Clinica Chimica Acta 496 (2019) 13–17

Contents lists available at ScienceDirect

Clinica Chimica Acta

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

A simple method to allow for guanine-cytosine amplification error in prenatal DNA screening for trisomy 18

Nicholas J. Wald*,[1], Jonathan P. Bestwick[1], King Wai Lau, Wayne J. Huttly, Weilin Ke, Ray Cheng, Robert W. Old

Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

ARTICLE INFO

Keywords:
Prenatal screening
Trisomy 18
Edwards syndrome
Cell-free DNA
GC content

ABSTRACT

Background: A source of error in prenatal screening for trisomies is PCR amplification error associated with guanine-cytosine (GC) content of DNA fragments in maternal plasma. We describe a simple method of allowing for this.

Methods: Data from a Reflex DNA screening programme (67 trisomy 18 and 83 unaffected pregnancies) were used to compare the ratio of chromosome 18 DNA fragment counts to chromosome 8 DNA fragment counts (because chromosome 8 has a similar GC content to chromosome 18) with the percentage of chromosome 18 DNA counts using counts from all autosomes in the denominator, with and without an all autosome correction for the GC content of the DNA fragments.

Results: A chromosome 18 to 8 ratio of DNA fragment counts was more discriminatory than the percentage of all autosome counts arising from chromosome 18 without, or with an all autosome correction for GC content bias. It achieves a high screening performance, eg. for a 0.25% false-positive rate, a 97% detection rate instead of 49% without a correction for GC content, and 91% with an all autosome correction for GC content.

Conclusion: Consideration can be given to using the ratio of chromosome 18 DNA fragment counts to chromosome 8 DNA fragment counts in cell-free DNA prenatal screening for trisomy 18, avoiding the need for more complex methods of making a correction for the GC content currently used.

1. Introduction

Analysis of maternal plasma DNA (also known as cell-free DNA) is an accurate method for prenatal screening for fetal trisomies 21, 18, and 13 [1]. However the screening performance for trisomy 18, is less than for trisomy 21 [2], and the reasons for this are unknown. This prompted us to examine possible sources of analytical error that might affect DNA screening for trisomy 18.

The DNA analysis most widely used in prenatal screening for trisomy 18 is massively parallel sequencing. This involves sequencing several million DNA fragments in maternal plasma and then calculating the proportion of sequences that map to chromosome 18. The denominator of the proportion is usually the number of DNA fragments that map to all autosomes. A correction for GC (guanine-cytosine) content of the DNA fragments is usually applied [3] to allow for GC associated error in the PCR copying number of DNA fragments. The usual method for allowing for GC error relies on a plot of the number of sequenced DNA fragments from all chromosomes against the GC content of the fragments [4]. Ideally, there should be no association between the GC content of a fragment and the fragment counts sequenced so that the plot is horizontal. In practice, however, the plot is bell-shaped, indicating underestimation with DNA fragments with high and low GC content and overestimation in between. Deviations from the overall average (ie. expectation) can be used to standardize (ie. correct) the error. The method has the advantage of generalizability (eg. applicable to DNA fragments from all chromosomes) but it has several disadvantages. The method is prone to variation from analytical run to run, and corrections, vary according to the pre-sequencing steps (eg. how the PCR is performed), and according to the sequencing methods used, all of which impair analytical precision. This all autosome GC correction method is complex, not transparent, and requires a large dataset, preferably linked to a particular sequencing method and laboratory.

Sehnert and colleagues [5] indicate that it may be better to use a

* Corresponding author.
E-mail address: n.j.wald@qmul.ac.uk (N.J. Wald).
[1] Joint first authors.

https://doi.org/10.1016/j.cca.2019.06.015
Received 25 March 2019; Received in revised form 5 June 2019
Available online 15 June 2019

single or a small number of chromosomes in the denominator, instead of all autosomes when calculating the proportion of DNA fragments aligning to chromosome 18. Empirical testing of different chromosome denominators indicated that chromosome 8 was the most discriminatory for trisomy 18. We explored this strategy as a way of improving DNA screening performance for trisomy 18, using a larger data set obtained from the Wolfson Institute (London) prenatal screening programme for trisomy 21, 18, and 13 from 2015 to 2018.

