Paediatric Respiratory Reviews 15 (2014) 348–354

Contents lists available at ScienceDirect

Paediatric Respiratory Reviews

Review

An update on epigenetics and childhood respiratory diseases

Olena Gruzieva [1], Simon Kebede Merid [1], Erik Melén [1,2,*]

[1] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[2] Sachs' Children's Hospital, Stockholm, Sweden

EDUCATIONAL AIMS:

The reader will come to appreciate that:

- Epigenetic mechanisms are an important potential link between genetic and environmental factors that influence complex traits.
- Children are particularly sensitive to the adverse effects of certain exposures predisposing to respiratory disease [e.g. diet, tobacco smoke, stress] and that these effects could be more profound if experienced in utero or early life.
- Epigenetics plays an important role in regulating the expression of genes involved in the immune-mediated inflammatory response.
- Methylation is the most widely investigated epigenetic mechanism to explain how gene-environment interactions may contribute to asthma and atopy.

ARTICLE INFO

Keywords:
Asthma
Children
Genetics
Epigenetics
Lung development
Methylation
Respiratory disease

SUMMARY

Epigenetic mechanisms, defined as changes in phenotype or gene expression caused by mechanisms other than changes in the underlying DNA sequence, have been proposed to constitute a link between genetic and environmental factors that affect complex diseases. Recent studies show that DNA methylation, one of the key epigenetic mechanisms, is altered in children exposed to air pollutants and environmental tobacco smoke early in life. Several candidate gene studies on epigenetics have been published to date, but it is only recently that global methylation analyses have been performed for respiratory disorders such as asthma and chronic obstructive pulmonary disease. However, large-scale studies with adequate power are yet to be presented in children, and implications for clinical use remain to be evaluated. In this review, we summarize the recent advances in epigenetics and respiratory disorders in children, with a main focus on methodological challenges and analyses related to phenotype and exposure using global methylation approaches.

INTRODUCTION

Technology development within the –omics field of genomics, epigenetics, transcriptomics and proteomics has enabled large-scale analyses of unprecedented speed and precision. These methods offer the potential to identify "biomarker fingerprints" that may improve the understanding of molecular mechanisms of disease, identify new disease pathways and predict models of complex diseases possibly leading to personalized medicine. Because access to lung tissue and bronchoalveolar lavage fluid is very limited in children, measurements of biomarkers in blood, urine and other bio-fluids have received increasing attention lately. Today, state-of-the-art analyses on disease mechanisms related to genetics and heredity include single-level analyses of DNA variants, epigenetics changes, transcriptomics and proteomics. More particularly, an integrative genomics approach needs to be applied where all these levels are considered, from gene to protein [1].

Epigenetic mechanisms, defined as changes in phenotype or gene expression caused by mechanisms other than changes in the

* Corresponding author. Institute of Environmental Medicine, Karolinska Institutet, Nobels väg 13, 171 77 Stockholm, Sweden. Tel.: +46 8 524 875 08.
E-mail address: erik.melen@ki.se (E. Melén).

http://dx.doi.org/10.1016/j.prrv.2014.07.003
1526-0542/

underlying DNA sequence, is one of the latest emerging –omics areas. High expectations remain for epigenetic mechanisms being an important link between genetic and environmental factors that affect complex traits, including respiratory diseases. As such, epigenetics has been suggested as a mediating factor that associates environmental exposures to childhood asthma-related phenotypes [2]. Since epigenetic alterations are inheritable through cell division and may be transferred to future generations, it is very important to clarify their role in the association between environmental risk factors and disease.

Several candidate gene studies on epigenetics have been published to date. It is only recently that global methylation analyses have been performed for respiratory disorders such as asthma and chronic obstructive pulmonary disease (COPD) [3,4]. However, large-scale studies with adequate power are yet to be presented in children and implications for clinical use remain to be evaluated. It has become evident that a rigorous quality control (QC) protocol needs to be applied before further analysis of global methylation data can take place, and the challenges therein cannot be underestimated.

In this review, we summarize the recent advances in epigenetics and respiratory disorders in children, and outline a roadmap for future studies. Epigenetics is often used as a broad term that includes mechanisms related to DNA methylation, posttranslational histone modifications and chromatin remodeling, and microRNAs. Most of the published epigenetic studies on children to date concern analyses of DNA methylation (binding of a methyl group to the DNA, usually where Cytosine (C) lies adjacent to Guanine (G), termed a CpG site). Given the rapid technology development with several global methylation studies recently published, and many more to come within the next few years, we focus here primarily on methodological challenges and analyses related to both phenotype and exposure using global methylation approaches.

