Placenta 84 (2019) 57–62

Contents lists available at ScienceDirect

Placenta

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

Considerations when processing and interpreting genomics data of the placenta

Chaini Konwar[a,b,1], Giulia Del Gobbo[a,b,1], Victor Yuan[a,b,1], Wendy P. Robinson[a,b,*]

[a] BC Children's Hospital Research Institute, 950 West 28th Ave, Vancouver, BC V5Z 4H4, Canada
[b] Department of Medical Genetics, University of British Columbia, 4500, Oak Street, Vancouver, BC V6H3N1, Canada

ARTICLE INFO

Keywords:
Placenta
Genomics
DNA methylation
Microarray
Gene expression

ABSTRACT

The application of genomic approaches to placental research has opened exciting new avenues to help us understand basic biological properties of the placenta, improve prenatal screening/diagnosis, and measure effects of *in utero* exposures on child health outcomes. In the last decade, such large-scale genomic data (including epigenomics and transcriptomics) have become more easily accessible to researchers from many disciplines due to the increasing ease of obtaining such data and the rapidly evolving computational tools available for analysis. While the potential of large-scale studies has been widely promoted, less attention has been given to some of the challenges associated with processing and interpreting such data. We hereby share some of our experiences in assessing data quality, reproducibility, and interpretation in the context of genome-wide studies of the placenta, with the aim to improve future studies. There is rarely a single "best" approach, as that can depend on the study question and sample cohort. However, being consistent, thoroughly assessing potential confounders in the data, and communicating key variables in the methods section of the manuscript are critically important to help researchers to collaborate and build on each other's work.

1. Introduction

A variety of high-throughput technologies can be applied to the placenta to better characterize its normal development, alterations associated with pathology, and adaptations associated with adverse *in utero* exposures. While publications using these approaches are on the rise, interpretation of results can be confounded by many factors including technical variables related to the arrays themselves and the variety of analysis approaches applied. This article summarizes a number of aspects—from sample procurement to data analysis—that should be considered when performing these experiments and evaluating the resulting data. While we focus our discussion on microarrays, most of our comments apply also to other high-throughput approaches, such as sequencing-based analysis of gene expression.

1.1. Genome-wide microarray technologies

Microarrays typically rely on the hybridization of test DNA or RNA to short oligonucleotides arrayed on a chip. Microarrays in which the DNA oligonucleotides (probes) are complementary to specific single nucleotide polymorphisms (SNPs) can be used for comprehensive genotyping of samples, as well as for the detection of chromosome abnormalities such as copy number variants (CNVs). Other types of microarrays are used to quantify levels of DNA methylation (DNAme) at hundreds of thousands of sites across the genome, such as the Illumina HumanMethylation450 (450 K) or EPIC (850 K) arrays. The profiles generated can help identify homogenous patient sub-groups based on molecular signatures and refine existing clinical diagnoses [1]; evidence of *in utero* exposure, such as maternal lifestyle (diet, smoking) or health (diabetes, obesity) factors [2]; or placenta-derived biomarkers in maternal blood that are predictive of perinatal complications [3]. Gene expression can also be measured using microarrays and can similarly detect changes in the placenta associated with pathology or exposures [4,5]. As changes to DNAme are often associated with changes in gene expression, these methods can be complementary [5].

1.2. Obtaining representative placental samples

A challenge somewhat unique to placental research is that even healthy placentas can show extensive variation in structure and cell composition, including localized areas of pathology. As each cell type has a unique epigenetic and gene expression profile, it is important to

* Corresponding author. BC Children's Hospital Research Institute, 950 West 28th Ave, Vancouver, BC V5Z 4H4, Canada.
E-mail addresses: chaini@alumni.ubc.ca (C. Konwar), gdelgobbo@bcchr.ca (G. Del Gobbo), vyuan@bcchr.ca (V. Yuan), wrobinson@bcchr.ca (W.P. Robinson).
[1] These authors contributed equally.

https://doi.org/10.1016/j.placenta.2019.01.006
Received 31 October 2018; Received in revised form 28 December 2018; Accepted 4 January 2019

obtain samples in a consistent way that minimizes this variation and allows for comparison across cohorts. Several previous reviews provide helpful guidance in following a standardized approach to sampling the placenta (chorionic villi) [6,7]. Taking samples from areas of local pathology or the maternal side of the placenta should be avoided, unless critical to the study question, for example, when studying gene expression at the maternal-placental interface in terminal villi. Areas where maternal decidua is present not only contaminate the placental (fetal) sample with a different (maternal) genome, but may also result in significant contributions of additional cell types that co-localize with maternal tissue, such as extravillous trophoblast and maternal dendritic cells [8]. Screening such samples for maternal contamination by genotyping polymorphic DNA markers, such as microsatellite loci, and comparing the ratio of maternal:fetal alleles allows for an estimation of the contribution of maternal genome to the placental sample [9]. For sequence-based studies such as RNAseq, maternal transcripts can directly be detected in the data if there are sequence variants between mother and fetus. This information can be kept into consideration when analyzing 'omics data and interpreting results from such samples, or it can be used to exclude highly contaminated samples. Additionally, it is important to wash the sample well to remove contaminating blood or amniotic fluid.

