1. Motivation

Eosinophils are white blood cells that have a variety of functions. They constitute an important immune system component responsible for combating multicellular parasites and certain infections in vertebrates. Eosinophils are a type of granulocyte cell, which have granules present in the cytoplasm of the cells and have varying shapes of nuclei. In addition to eosinophils, there are three other types of granulocyte cells: neutrophils (the most abundant of the granulocytes), basophils, and mast cells. Eosinophils change in response to environmental exposures (Cardenas, et al., 2015), and are associated with parasitic infections and some inflammatory diseases, such as allergies and asthma (Berek, 2016). Correct identification of eosinophils and estimation of their proportions in biological samples are thus crucial in many studies.

Umbilical cord blood, samples collected after birth without maternal contamination, have been used extensively in biological research and assessment of early life disease risk in recent years. The proportion of eosinophils relative to other cells in cord blood has the ability to predict respiratory illnesses in high risk infants: elevated cord eosinophils increased the risk for wheezing in children two years or younger in a sub-cohort of the LINA study (Junge, et al., 2014). In a prospective cohort study of 917 mother-child pairs in South Korea, higher cord blood eosinophils were associated with lower serum folate levels in mothers during mid-pregnancy. In turn, lower levels of serum folate levels were associated with elevated risks of lower respiratory tract infections and atopic dermatitis (Kim, et al., 2015). These findings indicate the importance of understanding inter-individual heterogeneity in eosinophils in cord blood for childhood health.

Several laboratory methods have been proposed to estimate eosinophilic counts in biological samples including staining with eosin followed by cell counting or using antibodies and counting by flow cytometry (Ethier, et al., 2014; Lavigne, et al., 1997; Thurau, et al., 1996). The proportion of eosinophils is usually quite low, less than 2% in blood. Hence, cell counting for eosinophils requires large amounts of fresh blood samples to process or sort sufficient eosinophils. However, the total amount of cord blood samples is often limited, or if the sample was previously frozen, resulting rupturing of the cell membranes could make cell counting inaccurate. In addition, discrimination of eosinophils from other granulocytes can also be difficult (Ethier, et al., 2014). Taken together, direct cell counting is not feasible in many settings.

# More recently, an indirect method for estimating cell composition, which uses DNA methylation (DNAm) profiles to infer cell proportions, has been developed and widely used. The underlying principle of the indirect method is in the uniqueness of DNAm profiles for each individual cell type (Houseman, et al., 2015). Using the *minfi* package in Bioconductor for R statistical software, we are now able to estimate eosinophil cell compositions in blood of adults using reference databases derived from peripheral blood of six adult men (Aryee, et al., 2014; Fortin, et al., 2017). Sorted cord blood cell methylation reference panels are available, however, they lack DNAm information for eosinophils (Bakulski, et al., 2016). Overall, the R package *minfi* is able to estimate proportions of eosinophils in peripheral blood samples found in males, but not for samples from cord blood and other tissues (Aryee, et al., 2014).

# 2. Prior distributions for individual-level measurement error and systematic error

The parameters and determine the size of measurement errors. Thus, the choice of hyper-prior parameters *a* and *b* in the prior of , and *c* and *d* in the prior of is critical and should reasonably represent the range of errors. We select these parameters using information from cell types with both cord blood references and adult blood references available.

In total, six types of cells have both adult blood reference and cord blood references: B cells, CD4+ T cells, granulocytes, monocytes, CD8+ T cells, and natural killer cells (abbreviated as Bcell, CD4T, Gran, Mono, CD8T and NK, respectively). These references use multiple DNAm markers that characterize these cells. Therefore, using the existing method available in the *minfi* package (in the function *estimateCellCounts*), we are able to infer their cell type proportions for each sample based on DNAm cord blood references (this gives for a cell) as well as based on adult blood references (this gives for that cell). Consequently, we are able to estimate the magnitude of measurement errors of each sample for each of the six cells (by calculating the difference between and ). These measurement errors are then used to determine the values of *a* and *b* in the uniform (*a*, *b*). In particular, we estimate for each of the six cell types by calculating the variance of (, denoted as , , …, , from which, we choose *a*=min (, , …, ) and *b*=max(, , …, ), assuming the size of measurement errors for eosinophil cells is not far from the size of measurement errors for other cell types.

Systematic error () in our situation refers to bias on average with respect to the estimation of logit-transformed cell type proportion for eosinophil cells. To objectively and informatively specify *c* and *d* in its prior distribution, we utilize two strategies. First, we use information on cells that are in the same or similar category of eosinophils. Here we choose granulocytes as eosinophilic cells are one of this type. Second, we take into account variations between different studies by utilizing information from previous studies. In particular, we use cord blood DNAm profiles analyzed using Illumina 450K from six independent studies. The first is the Isle of Wight (IOW) cohort study, which was initiated at the David Hide Asthma and Allergy Research Centre on the Isle of Wight, UK by Dr. Arshad in 1989/1990 (Arshad, et al., 2018; Hide, et al., 1991). About 320 offspring of the cohort were recruited between 2010 and 2015, and cord blood samples of 129 participants were analyzed for DNAm and used in the current study (Arshad, et al., 2017). The second study is 175 cord blood samples from the Early Autism Risk Longitudinal Investigation (EARLI) (Newschaffer, et al., 2012). The third dataset is 15 whole cord blood samples available publicly in the R package “FlowSorted.CordBlood.450k” (the Hopkins study) (Bakulski, et al., 2016). These were convenience cord blood samples collected in 2015 from the Johns Hopkins University Hospital. Additional aliquots of these samples were used for the sorting experiments to create the cord blood reference panel. The fourth study is 90 National Children’s Study Initial Vanguard cord blood samples, a pilot study involving mothers and infants recruited from seven different locations across the USA (Duplin County, NC; Queens County, NY; Orange County, CA; Waukesha County, WI; Salt Lake County, UT; Montgomery County, PA; and, a composite location of four adjacent counties in South Dakota and Minnesota) and cord bloods were collected between 2009 and 2011 (Bakulski, et al., 2015). The fifth dataset is extracted from GEO database, GSE69636. This study, which contains 45 cord blood DNAm samples, investigates lead exposure induced in human embryonic stem cells and umbilical cord blood. The last data set (the 6th) also comes from the GEO database (GSE85042). The data is from a study examining whether sex-induced differences in DNAm of cord blood samples are related to sex-biases in psychiatric disease (Maschietto, et al., 2017). This data set has in total 71 samples. Across all the six studies, DNAm measurements using the Illumina 450k array were available from 525 whole cord blood samples.

In all of the six studies, DNAm IDAT files were processed into RGchannelsets used to estimate cell proportions using *estimateCellCounts* function in *minfi* package. The estimation was performed twice: once with the cord blood reference (without eosinophils) and once with the adult reference (with eosinophils). The proportions of granulocytes in cord blood were estimated using the existing *estimateCellCounts* function with cord blood reference (the correct reference database) and adult blood reference (the incorrect reference database), respectively. The results are summarized in SFigure 1. The bias was calculated as the difference between the two estimates of granulocytes cell proportions after a logit transformation for each sample in each of the six studies. It was elicited as prior information for systemic errors. We set *c=*0.028 as the weighted mean of calculated differences from granulocytes in the above six studies and *d=* 0.168 is the max variance of the calculated differences from the six studies. Note that *c* and *d* are cell-specific, that is, these two default values included in the R package are for eosinophils. If the cell proportion for a different cell type needs to be inferred, then different *c* and *d* need to be specified as seen in the example below.

# 3. Evaluation of the method

We use B lymphocytes (B cells) to demonstrate the effectiveness of the proposed method and assess the accuracy of estimated cells proportions in the aforementioned six studies. For B cells, references based on both cord blood and blood in adults are available. The proportions of B cells in cord blood for each sample are first estimated using the reference database inferred based on cord blood samples. With the correct reference database for our cord blood samples, the estimated cell proportions are treated as the “true” cell proportion of B cells, denoted as *Ptruthij* for subject *j* in study *i*, *i=1, …, 6*, and j=1, …, ni. The proportions of B cells are also estimated using reference profiles inferred from blood samples in adults (the incorrect reference database) without correcting measurement errors, and we denote the estimates as *PAij*.

Next, pretending we did not know *Ptruth*, we estimate the proportions of B cells using the proposed Bayesian method. The key parameters in the proposed method that need to be determined are *c* and *d*. Using the method discussed earlier, parameters *c* and *d* are calculated as -0.44 and 0.73 for the uniform prior, and -0.076 and 0.136 for the normality prior respectively. We denote the cell proportions estimated using the proposed method by *PMij*.

To assess the quality of cell proportions estimated using our method, we use two statistics, estimation bias and mean squared errors (MSE). Estimation bias is defined as the difference between estimated cell proportions and “true” cell proportions (*Ptruthij*), i.e., *PAji - Ptruthji*, and *PMji - Ptruthji*. SFigure 2 shows the bias of cell proportions of B cells in the six studies. In comparison to cell proportions estimated without correcting for measurement errors, the proposed method shows a significant reduction in bias in all the studies except for the Hopkins study (labeled as M2 and Adult in SFigure 2).

The MSE for cell proportion without measurement error correction is defined as

,

and similarly, for cell proportion with measurement error corrected,

.

The MSE for each study is shown in STable 1 (the column labeled as “Normal”). Using the Bayesian measurement error model with reference profiles inferred from blood samples in adults, the mean MSE from the six studies is 1.22 (the second column of STable 1. The mean MSEs are in the last row), which is reduced by nearly 30% compared to the average MSE without measurement error correction (MSE=1.73, the last column of STable 1). Overall, the proposed Bayesian measurement error model correction improves the accuracy of cell proportion estimation compared to the method without measurement error correction. As an illustration, via the proposed method and using the GSE85042 dataset, we estimated cell type proportions of seven cells, Bcell, CD4T, CD8T, Mono, Gran, NK, nRBC, and Eos (STable 2).

4. Impact of prior selection

Because the prior on and are informative, using different priors is likely to influence posterior inferences. To examine the impact of selecting different priors, we consider different prior distributions of , in addition to the normality prior distribution proposed in our study

4.1. Informative uniform prior

To specify an informative uniform prior, we set uniform (*c*, *d*) with *c* being the minimum value of the differences and *d* being the maximum difference. The derived posterior distribution and Gibbs sampler process is as the following:

1) Sampling the posterior distribution of , conditional on data and other parameters (denoted as rest), ,)

2) Sampling the conditional posterior distribution of , with ;

3) Sampling the conditional distribution of , , with and , with .

4.2. None informative normal prior

As an alternative approach, we also used none informative normal prior. That is, instead of specifying *c* and *d* based on existing studies, we choose *c=0* and *d=100*. The Gibbs sampling process is the same as in the case of using other priors,

1) Sampling the posterior distribution of , conditional on data and other parameters, ,

2) Sampling the conditional posterior distribution of , with ;

3) Sampling the conditional distribution of , , with and , with .

4.3. Results

For each of the two priors noted in 4.1 and 4.2, we calculated estimation bias and MSE. The bias from using the informative uniform prior is illustrated in SFigure 2 (labeled as M1), which is always bigger than the bias from the normal prior (labeled as M2) and the bias if assuming no measurement errors. The bias based on the non-informative normal prior is significantly larger than the bias from all the other methods and thus is not included in SFigure 2. Similar patterns are observed in the MSEs. That is, the normal informative prior performs the best while the non-informative normal (flat) prior performs the worst (STable 1).

These analyses demonstrate the importance of selecting informative priors and support our careful design of using as many as six studies to infer the hyper-parameters in the prior distributions.

**SFigure 1.** The percentages of cells estimated to be granulocytes using the cord blood reference panel (gray) and adult blood reference panel (white) in each of the six studies. IOW: Isle of Wight (IOW). EARLI: Early Autism Risk Longitudinal Investigation., Hopkins: Hopkins cell reference experiment convenience. NCS: National Children’s Study Vanguard.



**SFigure 2.** Bias of B lymphocytes cell proportion estimation using the Bayesian measurement error model with uniform prior (“**M1**”) and normal prior (“**M2**”), and using the traditional adult reference panel only (“**Adult**”) in each of the six studies. Top left: Isle of Wight (IOW), n=121; top right: Early Autism Risk Longitudinal Investigation (EARLI), n=175; middle left: Hopkins cell reference experiment convenience sample, n=15; middle right: National Children’s Study Vanguard (NCS), n=90; bottom left: GSE85042 study, n=45; and bottom right: GSE69636, n=71.



**STable 1:** Mean square error (MSE) in each of the six studies; columns 1 to 3: different prior distributions of *ν* in the measurement error model, column 4: traditional adult reference panel.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Prior of *ν* | | |  |
| **Study name** | Uniform | Normal | Flat | Adult |
| **GSE85042** | 15.674 | 2.415 | 1137.688 | 4.500 |
| **GSE69636** | 8.164 | 1.718 | 11.514 | 2.633 |
| **IOW** | 7.823 | 1.837 | 97.484 | 2.152 |
| **EARLI** | 2.456 | 0.285 | 0.472 | 0.411 |
| **Hopkins** | 2.869 | 0.865 | 137.175 | 0.211 |
| **NCS** | 7.225 | 0.217 | 3.799 | 0.493 |
| Mean MSE | 7.369 | 1.223 | 231.355 | 1.733 |

**STable 2:** Estimated cell proportions in the GSE85042 dataset (n=45) via the proposed method.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Bcell | CD4T | CD8T | Mono | Gran | NK | nRBC | Eos |
| GSM2256724\_9285451112\_R02C02 | 0.085 | 0.108 | 0.170 | 0.107 | 0.402 | 0.010 | 0.168 | 0.059 |
| GSM2256725\_9285451144\_R06C02 | 0.080 | 0.110 | 0.127 | 0.102 | 0.558 | 0.011 | 0.052 | 0.009 |
| GSM2256726\_9285451144\_R02C02 | 0.142 | 0.084 | 0.228 | 0.103 | 0.303 | 0.015 | 0.164 | 0.058 |
| GSM2256727\_9285451193\_R03C02 | 0.112 | 0.198 | 0.133 | 0.054 | 0.424 | 0.000 | 0.120 | 0.096 |
| GSM2256728\_9285451120\_R01C01 | 0.063 | 0.140 | 0.130 | 0.045 | 0.427 | 0.000 | 0.213 | 0.087 |
| GSM2256729\_9285451142\_R02C02 | 0.118 | 0.138 | 0.216 | 0.082 | 0.367 | 0.013 | 0.107 | 0.019 |
| GSM2256730\_9285451144\_R04C02 | 0.073 | 0.235 | 0.116 | 0.056 | 0.505 | 0.000 | 0.061 | 0.011 |
| GSM2256731\_9285451194\_R04C02 | 0.101 | 0.167 | 0.113 | 0.081 | 0.498 | 0.004 | 0.086 | 0.028 |
| GSM2256732\_9285451120\_R03C01 | 0.104 | 0.100 | 0.167 | 0.065 | 0.491 | 0.005 | 0.117 | 0.036 |
| GSM2256733\_9285451112\_R06C02 | 0.134 | 0.146 | 0.173 | 0.083 | 0.426 | 0.000 | 0.083 | 0.058 |
| GSM2256734\_9285451193\_R02C02 | 0.108 | 0.107 | 0.127 | 0.098 | 0.494 | 0.000 | 0.104 | 0.063 |
| GSM2256735\_9285451144\_R01C01 | 0.095 | 0.079 | 0.134 | 0.082 | 0.387 | 0.000 | 0.267 | 0.105 |
| GSM2256736\_9285451194\_R05C01 | 0.163 | 0.128 | 0.188 | 0.109 | 0.322 | 0.000 | 0.141 | 0.049 |
| GSM2256737\_9285451144\_R02C01 | 0.096 | 0.139 | 0.128 | 0.093 | 0.529 | 0.000 | 0.063 | 0.000 |
| GSM2256738\_9285451194\_R06C02 | 0.137 | 0.192 | 0.164 | 0.112 | 0.348 | 0.007 | 0.105 | 0.071 |
| GSM2256739\_9285451194\_R06C01 | 0.104 | 0.144 | 0.054 | 0.068 | 0.529 | 0.058 | 0.098 | 0.056 |
| GSM2256740\_9285451126\_R01C01 | 0.114 | 0.239 | 0.207 | 0.091 | 0.318 | 0.000 | 0.081 | 0.068 |
| GSM2256741\_9285451126\_R02C02 | 0.136 | 0.143 | 0.148 | 0.092 | 0.469 | 0.000 | 0.061 | 0.031 |
| GSM2256742\_9285451126\_R04C02 | 0.088 | 0.055 | 0.082 | 0.063 | 0.555 | 0.019 | 0.176 | 0.055 |
| GSM2256743\_9285451120\_R03C02 | 0.081 | 0.076 | 0.162 | 0.079 | 0.448 | 0.007 | 0.184 | 0.064 |
| GSM2256744\_9285451142\_R04C01 | 0.115 | 0.232 | 0.125 | 0.076 | 0.399 | 0.000 | 0.103 | 0.046 |
| GSM2256745\_9285451142\_R03C01 | 0.179 | 0.154 | 0.163 | 0.052 | 0.252 | 0.035 | 0.225 | 0.111 |
| GSM2256746\_9285451144\_R05C01 | 0.067 | 0.045 | 0.082 | 0.121 | 0.595 | 0.021 | 0.104 | 0.000 |
| GSM2256747\_9285451193\_R01C01 | 0.093 | 0.151 | 0.134 | 0.076 | 0.526 | 0.000 | 0.053 | 0.015 |
| GSM2256748\_9285451120\_R02C01 | 0.152 | 0.182 | 0.202 | 0.094 | 0.260 | 0.033 | 0.124 | 0.041 |
| GSM2256749\_9285451120\_R05C02 | 0.098 | 0.296 | 0.136 | 0.047 | 0.416 | 0.000 | 0.055 | 0.020 |
| GSM2256750\_9285451114\_R02C02 | 0.071 | 0.055 | 0.175 | 0.127 | 0.388 | 0.015 | 0.210 | 0.062 |
| GSM2256751\_9285451142\_R05C02 | 0.117 | 0.107 | 0.151 | 0.092 | 0.493 | 0.014 | 0.081 | 0.021 |
| GSM2256752\_9285451194\_R03C01 | 0.135 | 0.261 | 0.200 | 0.079 | 0.265 | 0.000 | 0.100 | 0.086 |
| GSM2256753\_9285451112\_R06C01 | 0.140 | 0.107 | 0.156 | 0.091 | 0.443 | 0.001 | 0.126 | 0.055 |
| GSM2256754\_9285451114\_R03C02 | 0.091 | 0.178 | 0.260 | 0.070 | 0.269 | 0.000 | 0.169 | 0.067 |
| GSM2256755\_9285451112\_R05C01 | 0.073 | 0.196 | 0.162 | 0.099 | 0.450 | 0.000 | 0.069 | 0.003 |
| GSM2256756\_9285451126\_R03C01 | 0.163 | 0.119 | 0.170 | 0.095 | 0.263 | 0.000 | 0.243 | 0.138 |
| GSM2256757\_9285451142\_R01C02 | 0.081 | 0.094 | 0.155 | 0.090 | 0.509 | 0.000 | 0.108 | 0.049 |
| GSM2256758\_9285451112\_R02C01 | 0.099 | 0.065 | 0.156 | 0.084 | 0.390 | 0.000 | 0.253 | 0.104 |
| GSM2256759\_9285451112\_R03C02 | 0.083 | 0.055 | 0.149 | 0.090 | 0.465 | 0.000 | 0.195 | 0.063 |
| GSM2256760\_9285451194\_R01C02 | 0.094 | 0.095 | 0.122 | 0.128 | 0.480 | 0.034 | 0.080 | 0.042 |
| GSM2256761\_9285451114\_R04C02 | 0.078 | 0.103 | 0.084 | 0.087 | 0.571 | 0.000 | 0.104 | 0.034 |
| GSM2256762\_9285451144\_R01C02 | 0.115 | 0.117 | 0.149 | 0.105 | 0.486 | 0.000 | 0.054 | 0.004 |
| GSM2256763\_9285451144\_R03C02 | 0.087 | 0.070 | 0.130 | 0.073 | 0.367 | 0.018 | 0.300 | 0.106 |
| GSM2256764\_9285451194\_R04C01 | 0.061 | 0.097 | 0.104 | 0.056 | 0.678 | 0.000 | 0.037 | 0.014 |
| GSM2256765\_9285451114\_R06C01 | 0.063 | 0.076 | 0.088 | 0.077 | 0.640 | 0.028 | 0.063 | 0.063 |
| GSM2256766\_9285451193\_R06C02 | 0.097 | 0.297 | 0.135 | 0.053 | 0.405 | 0.000 | 0.068 | 0.000 |
| GSM2256767\_9285451120\_R06C01 | 0.060 | 0.079 | 0.107 | 0.100 | 0.561 | 0.038 | 0.095 | 0.000 |
| GSM2256768\_9285451112\_R04C02 | 0.130 | 0.137 | 0.141 | 0.080 | 0.436 | 0.023 | 0.096 | 0.038 |
| GSM2256769\_9285451144\_R06C01 | 0.140 | 0.414 | 0.267 | 0.020 | 0.047 | 0.000 | 0.160 | 0.105 |
| GSM2256770\_9285451144\_R04C01 | 0.082 | 0.113 | 0.202 | 0.076 | 0.336 | 0.052 | 0.185 | 0.075 |
| GSM2256771\_9285451193\_R06C01 | 0.103 | 0.077 | 0.127 | 0.077 | 0.305 | 0.024 | 0.325 | 0.120 |
| GSM2256772\_9285451114\_R03C01 | 0.151 | 0.116 | 0.170 | 0.092 | 0.455 | 0.000 | 0.057 | 0.016 |
| GSM2256773\_9285451126\_R05C02 | 0.112 | 0.124 | 0.146 | 0.080 | 0.463 | 0.000 | 0.121 | 0.036 |
| GSM2256774\_9285451114\_R05C02 | 0.121 | 0.113 | 0.202 | 0.161 | 0.179 | 0.042 | 0.236 | 0.139 |
| GSM2256775\_9285451194\_R03C02 | 0.176 | 0.096 | 0.187 | 0.068 | 0.391 | 0.027 | 0.100 | 0.030 |
| GSM2256776\_9285451120\_R02C02 | 0.129 | 0.182 | 0.135 | 0.080 | 0.420 | 0.000 | 0.093 | 0.052 |
| GSM2256777\_9285451126\_R04C01 | 0.074 | 0.042 | 0.224 | 0.088 | 0.191 | 0.010 | 0.413 | 0.161 |
| GSM2256778\_9285451126\_R03C02 | 0.178 | 0.136 | 0.205 | 0.071 | 0.386 | 0.002 | 0.073 | 0.000 |
| GSM2256779\_9285451194\_R01C01 | 0.163 | 0.137 | 0.214 | 0.059 | 0.277 | 0.000 | 0.204 | 0.080 |
| GSM2256780\_9285451126\_R06C02 | 0.104 | 0.096 | 0.180 | 0.086 | 0.510 | 0.000 | 0.073 | 0.004 |
| GSM2256781\_9285451193\_R02C01 | 0.090 | 0.131 | 0.132 | 0.095 | 0.543 | 0.000 | 0.039 | 0.000 |
| GSM2256782\_9285451193\_R05C02 | 0.100 | 0.130 | 0.123 | 0.097 | 0.535 | 0.000 | 0.052 | 0.000 |
| GSM2256783\_9285451142\_R03C02 | 0.130 | 0.100 | 0.161 | 0.107 | 0.348 | 0.036 | 0.160 | 0.085 |
| GSM2256784\_9285451126\_R05C01 | 0.120 | 0.126 | 0.102 | 0.070 | 0.537 | 0.015 | 0.082 | 0.011 |
| GSM2256785\_9285451144\_R03C01 | 0.095 | 0.121 | 0.173 | 0.076 | 0.448 | 0.048 | 0.093 | 0.015 |
| GSM2256786\_9285451193\_R04C02 | 0.139 | 0.189 | 0.166 | 0.126 | 0.122 | 0.000 | 0.288 | 0.147 |
| GSM2256787\_9285451120\_R04C02 | 0.030 | 0.034 | 0.027 | 0.057 | 0.775 | 0.003 | 0.088 | 0.059 |
| GSM2256788\_9285451126\_R06C01 | 0.132 | 0.213 | 0.159 | 0.085 | 0.350 | 0.000 | 0.122 | 0.093 |
| GSM2256789\_9285451114\_R01C01 | 0.115 | 0.147 | 0.202 | 0.085 | 0.358 | 0.012 | 0.063 | 0.086 |
| GSM2256790\_9285451126\_R01C02 | 0.037 | 0.146 | 0.086 | 0.029 | 0.345 | 0.000 | 0.383 | 0.126 |
| GSM2256791\_9285451193\_R03C01 | 0.136 | 0.205 | 0.173 | 0.060 | 0.393 | 0.000 | 0.065 | 0.012 |
| GSM2256792\_9285451114\_R05C01 | 0.124 | 0.183 | 0.149 | 0.078 | 0.434 | 0.001 | 0.074 | 0.024 |
| GSM2256793\_9285451120\_R05C01 | 0.091 | 0.175 | 0.167 | 0.089 | 0.435 | 0.000 | 0.093 | 0.037 |
| GSM2256794\_9285451112\_R01C01 | 0.047 | 0.089 | 0.078 | 0.041 | 0.182 | 0.000 | 0.595 | 0.236 |

References

1. Arshad, S.H.*, et al.* Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol* 2018;47(4):1043-1044i.
2. Arshad, S.H.*, et al.* Multigenerational cohorts in patients with asthma and allergy. *J Allergy Clin Immunol* 2017;139(2):415-421.
3. Aryee, M.J.*, et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30(10):1363-1369.
4. Bakulski, K.M.*, et al.* DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* 2016;11(5):354-362.
5. Bakulski, K.M.*, et al.* Prenatal mercury concentration is associated with changes in DNA methylation at TCEANC2 in newborns. *Int J Epidemiol* 2015;44(4):1249-1262.
6. Berek, C. Eosinophils: important players in humoral immunity. *Clin Exp Immunol* 2016;183(1):57-64.
7. Cardenas, A.*, et al.* Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. *Epigenetics* 2015;10(6):508-515.
8. Ethier, C., Lacy, P. and Davoine, F. Identification of human eosinophils in whole blood by flow cytometry. *Methods Mol Biol* 2014;1178:81-92.
9. Fortin, J.P., Triche, T.J., Jr. and Hansen, K.D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* 2017;33(4):558-560.
10. Hide, D.W.*, et al.* Cord serum IgE: an insensitive method for prediction of atopy. *Clin Exp Allergy* 1991;21(6):739-743.
11. Houseman, E.A.*, et al.* DNA Methylation in Whole Blood: Uses and Challenges. *Current environmental health reports* 2015;2(2):145-154.
12. Junge, K.M.*, et al.* The LINA cohort: Cord blood eosinophil/basophil progenitors predict respiratory outcomes in early infancy. *Clin Immunol* 2014;152(1-2):68-76.
13. Kim, J.H.*, et al.* Relationship between prenatal and postnatal exposures to folate and risks of allergic and respiratory diseases in early childhood. *Pediatr Pulmonol* 2015;50(2):155-163.
14. Lavigne, S.*, et al.* Identification and analysis of eosinophils by flow cytometry using the depolarized side scatter-saponin method. *Cytometry* 1997;29(3):197-203.
15. Maschietto, M.*, et al.* Sex differences in DNA methylation of the cord blood are related to sex-bias psychiatric diseases. *Sci Rep* 2017;7:44547.
16. Newschaffer, C.J.*, et al.* Infant siblings and the investigation of autism risk factors. *J Neurodev Disord* 2012;4(1):7.
17. Thurau, A.M.*, et al.* Identification of eosinophils by flow cytometry. *Cytometry* 1996;23(2):150-158.