2. Methods

Maternal plasma DNA from 67 trisomy 18 (affected) pregnancies and 83 unaffected pregnancies was sequenced using a semiconductor sequencing platform and software [6]. Typically about 10 million DNA fragments were analysed in each plasma sample. Data from the BAM (Binary Alignment Map) files that plasma DNA analysis generated for each pregnancy were aligned to the human reference genome (hg19). DNA fragments that mapped to individual chromosomes were counted. The fetal fraction of individual samples was estimated by proprietary (Premaitha) software within the test platform.

DNA fragments from chromosome 18 were expressed as (i) a percentage using counts from all autosomes as the denominator, without a correction for GC content, (ii) a percentage using counts from all autosomes as the denominator, with an all autosome correction for GC content and (ii) a ratio using counts from chromosome 8 as the denominator. Whether a percentage or ratio is used with all autosomes as the denominator will make little difference because the proportion of DNA fragments from chromosome 18 is small compared to fragments from all autosomes. However, when only considering chromosomes 18 and 8 it does matter because the random error arising from combining the DNA fragment counts from chromosomes 18 and 8 in the denominator is greater than with fragments from chromosome 8 alone. Dot plots, that categorise observations in a manner that avoids overlapping dots, were used to visually compare the distributions in affected and unaffected pregnancies with DNA fragments from chromosome 18 expressed as (i) a percentage using counts from all autosomes as the denominator, without a correction for GC content, (ii) a percentage using counts from all autosomes as the denominator, with an all autosome correction for GC content and (ii) a ratio using counts from chromosome 8 as the denominator.

The percentages and ratio were converted into multiple of the median (MoM) values by dividing by the respective median percentages and ratio in unaffected pregnancies. A regression of $\log_{10}$ MoM values against fetal fraction in affected pregnancies was performed to estimate the fetal-fraction specific median MoM in affected pregnancies. To illustrate the fit of chromosome 18 MoM values (using all autosomes as the denominator and using chromosome 8 as the denominator) to Gaussian distributions, probability plots were generated separately for affected and unaffected pregnancies, with affected pregnancies standardised to a fetal fraction of 6% by adjusting the MoM values according to the slopes of the regression lines. Points on the probability plots lying on a straight line indicate a good fit to a Gaussian distribution. The standard deviations of MoM values in affected and unaffected pregnancies were taken as the slopes of regression lines of the points on each probability plot between the 10th and 90th centiles; a standard method of estimating the standard deviation that avoids the undue influence of outliers. [7] The risk of each of the 67 affected and 83 unaffected pregnancies being affected with trisomy 18 was estimated as the maternal age-specific odds of an affected livebirth [8], adjusted to the first trimester by the fetal loss rate from this time in pregnancy until term [9], multiplied by the likelihood ratio (the height of the fetal-fraction specific Gaussian distribution in affected pregnancies divided by the height of the Gaussian distribution in unaffected pregnancies). Screening performance was estimated as the detection rate (DR; the proportion of affected pregnancies with a positive result) for a specified false-positive rate (FPR; the proportion of unaffected pregnancies with a positive result), (FPR for a specified DR and DR and FPR for a specified risk cut-off). Modelling based on multivariate Gaussian analyses provides a more robust estimate of screening performance, provided the underlying distribution of the markers are approximately Gaussian, a method that is routinely used in prenatal screening and has been empirically validated. [10–12] It avoids random error in estimation that arises from using directly observed results when the number of observations is not large, for example if there were no false-positives among 100 unaffected pregnancies in a study sample, it should not be taken to mean that there will be no false-positives in the population at large. We provide observed and modelled results so that a comparison can be made between the two methods. Screening performance using modelled DNA counts was estimated by simulating data on 100,000 affected and unaffected pregnancies based on (i) the distribution of live births in England and Wales 2014–16 [13], (ii) the distributions of fetal fraction in affected and unaffected pregnancies and (iii) the fetal-fraction-specific Gaussian distributions of chromosome 18 MoM values using (a) all autosomes as the denominator, with and without an all autosome correction for GC content, and (b) chromosome 8 as the denominator. The likelihood ratio and risk of being affected was then calculated as with the observed data from the 67 affected and 83 unaffected pregnancies i.e. using observed DNA counts.