As cell composition variation can be a large driver of variation in DNAme and gene expression, it is popular in studies of blood and other tissues to obtain isolated cell populations or apply approaches to adjust measurements for altered cell ratios [10,11]. Currently, isolating uniform populations of single placental cell types has practical challenges. As a result, most placental studies are done in whole chorionic villus samples and thus, would benefit from cell deconvolution algorithms, statistical approaches to estimate and account for cell composition variation [9,10]. Reference-based cell deconvolution approaches require DNAme signatures of isolated placental cell types and therefore cannot be applied as of yet. Until then, reference-free approaches that allow the adjustment of cell type ratios without direct estimation of the cell proportions can be used. However, researchers should consider the purpose of the study before removing effects of cell composition variation with reference-free cell deconvolution. For example, consistent changes in placental cell ratios that are associated with adverse perinatal outcomes are of interest to identify (not adjust out of the analysis), as these are relevant for generating biomarkers of disease that are useful for diagnosis or reflected in maternal serum (which can reflect more expressing cells, as well as up-regulation of expression) [12].

Genetic or epigenetic mutations can arise with each cell division and may be confined to individual villous trees due to the clonal nature of their development [13,14]. This genetic mosaicism, along with variation in cell composition, local pathology and random effects, contributes to the high degree of gene expression and epigenetic variation observed between samples of chorionic villi taken from distinct sites from the placenta [15–17]. These sites are often pooled together to provide a more representative sample of the placenta for gene expression or DNAme studies. However, samples taken from distinct cotyledons can also be used as independent biological replicates to estimate intra-placental variability [17]. The average intra-placental variation can then be compared to between-group comparisons and technical replicates to better interpret the biological significance of results. For example, based on RNA sequencing and representing each placenta with three sets of pooled samples, each from a different placental quadrant, it was observed that 33% of gene expression variation was explained by intra-placental variation, while 59% was attributed to between-individual variation and 8% to between-population groups [17]. If the between-group differences at a particular gene of interest do not exceed normal biological variability within a placenta, then the result is less likely to be clinically meaningful (Fig. 1).

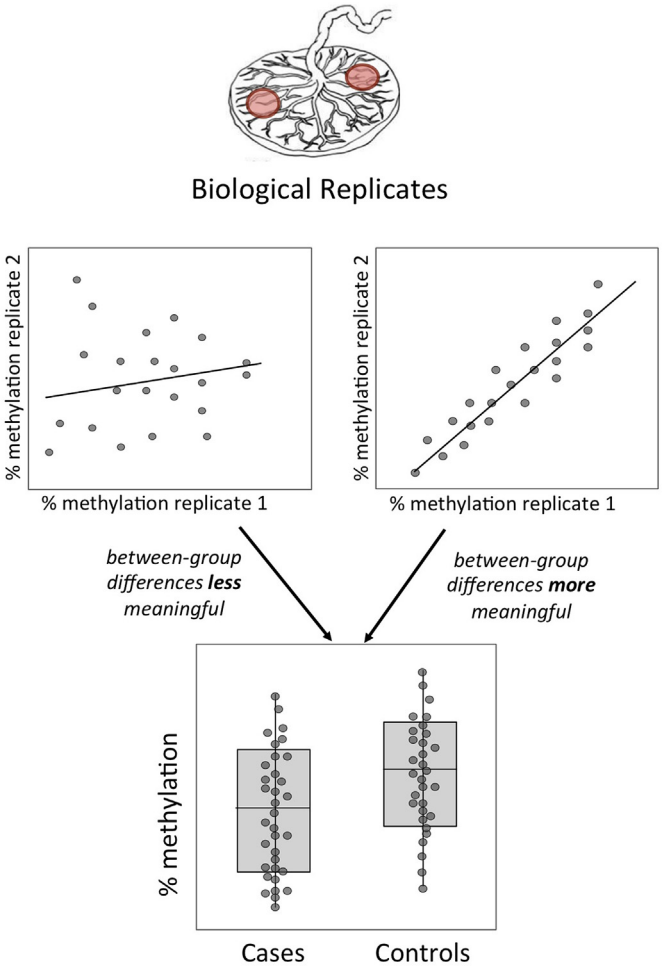

Fig. 1. Variation between biological replicates provides context for intersample comparisions. Correlation between two biological replicates may be high (right) or low (left) depending on the site analyzed. Differences observed when comparing cases and controls (lower figure) are more likely to be meaningful when biological replication is high.

