary macrophages from smokers. Am J Genet B Beuropsychiatr Genet 159B(2): 141–151.
16. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, et al. (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 12: R10.
17. Rakyan VK, Down TA, Balding DJ, and Beck S. (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12(8): 529-541. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3508712/

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 270-2013-002 Current as of 27 November 2013

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA