Placenta 88 (2019) 8–11

Contents lists available at ScienceDirect

Placenta

journal homepage: www.elsevier.com/locate/placenta

Technical note

Quality control measures for placental sample purity in DNA methylation array analyses

Qianhui Wan[a,b], Shalem Yiner-Lee Leemaqz[a,b], Stephen Martin Pederson[c], Dylan McCullough[a,b], Dale Christopher McAninch[a,b], Tanja Jankovic-Karasoulos[a,b], Melanie Denise Smith[a,b], Konstantinos Justinian Bogias[a,b], Ning Liu[b,c], James Breen[a,c,d], Claire Trelford Roberts[a,b], Tina Bianco-Miotto[a,e,*]

[a] Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia
[b] Adelaide Medical School, University of Adelaide, Adelaide, SA, Australia
[c] Bioinformatics Hub, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia
[d] South Australian Health & Medical Research Institute (SAHMRI), Adelaide, SA, Australia
[e] School of Agriculture, Food and Wine, University of Adelaide, Adelaide, SA, Australia

ARTICLE INFO

Keywords:
Human placenta
DNA methylation
Methylation Bead Chip
Quality control

ABSTRACT

The purity of tissue samples can affect the accuracy and utility of DNA methylation array analyses. This is particularly important for the placenta which is globally hypomethylated compared to other tissues. Placental villous tissue from early pregnancy terminations can be difficult to separate from non-villous tissue, resulting in potentially inaccurate results. We used several methods to identify mixed placenta samples using DNA methylation array datasets from our laboratory and those contained in the NCBI GEO database, highlighting the importance of determining sample purity during quality control processes.

1. Introduction

The number of studies investigating genome-scale DNA methylation is growing rapidly. The most popular platforms are DNA methylation arrays [1]. The placenta is a transient organ that primarily acts to support fetal differentiation and growth and orchestrates maternal adaptations to pregnancy [2,3]. Placental DNA is uniquely and globally hypomethylated, reflecting this organ's early developmental origin and distinct functions [4,5]. Analysing the placental methylome could provide information relevant to understanding pregnancy health and disease [6]. Existing pipelines for processing DNA methylation array data do not include a quality control step for detecting tissue impurities [7,8]. We demonstrate that mixed placenta samples, that is placenta samples that contain a significant amount of surrounding contaminating tissue, can be identified from DNA methylation array data, allowing for sample removal or incorporation of other analytic changes as appropriate.

2. Methods

2.1. Array data

Thirteen human tissue datasets containing raw IDAT files from Illumina Infinium HumanMethylation450 (450 K) or HumanMethylationEPIC (EPIC) array platforms (GPL13534 or GPL21145) were acquired from NCBI GEO database and our laboratory. Ten datasets used the 450 K platform (GSE66210, GSE74738, GSE69502, GSE75196, GSE75248, GSE120250, GSE100197, GSE98224, GSE71678, GSE66459) [9–18] and 3 the EPIC platform (GSE115508, GSE113600 [19,20] plus 10 samples from our study, GSE131945). 410 samples (394 from 450 K platform and 16 from EPIC platform) were selected: 380 placentas [22 from first trimester, 16 from second trimester (pregnancy terminations) and 342 from term (uncomplicated pregnancies)], 12 maternal whole blood samples, 11 umbilical cord blood, 3 decidua, 2 amnion and 2 chorion (Supplementary Table 1). For the term samples, only samples from uncomplicated pregnancies were used in the downstream analyses.

* Corresponding author. Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia.
E-mail address: tina.bianco@adelaide.edu.au (T. Bianco-Miotto).

https://doi.org/10.1016/j.placenta.2019.09.006
Received 12 June 2019; Received in revised form 3 September 2019; Accepted 15 September 2019
Available online 17 September 2019
0143-4004/

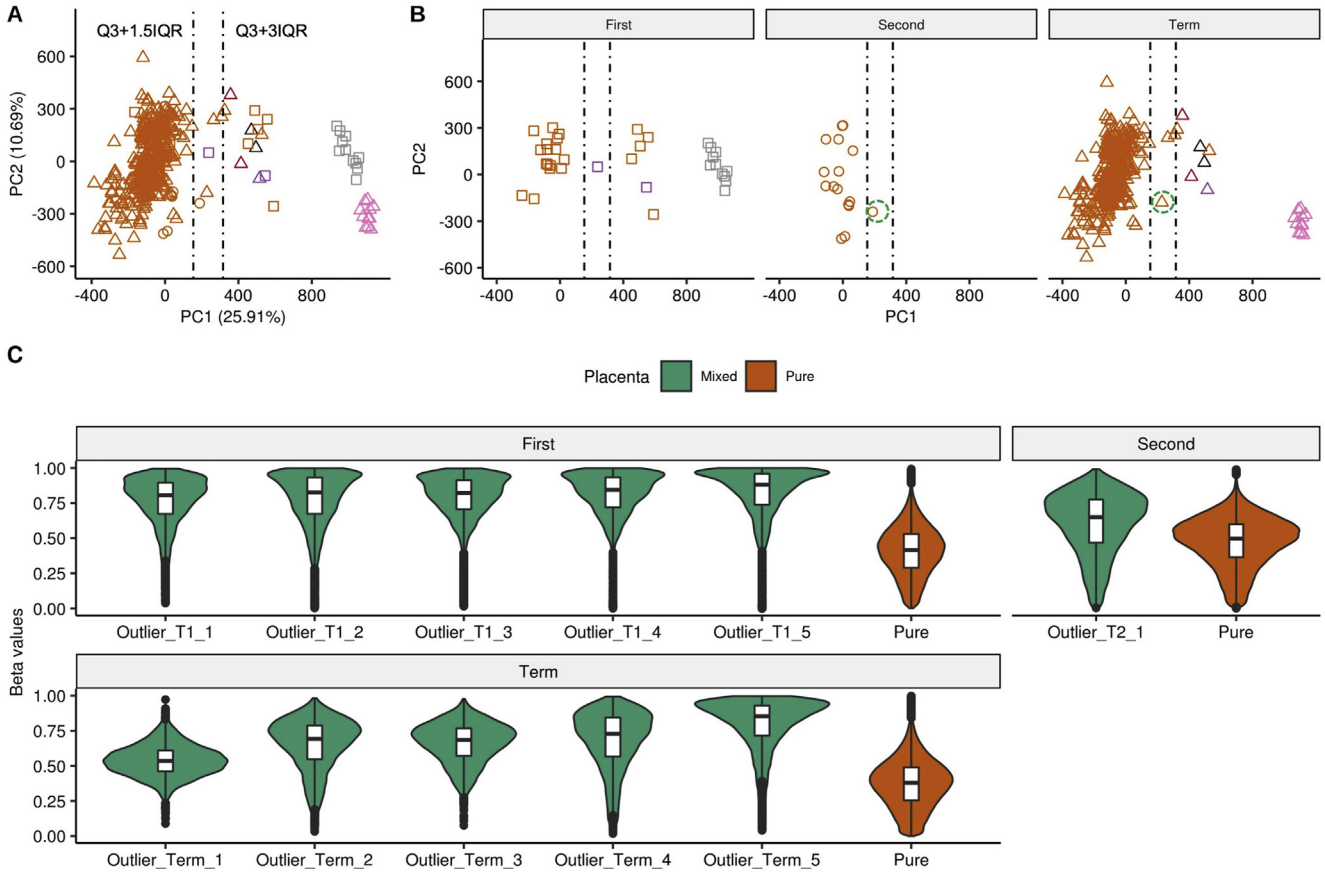

Fig. 1. Analysis of 408 samples from different tissue types. (A) PCA plot using M values for all 408 samples. Trimester: □first, ○second, △term. Tissue types: placenta (orange), amnion (black), chorion (red), maternal whole blood (grey), decidua (purple), umbilical cord blood (pink). Dashed vertical line marked Q3 + 1.5IQR (153.88) and Q3 + 3IQR (316.01) of PC1. Outliers are at the right side of dashed vertical line (Q3 + 1.5IQR). (B) PCA plot of samples from first trimester, second trimester and term were shown respectively. Mixed placenta samples tend to be similar to non-placenta tissues. Samples in dashed green ellipses were Outlier_T2_1 and Outlier_Term_1. (C) DNA methylation difference between outlier and pure placenta from first trimester, second trimester and term. T1: first trimester; T2: second trimester.

2.2. Array quality control

Two samples were removed due to low array intensities defined as < 10.5, one first trimester placenta sample (GSM1617002) and one maternal whole blood sample (GSM1616993). Probes filtered out included: failed probes with detection P > 0.01, probes with fewer than 3 beads in > 5% of all samples [8]; cross-reactive probes [21]; probes on sex chromosomes and probes with biased DNA methylation signal due to SNPs at CpG sites, single base extension (SBE) sites and probe body [22,23]. In all, 408 samples and 301,496 probes common to both 450 K and EPIC arrays were used for downstream analyses. See Supplementary Methods for additional information.

2.3. Principal component analysis (PCA) and sample clustering

The dye bias correction for filtered data was performed using the method regression on logarithm of internal control probes (RELIC) from *ENmix* package and background subtraction according to tissue types were performed with the method from *ENmix* package which models methylation signal intensities with a flexible exponential-normal mixture distribution, and use out-of-band Infinium I intensities ("OOB") to estimate normal distribution parameters to model background noise. Datasets were normalised using the beta-mixture quantile normalization method [24] to correct for type I and type II probe bias. Control probes were used to check batch effects between studies. All 408 samples were used to generate the PCA plot which gave an overview of similarities and differences between samples from different tissue types. Interquartile range (IQR) of values for samples at PC1 was calculated to estimate outliers. Using PC1, placenta samples with less than Q3 + 1.5IQR were considered pure placenta samples and those samples with greater than Q3 + 3IQR were mixed placenta samples. The area in between Q3 + 1.5IQR and Q3 + 3IQR was designated a 'grey' zone. Multivariate unsupervised clustering with mixtures of Gaussian distributions [25] was performed to estimate the probability of each sample being a pure placenta sample. The top 2% of probes that were different between pure and mixed samples (PC1) were selected to investigate DNA methylation differences between them. M values were used for the statistical tests based on linear models.

2.4. DNA methylation analyses at partially methylated domains (PMDs) and placenta-specific imprinting control regions (ICRs)

The genome was tailed by 10 kb non-overlapping bins and mean Beta values (methylation percentage) of each bin were calculated [5]. Mean Beta values of bins that overlapped with PMDs for pure and mixed placenta samples were plotted to show the difference between these samples. Beta values at ICRs [26] for pure and mixed samples were also plotted. Mann–Whitney *U* test was used to test the significance of change of DNA methylation at PMDs and ICRs between potentially mixed and pure placenta sample groups. The codes for all analyses can be found here: https://github.com/QianhuiWan/MethylationArray_Placenta.

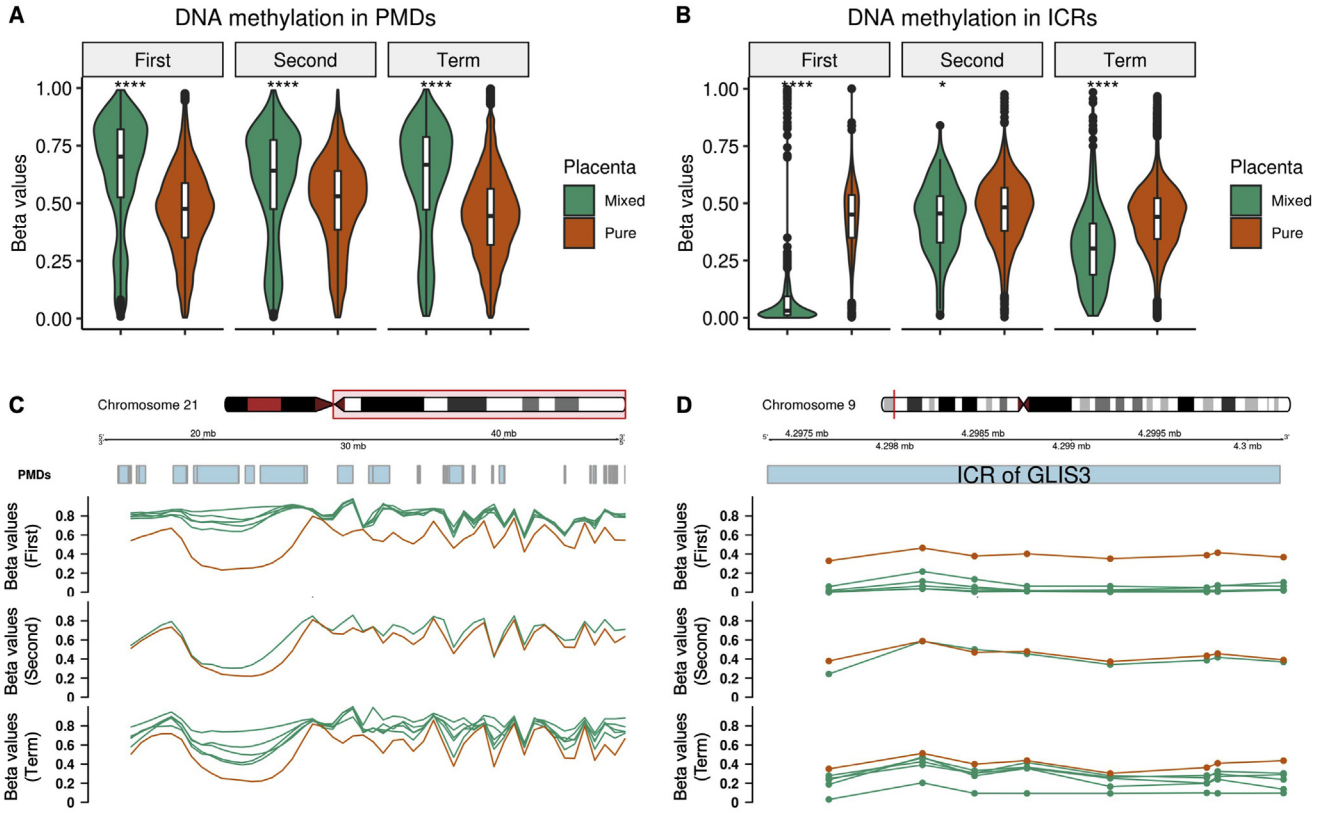

Fig. 2. DNA methylation in partially methylated domains (PMDs) (A and C) and imprinting control regions (ICRs) (B and D). (A) Average DNA methylation in PMDs was higher in mixed (green) than pure placenta samples (orange) ($p < 2.2e-16$), Chr21 was used as an example as shown in (C). (B) DNA methylation at placenta-specific ICRs was overall lower in mixed (green) than in pure placenta samples (orange) ($p < 0.05$), ICR on Chr9 is an example shown in (D). Green lines in (C) and (D) represent individual mixed samples from different trimesters, while orange lines in (C) and (D) represent mean beta values of pure samples from different trimesters. Mann–Whitney *U* test was used to test the significance between groups in plot (A) and (B). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 2.2e-16$.

3. Results

The PCA plot of 408 samples identified 11 placenta samples as outliers (Fig. 1A). Clustering of a subset of placental samples to non-placental samples suggests that these samples may be mixed with DNA from non-placenta tissue. These samples will be called “mixed placenta samples” in this report (Supplementary Fig. 1). Five were from first trimester, one from second trimester and the remaining five were from term samples (Fig. 1B). Four of the term outliers were previously identified [14]. Nine placenta sample outliers clustered with decidua, amnion and chorion and are likely to contain more non-placenta tissue than the other two samples (Outlier_T2_1 and Outlier_Term_1, dashed green circle; Fig. 1B). The estimated probability of each placenta sample being a pure sample is listed in Supplementary Table 2. The top 2% of most variable probes according to PC1, showed that all had higher DNA methylation in the putative mixed placenta samples compared to the remaining placenta samples (Fig. 1C).

Mixed placenta samples from first trimester, second trimester and term, showed greater DNA methylation than pure placenta samples at PMDs ($P < 2.2e-16$) (Fig. 2A and C). Mixed placenta samples showed altered DNA methylation compared to pure placenta samples at placenta-specific ICRs ($P < 0.05$) (Fig. 2B and D and Supplementary Fig. 2).

4. Discussion

Using publicly available DNA methylation array datasets, as well as 10 samples of our own, we identified outlier placenta villous tissue samples which clustered with other tissue types and were likely to contain other tissues, most likely decidua. These outliers had different DNA methylation profiles at both PMDs and ICRs, indicating they may contain non-villous tissue, such as hypermethylated maternal decidua [27]. Placenta samples from early gestation are at particular risk of being mixed with other tissue types, possibly because the termination procedures inevitably macerate the tissue. Since these outliers had quite distinct methylation profiles compared to other placenta samples, they could potentially influence results, necessitating their removal or down weighting before further analysis.

Before processing high dimension data from placenta villous tissues, we recommend checking sample purity in three steps. First, download the DNA methylation array data listed in Supplementary Table 1 from NCBI GEO database. Second, apply PCA and unsupervised clustering using pure placenta samples from GEO database as positive controls and samples from other tissue types from GEO database as negative controls. Third, if any samples cluster with other tissue types, and the estimated probability of the sample being pure placenta is low, the DNA methylation of placenta PMDs and ICRs needs to be checked to verify whether these samples should be removed or down weighted.

Conflicts of interest

The authors confirm that there are no conflicts of interest.

Acknowledgments

We would like to thank all the women who have kindly donated their placentas for our research and the clinical staff who collected them.

CTR was supported by a Lloyd Cox Professorial Research Fellowship from the University of Adelaide. The research was funded by a NICHD

NIH Project Grant (NIH oppRFA-HD-16-036 1 R01 HD089685-01): Maternal molecular profiles reflect placental function and development across gestation awarded to CTR, TBM and JB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.placenta.2019.09.006.

References

[1] A.E. Teschendorff, C.L. Relton, Statistical and integrative system-level analysis of DNA methylation data, Nat. Rev. Genet. 19 (2017) 129–147, https://doi.org/10.1038/nrg.2017.86.
[2] N.M. Gude, C.T. Roberts, B. Kalionis, R.G. King, Growth and function of the normal human placenta, Thromb. Res. 114 (2004) 397–407, https://doi.org/10.1016/j.thromres.2004.06.038.
[3] S. Zhang, T. Regnault, P. Barker, K. Botting, I. McMillen, C. McMillan, C. Roberts, J. Morrison, S. Zhang, T.R.H. Regnault, P.L. Barker, K.J. Botting, I.C. McMillen, C.M. McMillan, C.T. Roberts, J.L. Morrison, Placental adaptations in growth restriction, Nutrients 7 (2015) 360–389, https://doi.org/10.3390/nu7010360.
[4] T. Bianco-Miotto, B.T. Mayne, S. Buckberry, J. Breen, C.M. Rodriguez Lopez, C.T. Roberts, Recent progress towards understanding the role of DNA methylation in human placental development, Reproduction 152 (2016) R23–R30, https://doi.org/10.1530/REP-16-0014.
[5] D.I. Schroeder, J.D. Blair, P. Lott, H.O.K. Yu, D. Hong, F. Crary, P. Ashwood, C. Walker, I. Korf, W.P. Robinson, J.M. LaSalle, The human placenta methylome, Proc. Natl. Acad. Sci. 110 (2013) 6037–6042, https://doi.org/10.1073/PNAS.1215145110.
[6] W.P. Robinson, E.M. Price, The human placental methylome, Cold Spring Harb. Perspect. Med. 5 (2015) a023044, https://doi.org/10.1101/cshperspect.a023044.
[7] J.-P. Fortin, T.J. Triche, K.D. Hansen, Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi, Bioinformatics 33 (2016) btw691, https://doi.org/10.1093/bioinformatics/btw691.
[8] Y. Tian, T.J. Morris, A.P. Webster, Z. Yang, S. Beck, A. Feber, A.E. Teschendorff, ChAMP: updated methylation analysis pipeline for Illumina Bead Chips, Bioinformatics 33 (2017) 3982–3984, https://doi.org/10.1093/bioinformatics/btx513.
[9] L. Hatt, M.M. Aagaard, J. Graakjaer, C. Bach, S. Sommer, I.E. Agerholm, S. Kølvraa, A. Bojesen, Microarray-based analysis of methylation status of CpGs in placental DNA and maternal blood DNA – potential new epigenetic biomarkers for cell free fetal DNA-based diagnosis, PLoS One 10 (2015) e0128918, , https://doi.org/10.1371/journal.pone.0128918.
[10] C.W. Hanna, M.S. Peñaherrera, H. Saadeh, S. Andrews, D.E. McFadden, G. Kelsey, W.P. Robinson, Pervasive polymorphic imprinted methylation in the human placenta, Genome Res. 26 (2016) 756–767, https://doi.org/10.1101/gr.196139.115.
[11] E.M. Price, M.S. Peñaherrera, E. Portales-Casamar, P. Pavlidis, M.I. Van Allen, D.E. McFadden, W.P. Robinson, Profiling placental and fetal DNA methylation in human neural tube defects, Epigenet. Chromatin 9 (2016) 6, https://doi.org/10.1186/s13072-016-0054-8.
[12] K.R. Yeung, C.L. Chiu, R. Pidsley, A. Makris, A. Hennessy, J.M. Lind, DNA methylation profiles in preeclampsia and healthy control placentas, Am. J. Physiol. Cell Physiol. 310 (2016) H1295–H1303, https://doi.org/10.1152/ajpheart.00958.2015.
[13] A.G. Paquette, E.A. Houseman, B.B. Green, C. Lesseur, D.A. Armstrong, B. Lester, C.J. Marsit, Regions of variable DNA methylation in human placenta associated with newborn neurobehavior, Epigenetics 11 (2016) 603–613, https://doi.org/10.1080/15592294.2016.1195534.
[14] S. Choufani, A.L. Turinsky, N. Melamed, E. Greenblatt, M. Brudno, A. Bérard, W.D. Fraser, R. Weksberg, J. Trasler, P. Monnier, For the 3D cohort study group, impact of assisted reproduction, infertility, sex, and paternal factors on the placental DNA methylome, Hum. Mol. Genet. (2018), https://doi.org/10.1093/hmg/ddy321.