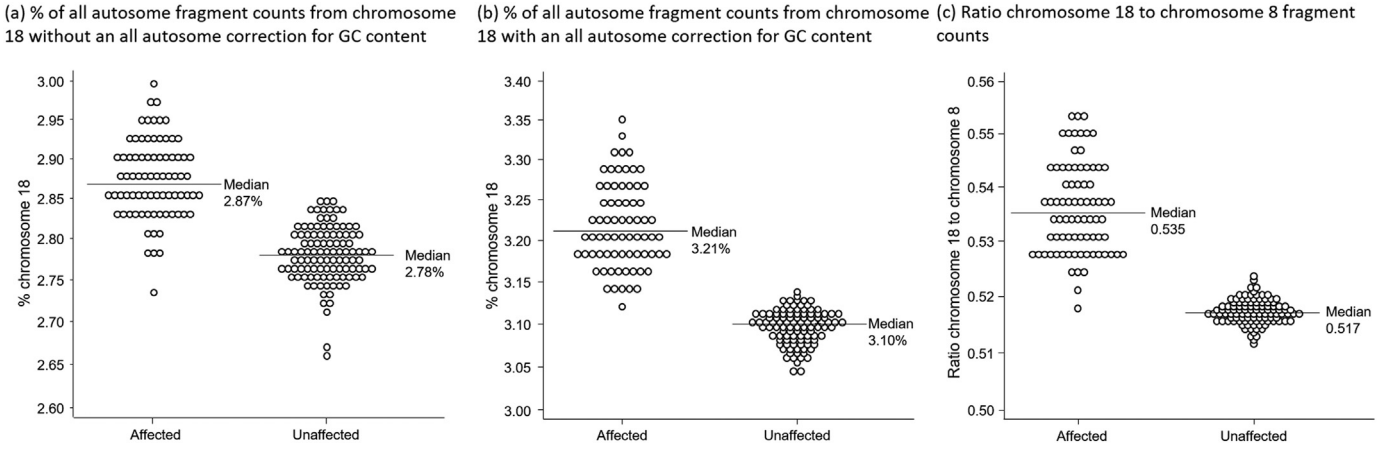

Fig. 1. Percentage of DNA fragments in maternal plasma from chromosome 18 (expressed as a percentage of all autosomes) without (a) and with (b) an all autosome correction for GC content and (c) ratio of DNA fragment counts in maternal plasma from chromosome 18 to fragment counts from chromosome 8 in 67 affected and 83 unaffected pregnancies.

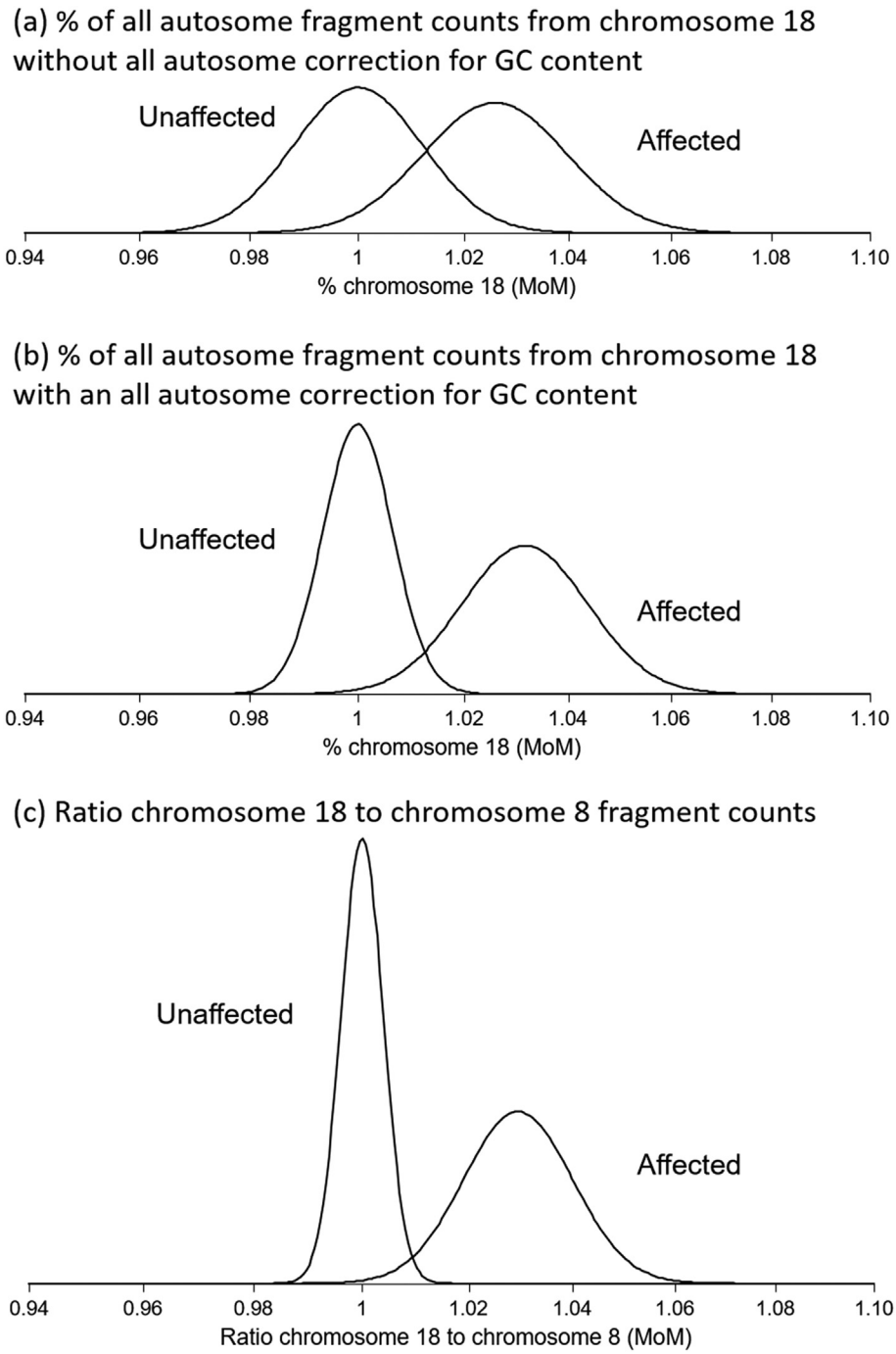

Fig. 2. Relative frequency distributions of DNA fragment counts in maternal plasma from chromosome 18 expressed as the percentage of all autosome fragment counts from chromosome 18 without (a) and with (b) an all autosome correction for GC content and (c) the ratio of chromosome 18 to chromosome 8 counts (all standardised to a fetal fraction of 6%; in multiples of the unaffected median [MoM] values; on a log scale).

3. Results

Fig. 1a shows the percentage of chromosome 18 DNA fragments in maternal plasma (expressed as a percentage of fragments from all autosomes) without any correction for GC content based on 67 affected and 83 unaffected pregnancies. The figure shows higher values in affected pregnancies (median 2.87%) than in unaffected pregnancies (median 2.78%), but there is considerable overlap in values. Fig. 1b shows the same, but with an all autosome correction for GC content. The range of values in unaffected pregnancies is much reduced and there is little overlap between the values for affected (median 3.21%) and unaffected pregnancies (median 3.10%). Fig. 1c shows the ratio of fragment counts from chromosome 18 to chromosome 8. The range of values in unaffected pregnancies is further reduced and there is little overlap between the values for affected (median 0.535) and unaffected pregnancies (median 0.517).

Appendix Fig. 1 shows the relationship between percent chromosome 18 expressed as MoM values using all autosomes as the denominator without a correction for GC content in affected pregnancies according to fetal fraction together with a regression line. MoM values increased by 0.26% for each percentage point increase in fetal fraction ($p < 0.001$). Appendix Fig. 2 shows the same, but with an all autosome correction for GC content. MoM values increased by 0.32% for each percentage point increase in fetal fraction ($p < 0.001$). Appendix Fig. 3

Table 1
Detection rates (DRs) and false-positive rates (FPRs) using observed DNA fragment counts and modelled counts according to trisomy 18 risk cut-off and expressing chromosome 18 (chr18) DNA fragments counts as the percentage of counts from all autosomes without and with all autosome correction for GC content or as the ratio to chromosome 8 (chr8) fragment counts.

Chromosome 18 MoM	DR and FPR for risk cut-off of:-													
	1 in 10		1 in 20		1 in 30		1 in 40		1 in 50		1 in 100		1 in 150	
	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)	DR (%)	FPR (%)
Observed DNA counts														
% of all autosome fragment counts from chr18														
Without correction for GC content	39	0	49	0	52	0	55	0	60	1.2	64	1.2	73	1.2
With all autosome correction for GC content	84	0	88	0	90	0	91	0	91	0	93	0	95	0
Ratio chr18 to chr8 fragment counts	97	0	97	0	97	0	97	0	97	0	97	0	97	0
Modelled DNA counts														
% of all autosome fragment counts from chr18														
Without correction for GC content	38	0.04	46	0.18	51	0.30	55	0.43	57	0.53	66	1.1	71	1.6
With all autosome correction for GC content	85	0.03	87	0.06	88	0.08	89	0.1	89	0.12	91	0.25	92	0.36
Ratio chr18 to chr8 fragment counts	94	0.01	95	0.02	95	0.03	95	0.04	95	0.05	96	0.1	96	0.15

shows the same, but using the ratio of chromosome 18 to chromosome 8 counts. MoM values increased by 0.33% for each percentage point increase in fetal fraction ($p < 0.001$). The regression line equations used to define the fetal fraction-specific median MoM in affected pregnancies, and the standard deviations of MoM values in affected pregnancies are given in Appendix Table 1. The standard deviation of $\log_{10}$ MoM values in unaffected pregnancies was statistically significantly smaller for the ratio of fragment counts from chromosome 18 to chromosome 8 than the percentage of chromosome 18 DNA fragments using all autosomes as the denominator with an all autosome correction for GC content ($p = 0.002$).

Appendix Figs. 4, 5 and 6 show probability plots of MoM values using chromosome 18 counts as a percentage with an all autosome denominator without and with correction for GC content and using the ratio of chromosome 18 to chromosome 8 counts respectively, with MoM values in affected pregnancies standardised to a fetal fraction of 6%. All the plots are reasonably linear indicating a good fit to Gaussian distributions. The plots for affected and unaffected pregnancies in Appendix Fig. 6 are "flatter", indicating smaller standard deviations using the ratio of chromosome 18 to chromosome 8 counts than chromosome 18 counts as a percentage with an all autosomal denominator either with or without a correction for GC content.