1.3. The importance of high-quality RNA or DNA

While most studies consider clinical factors that can affect study questions, for example maternal health parameters or perinatal complications, often less attention is given to sample processing parameters that can affect DNA/RNA quality. These can include time and mode of storage of the placenta prior to sampling, processing time and sample storage parameters (temperature, buffers, etc.) [16,18,19]. Even relatively small differences in RNA integrity number (RIN) significantly affected the quantification of 8% of studied transcripts in a study of RNA sequencing of the human placenta [19]. While RNA quality is typically more sensitive to processing times than are miRNA or DNA quality, any factor that affects the integrity of the sample can affect array hybridization or sequencing efficiency [20]. This is particularly a concern if cases and controls are being obtained in a disparate manner (i.e. varying by centre or gestational age), which can be associated with systematic differences in sample processing. Recording the various parameters that can potentially impact DNA/RNA quality or integrity, and correcting for them in subsequent analyses, is recommended. For example, in processing placental samples on the Illumina Omni2.5 SNP array for the detection of CNVs, we find that more CNVs are detected (especially losses) when DNA quality is low (Fig. 2). This is clearly an artifact of a high rate of probe failure and can be easily detected if the correct quality control measures are employed. Such associations

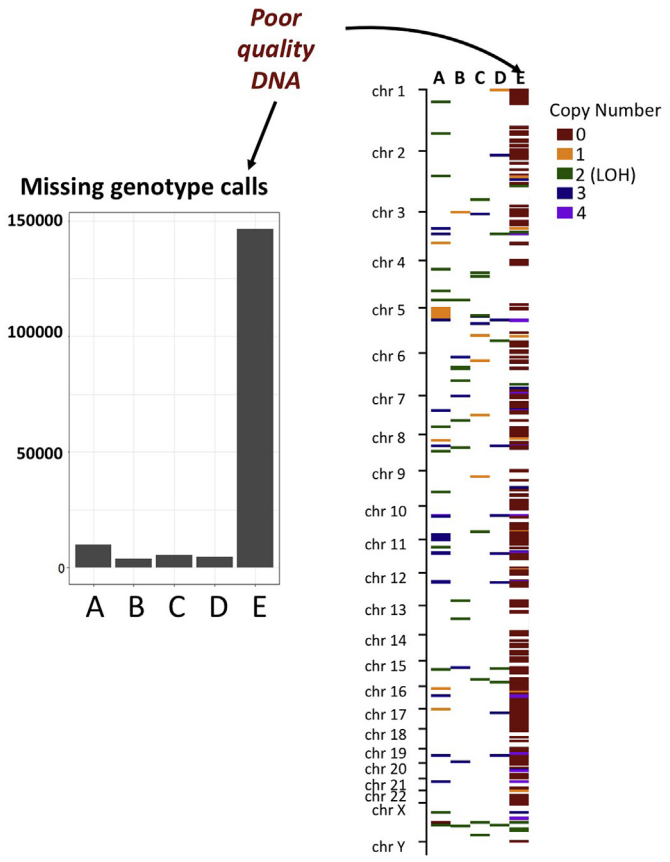

Fig. 2. Sample quality can affect array hybridization. In this example SNP array hybridization shows a high failure rate in sample E; this sample also shows an abnormally high number of CNV calls. LOH: loss-of-heterozygosity.

between DNA/RNA quality and data quality are inherent problems associated with any genomics analysis method, thus effective strategies are needed to identify poor quality samples prior to data collection or to filter them from results through efficient data quality checks (see below).