[15] S.L. Wilson, K. Leavey, B.J. Cox, W.P. Robinson, Mining DNA methylation alterations towards a classification of placental pathologies, Hum. Mol. Genet. 27 (2018) 135–146, https://doi.org/10.1093/hmg/ddx391.
[16] K. Leavey, S.L. Wilson, S.A. Bainbridge, W.P. Robinson, B.J. Cox, Epigenetic regulation of placental gene expression in transcriptional subtypes of preeclampsia, Clin. Epigenet. 10 (2018) 28, https://doi.org/10.1186/s13148-018-0463-6.
[17] B.B. Green, M.R. Karagas, T. Punshon, B.P. Jackson, D.J. Robbins, E.A. Houseman, C.J. Marsit, Epigenome-Wide assessment of DNA methylation in the placenta and arsenic exposure in the New Hampshire birth cohort study (USA), Environ. Health Perspect. 124 (2016) 1253–1260, https://doi.org/10.1289/ehp.1510437.
[18] F. Fernando, R. Keijser, P. Henneman, A.-M.F. van der Kevie-Kersemaekers, M.M. Mannens, J.A. van der Post, G.B. Afink, C. Ris-Stalpers, The idiopathic preterm delivery methylation profile in umbilical cord blood DNA, BMC Genomics 16 (2015) 736, https://doi.org/10.1186/s12864-015-1915-4.
[19] C. Konwar, E.M. Price, L.Q. Wang, S.L. Wilson, J. Terry, W.P. Robinson, DNA methylation profiling of acute chorioamnionitis-associated placentas and fetal membranes: insights into epigenetic variation in spontaneous preterm births, Epigenet. Chromatin 11 (2018) 63, https://doi.org/10.1186/s13072-018-0234-9.
[20] M. Yu, G. Du, Q. Xu, Z. Huang, X. Huang, Y. Qin, L. Han, Y. Fan, Y. Zhang, X. Han, Z. Jiang, Y. Xia, X. Wang, C. Lu, Integrated analysis of DNA methylome and transcriptome identified CREB5 as a novel risk gene contributing to recurrent pregnancy loss, EBioMedicine 35 (2018) 334–344, https://doi.org/10.1016/j.ebiom.2018.07.042.
[21] R. Pidsley, E. Zotenko, T.J. Peters, M.G. Lawrence, G.P. Risbridger, P. Molloy, S. Van Djik, B. Muhlhausler, C. Stirzaker, S.J. Clark, Critical evaluation of the Illumina Methylation EPIC BeadChip microarray for whole-genome DNA methylation profiling, Genome Biol. (2016), https://doi.org/10.1186/s13059-016-1066-1.
[22] D. Zhi, S. Aslibekyan, M.R. Irvin, S.A. Claas, I.B. Borecki, J.M. Ordovas, D.M. Absher, D.K. Arnett, SNPs located at CpG sites modulate genome-epigenome interaction, Epigenetics 8 (2013) 802–806, https://doi.org/10.4161/epi.25501.
[23] W. Zhou, P.W. Laird, H. Shen, Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes, Nucleic Acids Res. (2017), https://doi.org/10.1093/nar/gkw967.
[24] S. Horvath, DNA methylation age of human tissues and cell types, Genome Biol. 14 (2013) R115, https://doi.org/10.1186/gb-2013-14-10-r115.
[25] L. Scrucca, M. Fop, T.B. Murphy, A.E. Raftery, Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models, R J. 8 (2016) 289–317 http://www.ncbi.nlm.nih.gov/pubmed/27818791 , Accessed date: 23 May 2019.
[26] F. Court, C. Tayama, V. Romanelli, A. Martin-Trujillo, I. Iglesias-Platas, K. Okamura, N. Sugahara, C. Simón, H. Moore, J.V. Harness, H. Keirstead, J.V. Sanchez-Mut, E. Kaneki, P. Lapunzina, H. Soejima, N. Wake, M. Esteller, T. Ogata, K. Hata, K. Nakabayashi, D. Monk, Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment, Genome Res. 24 (2014) 554–569, https://doi.org/10.1101/gr.164913.113.
[27] T.J. Jensen, S.K. Kim, Z. Zhu, C. Chin, C. Gebhard, T. Lu, C. Deciu, D. van den Boom, M. Ehrich, Whole genome bisulfite sequencing of cell-free DNA and its cellular contributors uncovers placenta hypomethylated domains, Genome Biol. 16 (2015) 78, https://doi.org/10.1186/s13059-015-0645-x.