METHODS FOR ANALYZING METHYLATION IN RELATION TO DISEASE OR EXPOSURE

To date, the most common global methylation array on the market is the Illumina 450 Infinium array, and we will therefore mainly focus the QC section on issues related to this platform. The Illumina 450 array covers over 450,000 methylation data points at CpG sites spread across the whole genome, thereby enabling unbiased analyses of methylation status, disease and exposure.

Quality control (QC) of DNA samples

Similar to large-scale gene-expression and GWAS analyses, a rigorous QC protocol needs to be applied before further analysis of the methylation data can take place (Figure 1). The Infinium arrays include several control probes to detect poorly performing samples. In addition, control DNA samples with known methylation profiles are also included in the analyses, to ensure an overall high quality of the data. Different diagnostic plots of control probes can be generated in the Genome-Studio software [5]. Samples with poor raw signal intensity of the control probes (±3 SD from the median) are usually filtered out.

Probe filtering

The methylation levels are affected by probes targeting CpG loci, which include single nucleotide polymorphisms (SNPs) near or within the probe sequence or even in the target CpG dinucleotide. Typically, the investigators want to remove CpG sites that are directly influenced by genotype. The genetic or environmental interactions that affect the population-specific DNA methylation levels may in fact be due to ethnicity-specific genetic

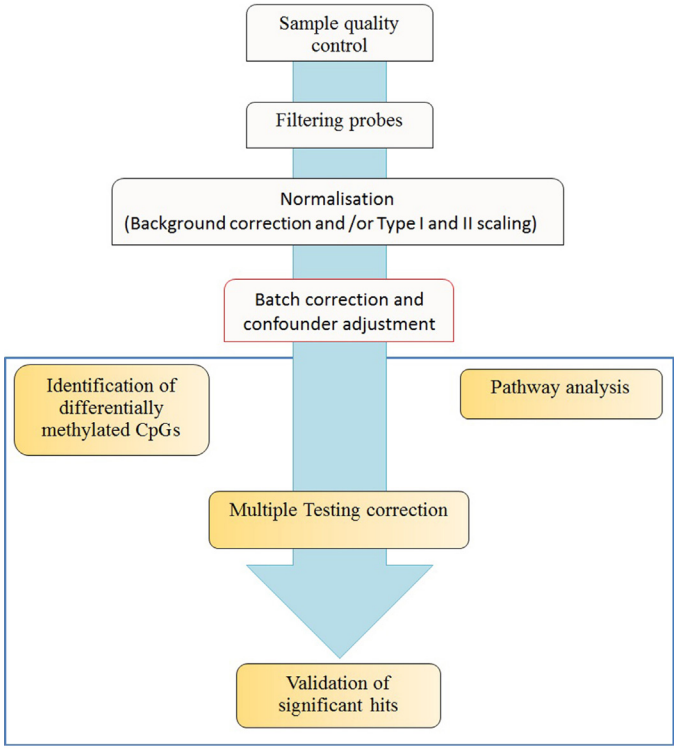

Figure 1. Overview of quality control (QC) pipeline for global methylation analyses.

variants. Additional probes that are usually removed include spurious cross-hybridisations and probes on the sex chromosomes (unless specifically under study). Probes with very low variation and extreme methylation levels (i.e., median=0% or median=100%) are filtered out. Additionally, the Illumina protocol recommends excluding probes that have a detection *P*-value (defined as 1 − *P*-value obtained from the model evaluating the chance that the signal was differentiated from negative controls) higher than 0.05 in a certain proportion of the samples. However, these filtering thresholds vary across the studies [6].

Normalisation (background correction and /or type I and II scaling)

Experimental artefacts, random noise and technical and systematic variation in microarray-based analyses may mask the true biological differences, or give rise to false-positive findings. Most sources of these technical variations are removed by a normalisation procedure [7]. There are two types of normalization protocols typically applied in global methylation studies; 1) between-array normalisation and 2) within-array normalisation. CpG sites and DNA methylation patterns are not randomly distributed through the genome and there is a link between CpG density and DNA methylation (i.e., CpG islands - regions with a high frequency of CpG sites - are often unmethylated whereas a non-CpG islands are methylated). Moreover, DNA methylation is inversely related to total fluorescence signal that complicates normalisation [8].

Genome-Studio provides an internal control normalisation method for global methylation data generated by the Illumina 450K array [9]. Different normalisation pipelines for Illumina 450K data have recently been evaluated systematically, including quantile normalisation, DASEN and SWAN methods [10]. One of the most commonly used normalisation techniques in methylation studies is quantile normalisation. It assumes similar total signals across samples, which may confer a limitation for global analyses when probes are examined at different subsets of the genome.