1.4. Creating a data acquisition plan

In performing genome-wide analysis (or reviewing such studies) it is important to ensure that true biological signal is not confounded with signal variation due to batch effects [21,22]. Batch effects can be varied in nature including date of run, different operators or labs where samples are processed, sample processing and storage methods, reagent batches, chip (or flow cell in the case of sequencing) and row-on-chip effects. Leek et al., 2010 [21] suggested that sample batch is typically one of the top sources of variation in high-throughput data and provided various strategies for detecting such unwanted effects through initial exploratory data analysis followed by the use of surrogate variables to adjust downstream analyses. However, application of batch-correction methods can introduce false biological signal (and false negative) if applied to an unbalanced study design [22,23]. Thus, we recommend that researchers design a data acquisition plan to ensure that batches are as balanced as possible across the primary variables of interest, as well as common confounders such as sex, processing time and gestational age. The study sample size needs to be large enough to allow for this. For example, if there are 6 rows on each array chip, the sample size in each group should be sufficient such that representative samples from each group can be dispersed semi-randomly across rows and chips (Fig. 3a). If samples vary by critical variables such as processing time, gestational age and sex, then these should also be randomized as best as

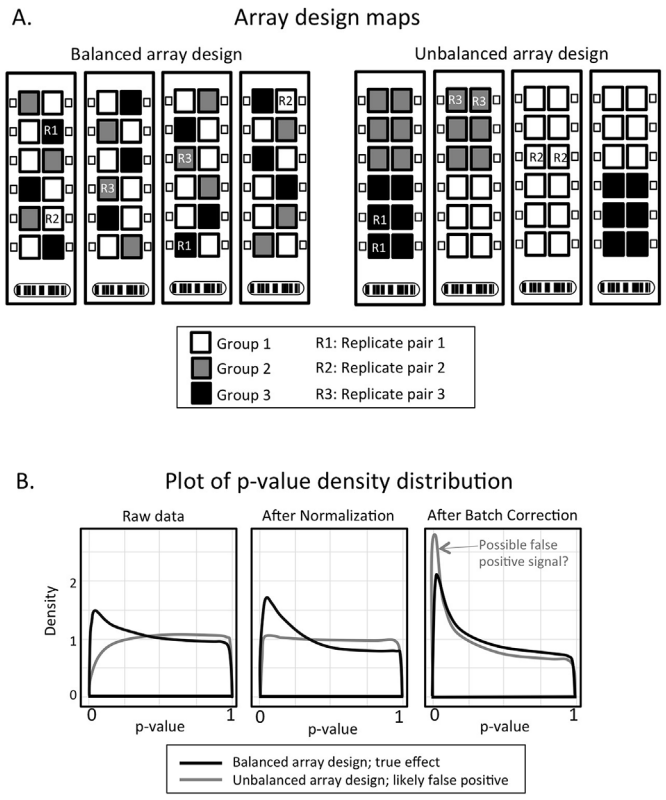

Fig. 3. A. Example of a simple array map with three study groups (e.g. three gestational ages). Samples should be balanced across rows, chips and batches for key variables (groups, sex, etc) and include technical and/or biological replicates. **B. P-Value density plots** of two theoretical sets data before and after data processing. Black: Normalization and batch correction will most likely improve significance of a true signal; Grey: The introduction of strong signal only after batch correction, with no prior evidence of an increased density of small p-values, may be the result of an unbalanced array design (after reference [23]).

possible across row, chip and batches. Incorporating technical replicates in the study (identical samples repeated more than once) also allows one to evaluate whether data processing methods are effective at reducing the level of background technical noise (i.e. increasing the correlation between the replicates). Thus, large sample sizes are not only important for increasing the power of the study, but for incorporating correction procedures that reduce technical noise.

As an example, in a recent study [23] we highlighted the issue that false positives can arise using a standard approach for batch correction, the function *ComBat* [24] in the R package *sva*, which is a valuable tool for correcting for batch effects such as chip and row in microarray data. Applying such correction in our study comparing placental DNAme by *MTHFR* genotype (N = 10 in each comparison group) resulted in nearly 20,000 CpGs significant at a false discovery rate of < 0.05, despite no evidence of a significant signal (based on the p-value distribution) in the raw data [23]. Because the unadjusted *p*-value distribution can be visualized at each stage of the data processing (see e.g. Fig. 3b), it was evident that the application of the batch correction step was generating this unexpectedly large number of “hits”. When reviewing the study design, it was found that samples were not equally represented across bisulfite conversion batches, chips, and rows, a consequence of the fact that this was a secondary analysis of previously collected data and not part of the original study plan. By taking advantage of additional samples not initially included in the batch correction step, the false positive signal was eliminated. This was fortunate, but in situations where batch variables are entirely confounded with study group, it is impossible to disentangle true biological signal from technical signal

due to batch effects.

1.5. Evaluating data quality through visualization approaches

Several types of probe and sample quality checks are commonly performed before any analysis is initiated, including identification and removal of poorly hybridizing probes for microarrays or low read counts for sequencing, removal of samples with a large number of failed probes, identification of mislabeled and/or outlier samples, and removal of samples with an excessive number of poor-quality measurements. Furthermore, visualization tools such as dendrograms, scatterplots, MA plots, heat maps, density plots, and principal components analysis (PCA) are often used in exploratory data analysis and as approaches to monitor data processing [25,26]. It is important to visualize high-dimensional data in multiple ways to check for data integrity and identify any unusual patterns. There is no single best method and there are often platform specific-approaches to data visualization. However, we list here some example data visualization approaches that we routinely employ in analysis of microarray data from the placenta.

- **Confirmation of the expected overall pattern in the data can be done by plotting a density distribution of the measured values** (Fig. 4A). For example, raw beta values from the Illumina 450 K and 850 K Methylation arrays should show a characteristic, tissue-specific pattern. As placenta has a unique epigenetic and gene expression pattern compared to other tissues [27], such profiles should be compared to expectations for placenta and not for other tissue types. In contrast to a bimodal beta value distribution observed in most somatic tissues, where CpGs typically are either < 10% methylated or > 90% methylated, the presence of partially methylated domains, imprinted genes, and methylation polymorphism in the placenta results in a unique trimodal distribution of DNAme measures (Fig. 4A). Deviations from the expected pattern could be indicative of, for example, chip problems, scanner problems, incomplete DNA bisulfite conversion, or contamination of placental samples with other cell/tissue types. Comparison of density plots by study cohort or batches can also be informative of systematic differences between cohorts, which one may try to reduce through batch correction and normalization procedures.
- **Identification of outlier samples can be done by plotting the inter-sample correlation** (Fig. 4B). Results from samples obtained from the same tissue should have relatively high correlations with each other, whereas samples with low correlations could be the result of contamination with other cells or poor quality [28], although abnormal pathology may also lead to poor correlation. Plotting inter-sample correlation can additionally indicate systematic differences between study cohorts or batch (Fig. 4B). One can also examine the correlation between individual samples, for example technical replicates, which can be used to measure background noise in the data.
- **Confirming expected relationships between samples can be done using clustering approaches** [29]**.** If different types of tissue are included in a study, e.g. chorionic villi, amnion, chorion, and/or cord blood, then samples should cluster cleanly by tissue [27]. Within a tissue, male and female sexes should be easily distinguishable when clustering on sex-chromosome specific genes/probes. This holds true for both methylation and gene expression data despite the fact that the inactive X chromosome is hypomethylated in female placenta compared to somatic tissues [30,31]. In most cases, samples that do not cluster with the correct tissue or sex are indicative of sample mix-ups or mislabeling and should be of concern. Hierarchical clustering can also be an approach to identify outliers that do not cluster with any group, suggesting technical issues (i.e. poor DNA quality or poor hybridization). In addition, technical and biological replicates should cluster together (Fig. 4C).
- **Primary sources of variation in the data can be identified with principal components analysis (PCA) and then visualized** [32]. PCA is a commonly used dimension reduction tool that compresses high dimensional data into fewer dimensions or principal components (PCs), while retaining the trends and patterns in the original data. A correlation analysis with the top PCs can identify the major sources of variation in the data, and should include both biological

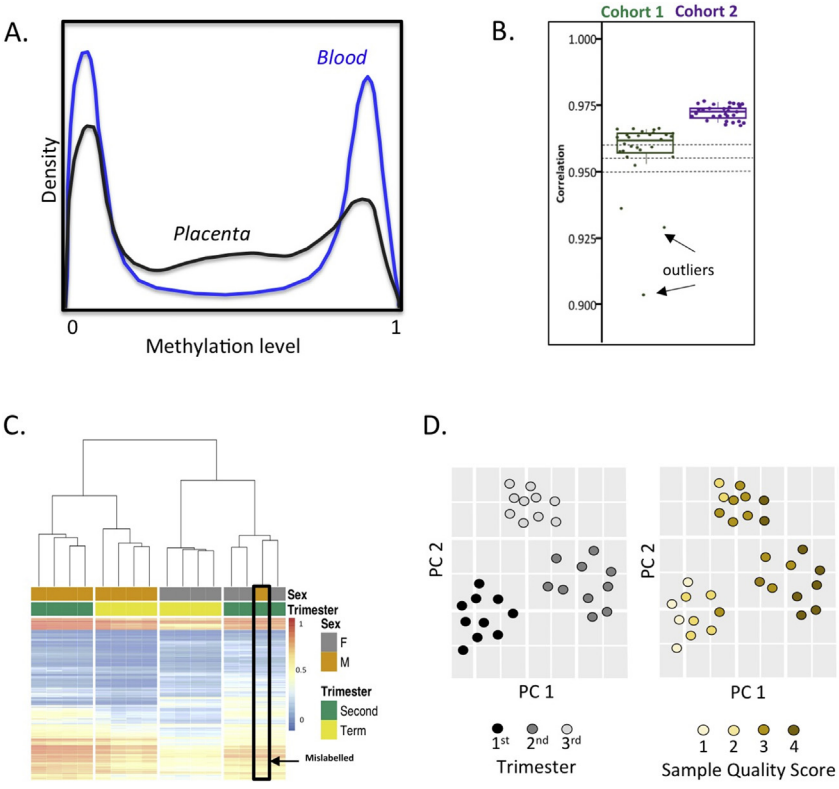

Fig. 4. Data visualization approaches to evaluate data integrity A. Density distribution from placenta and blood as expected utilizing all probes on the Illumina 450 K array; atypical patterns may suggest technical problems. **B. Intersample correlations** can be used to identify outlier samples that should be removed from the data. Differences in mean inter-sample correlation between cohorts suggests systemic differences that could potentially affect joint analysis of data. **C. Hierarchical clustering** of data shows clear separation by sex except for a single sample interpreted to be mislabeled; secondary clustering by gestational age is also observed. **D. PCA visualization** of two variables associated with PC1 shows that while there is a clear separation of samples by gestational age (left), visualization of the data by sample quality score (e.g. RNA RIN value; processing time etc) shows that sample quality is confounded with gestational age such that highest quality samples are from the 2nd trimester.

variables (sex, gestational age, ethnicity, pathology or phenotype of interest) as well as technical variables (platform, processing time, batch, chip, row, DNA/RNA quality measurements). Any variable correlated with the top PCs should then be carefully monitored and accounted for in downstream analyses to minimize potential confounding. In addition to performing the PCA itself to identify which variables are associated with main PCs, it is useful to plot the PC data while highlighting different variables (group, sex, gestational age, technical variables) to better understand their relationships. If multiple variables are associated with the same PCs, visualizing the data helps in understanding if (and how) these are confounded (Fig. 4D).

1.6. Replication of findings

Recent publications have brought attention to the lack of reproducibility in research [33,34], particularly in high-throughput genomic, epigenomic, and transcriptomic studies. Many factors can contribute to non-reproducible results, including failure to control for bias, low statistical power, poor quality control, p-value hacking, and publication bias [35]. Differences in analysis approaches may also contribute. For example, concordance of CNV calls using a variety of microarray platforms and algorithms is often < 50% when analyzing the same data and < 70% when analyzing data from technical replicates [36]. However, there is currently little consensus on what the best data processing and analytic approaches are for any given situation and new methods continue to be proposed. It has thus become increasingly important to replicate findings in a distinct study cohort before any conclusions are solidified. However, replication alone does not prove validity and many steps should be taken to show a result is meaningful [37]. Moreover as most studies show only associations, further investigation of the potential pathways that underlie these associations can add validity to the result. For example, functional studies using cellular and animal models can help to identify the causal mechanisms.

It is challenging to determine whether lack of reproducibility is due to research biases, technical issues, different analytical approaches or population differences. For example, in a recent study on preeclampsia, we collected Illumina 450 K array data on two distinct cohorts of samples including placentas associated with early-onset maternal preeclampsia compared to term and preterm normotensive samples [1]. Although samples from the two cohorts were obtained in different Canadian cities (Vancouver and Toronto), the population demographics were similar, the same sampling protocol was followed, and DNA samples were processed and run on the array at the same centre (Vancouver). While there was a significant overlap in differentially methylated hits between the two cohorts, including at many functionally interesting genes (*FN1, TEAD3, JUNB, PAPPA2, INHA* etc.), there was also a surprising number of sites that failed to replicate, particularly those statistically significant hits with smaller effect sizes. While the explanation is likely due to a combination of factors, the sample sizes may have been insufficient to account for all technical variables. While samples from different cohorts can be grouped together to increase power, this can also potentially increase false signal in the data due to an inability to correct for batch effects. On the other hand, using multiple cohorts for replication reduces the power to detect effects, but increases the likelihood that findings are true positives.

2. Conclusion

The placenta is a readily accessible organ, and with the widespread interest in understanding how pregnancy can influence fetal development and long-term health, increasing numbers of studies are now targeting this tissue. It is thus increasingly important to apply lessons from past research to improve the quality of such studies. A list of studies referenced in this review can be found organized by topic with description in Supplementary Table 1. While the solution to many issues around reproducibility may be large sample sizes with many replicates, the reality is such studies are costly and most investigators do not have the funds to perform truly optimally designed studies. Increased collaboration and sharing of public datasets provide sufficient sample sizes, as long as there is good attention to sampling protocols, technical variables, and confounders in such studies, in order to allow for a better integration of such shared data. Furthermore, each public dataset should be accompanied by clear information on sampling processes, relevant technical variables such as DNA/RNA quality measurements, batch/chip design maps, pre-processed data, and an honest appraisal of any concerns investigators had in data collection and processing. Working together in this way will help to harness the power of high-throughput technologies in order to further our understanding of placental biology and its role in developmental health and disease.

Authors' contributions

All authors contributed to ideas, editing of manuscript, producing figures and have read and approved the final manuscript. WPR drafted the manuscript based on ideas and discussions shared with the lab members, including GDG, CK, and VY.

Conflicts of interest

The authors declare that they have no competing interests.

Conflict of interests to declare

None.

Funding

This work was supported by grants from the Canadian Institutes of Health Research (CIHR) [WPR; PJT-152887] and National Institutes of Health (WPR; RFN 5R01HD089713-04). WPR receives salary support through an award from the BC Children's Hospital Research Institute. GDG receives support from a CIHR Doctoral Fellowship (GSD-157973).

Acknowledgements

We are thankful to current and alumnus Robinson lab members for their insightful discussions and advice. In particular, Dr. E. Magda Price, Dr. Maria Peñaherrera and Amy Inkster contributed to the ideas and edits of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.placenta.2019.01.006.

References

[1] 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 (2017) 135–146.
[2] C.S. Wilhelm-Benartzi, E.A. Houseman, M.A. Maccani, G.M. Poage, D.C. Koestler, S.M. Langevin, L.A. Gagne, C.E. Banister, J.F. Padbury, C.J. Marsit, In utero exposures, infant growth, and DNA methylation of repetitive elements and developmentally related genes in human placenta, Environ. Health Perspect. 120 (2012) 296–302.
[3] I. Manokhina, G.F. del Gobbo, C. Konwar, S.L. Wilson, W.P. Robinson, Placental biomarkers for assessing fetal health, Hum. Mol. Genet. 26 (2017) R237–R245.
[4] S.J. Benton, K. Leavey, D. Grynspan, B.J. Cox, S.A. Bainbridge, The clinical heterogeneity of preeclampsia is related to both placental gene expression and placental histopathology, Obstet. Gynecol. 219 (2018) 604e1:e25.
[5] 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.
[6] G. Burton, N. Sebire, L. Myatt, D. Tannetta, Y. Wang, Y. Sadovsky, A. Staff, C. Redman, Optimising sample collection for placental research, Placenta 35 (2014)

9–22.
[7] K. Hogg, E.M. Price, W.P. Robinson, Improved reporting of DNA methylation data derived from studies of the human placenta, Epigenetics 9 (2014) 333–337.
[8] J.C.H. Tsang, J.S.L. Vong, L. Ji, L.C.Y. Poon, P. Jiang, K.O. Lui, Y.B. Ni, K.F. To, Y.K.Y. Cheng, R.W.K. Chiu, Y.M.D. Lo, Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics, Proc. Natl. Acad. Sci. U. S. A. 14 (2017) E7786–E7795.
[9] E. Winsor, H. Akoury, D. Chitayat, L. Steele, T. Stockley, The role of molecular microsatellite identity testing to detect sampling errors in prenatal diagnosis, Prenat. Diagn. 30 (2010) 746–752.
[10] A.E. Teschendorff, C.L. Relton, Statistical and integrative system-level analysis of DNA methylation data, Nat. Rev. Genet. 19 (2018) 129.
[11] A.J. Titus, R.M. Gallimore, L.A. Salas, B.C. Christensen, Cell-type deconvolution from DNA methylation: a review of recent applications, Hum. Mol. Genet. 26 (2017) R216–R224.
[12] S.L. Wilson, J.D. Blair, K. Hogg, S. Langlois, P. von Dadelszen, W.P. Robinson, Placental DNA methylation at term reflects maternal serum levels of INHA and FN1, but not PAPPA, early in pregnancy, BMC Med. Genet. 16 (2015) 1.
[13] M.S. Penaherrera, R. Jiang, L. Avila, R.K. Yuen, C.J. Brown, W.P. Robinson, Patterns of placental development evaluated by X chromosome inactivation profiling provide a basis to evaluate the origin of epigenetic variation, Hum. Reprod. 27 (2012) 1745–1753.
[14] R. Yuen, W. Robinson, Review: a high capacity of the human placenta for genetic and epigenetic variation: implications for assessing pregnancy outcome, Placenta 32 (2011) S136–S141.
[15] R. Sood, J.L. Zehnder, M.L. Druzin, P.O. Brown, Gene expression patterns in human placenta, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5478–5483.
[16] L. Avila, R. Yuen, D. Diego-Alvarez, M. Peñaherrera, R. Jiang, W. Robinson, Evaluating DNA methylation and gene expression variability in the human term placenta, Placenta 31 (2010) 1070–1077.
[17] D.A. Hughes, M. Kircher, Z. He, S. Guo, G.L. Fairbrother, C.S. Moreno, P. Khaitovich, M. Stoneking, Evaluating intra-and inter-individual variation in the human placental transcriptome, Genome Biol. 16 (2015) 54.
[18] N. Vilahur, A.A. Baccarelli, M. Bustamante, S. Agramunt, H. Byun, M.F. Fernandez, J. Sunyer, X. Estivill, Storage conditions and stability of global DNA methylation in placental tissue, Epigenomics 5 (2013) 341–348.
[19] M. Reiman, M. Laan, K. Rull, S. Sõber, Effects of RNA integrity on transcript quantification by total RNA sequencing of clinically collected human placental samples, FASEB J. 31 (2017) 3298–3308.
[20] M. Forest, K.J. O'Donnell, G. Voisin, H. Gaudreau, J.L. MacIsaac, L.M. McEwen, P.P. Silveira, M. Steiner, M.S. Kobor, M.J. Meaney, Agreement in DNA methylation levels from the Illumina 450K array across batches, tissues, and time, Epigenetics 13 (2018) 19–32.
[21] J.T. Leek, R.B. Scharpf, H.C. Bravo, D. Simcha, B. Langmead, W.E. Johnson, D. Geman, K. Baggerly, R.A. Irizarry, Tackling the widespread and critical impact of batch effects in high-throughput data, Nat. Rev. Genet. 11 (2010) 733.
[22] O.D. Buhule, R.L. Minster, N.L. Hawley, M. Medvedovic, G. Sun, S. Viali, R. Deka, S.T. McGarvey, D.E. Weeks, Stratified randomization controls better for batch effects in 450K methylation analysis: a cautionary tale, Front. Genet. 5 (2014) 354.
[23] E.M. Price, W.P. Robinson, Adjusting for batch effects in DNA methylation microarray data, a lesson learned, Front. Genet. 9 (2018) 83.
[24] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883.
[25] S. Liu, D. Maljovec, B. Wang, P. Bremer, V. Pascucci, Visualizing high-dimensional data: advances in the past decade, IEEE Trans. Visual. Comput. Graph. (2017) 1249–1268.
[26] G. Grinstein, M. Trutschl, U. Cvek, High-dimensional Visualizations, (2001), p. 120.
[27] W.P. Robinson, E.M. Price, The human placental methylome, Cold Spring Harb. Perspect. Med. 5 (2015) a023044.
[28] R.D. Edgar, M.J. Jones, W.P. Robinson, M.S. Kobor, An empirically driven data reduction method on the human 450K methylation array to remove tissue specific non-variable CpGs, Clin. Epigenet. 9 (2017) 11.
[29] N. Altman, M. Krzywinski, Points of Significance: Clustering, (2017).
[30] A.M. Cotton, L. Avila, M.S. Penaherrera, J.G. Affleck, W.P. Robinson, C.J. Brown, Inactive X chromosome-specific reduction in placental DNA methylation, Hum. Mol. Genet. 18 (2009) 3544–3552.
[31] A.M. Cotton, L. Lam, J.G. Affleck, I.M. Wilson, M.S. Penaherrera, D.E. McFadden, M.S. Kobor, W.L. Lam, W.P. Robinson, C.J. Brown, Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation, Hum. Genet. 130 (2011) 187–201.
[32] J. Lever, M. Krzywinski, N. Altman, Points of significance: principal component analysis, Nat. Methods 14 (2017) 641–642.
[33] R. Nuzzo, Fooling ourselves, Nature 526 (2015) 182.
[34] Open Science Collaboration, Estimating the reproducibility of psychological science, Science 349 (2015) aac4716.
[35] M.R. Munafò, B.A. Nosek, D.V. Bishop, K.S. Button, C.D. Chambers, N.P. du Sert, U. Simonsohn, E. Wagenmakers, J.J. Ware, J.P. Ioannidis, A manifesto for reproducible science, Nat. Human Behav. 1 (2017) 0021.
[36] D. Pinto, K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, A.C. Lionel, B. Thiruvahindrapuram, J.R. MacDonald, R. Mills, Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants, Nat. Biotechnol. 29 (2011) 512.
[37] J.P. Ioannidis, M.J. Khoury, Improving validation practices in "omics" research, Science 334 (2011) 1230–1232.