Fig. 2 shows the modelled distributions (using medians and standard deviations derived from Appendix Figs. 4, 5 and 6) using chromosome 18 counts as a percentage with an all autosome denominator without (Fig. 2a) and with (Fig. 2b) an all autosome correction for GC content and using the ratio of chromosome 18 to chromosome 8 counts (Fig. 2c), with MoM values in affected pregnancies standardised to a fetal fraction of 6%. The figure shows greater discrimination (Fig. 2c) using the ratio of chromosome 18 to chromosome 8 which arises from the smaller standard deviation in unaffected pregnancies and hence reduced overlap between the distributions in affected and unaffected pregnancies.

Table 1 shows the trisomy 18 detection rates and false-positive rates based directly on the observed data and based on the modelled results using the distribution parameters in Appendix Table 1. The observed and modelled results are similar, for example at a 1 in 50 risk cut-off, using the ratio of chromosome 18 to chromosome 8 counts the detection rate is 97% (65/67) and false-positive rate is 0% (0/83) compared with the modelled estimates of 95% and 0.05% respectively. There is a clear improvement in performance when the ratio of chromosome 18 to chromosome 8 counts is used instead of using chromosome 18 counts as a percentage with an all autosome denominator either with or without an all autosome correction for GC content. For example, at a risk cut-off of 1 in 50 the DR is 95% and the FPR 0.05% using the ratio of chromosome 18 to chromosome 8 counts compared with a DR of 89% and an FPR of 0.12% using chromosome 18 counts as a percentage with an all autosomal denominator with correction for GC content and a DR of 57% and an FPR of 0.53% using chromosome 18 counts as a percentage with an all autosomal denominator without correction for GC content. Table 2 shows the detection rates for specified false-positive rates and false-positive rates for specified detection rates using the modelled data. For example, for a false-positive rate of 0.25% the use of the ratio of chromosome 18 to chromosome 8 counts yields a 97% detection rate compared with 91% and 49% using chromosome 18 counts as a percentage with an all autosome denominator with and without an all autosome correction for GC content respectively.

4. Discussion

Our results show that in prenatal screening for trisomy 18 the ratio of plasma DNA fragment counts that map to chromosome 18 to DNA fragment counts that map to chromosome 8 is a simple method of allowing for analytical error due to variation in DNA GC content. It is clearly better than making no adjustment for GC content and our results indicate that the method is also better than the conventional method of

Table 2
Detection rates (DRs) according to false-positive rates (FPRs) and FPRs according to DRs using modelled DNA fragment counts according to expressing chromosome 18 (chr18) DNA fragments counts as the percentage of counts from all autosomes without and with all autosome correction for GC content or as the ratio to chromosome 8 (chr8) fragment counts.

Chromosome 18 MoM	DR (%) for FPR of:-					FPR (%) for DR of:-				
	0.125%	0.25%	0.50%	1%	2%	85%	90%	95%	98%	99%
% of all autosome fragment counts from chr18										
Without correction for GC content	42	49	56	64	73	6.1	10	19	32	41
With all autosome correction for GC content	89	91	93	95	96	0.03	0.15	1.1	5.7	12
Ratio chr18 to chr8 fragment counts	96	97	97	98	99	< 0.01	< 0.01	0.03	0.96	3.8

all autosome GC adjustment. Sehnert et al. [5] observed such an advantage without linking it to DNA fragment GC content. Our data take the observation further by estimating the quantitative effect using the ratio of chromosome 18 to chromosome 8 fragment counts without GC correction and finding an improved screening performance; the improvement in performance is considerable. Chromosome 8 has a GC-content that is close to that of chromosome 18 [14]. Using the ratio of chromosome 18 to chromosome 8 fragment counts therefore directly allows for variability in the PCR part of the DNA analysis arising from GC associated error in amplifying the correct number of DNA fragments which can vary from sample to sample, which affect DNA fragment counts. Failure to take account of the GC content of DNA has a clinically significant effect on screening performance as illustrated in Tables 1 and 2.

The conventional method of allowing for GC content is based on algorithms using estimated associations between the number of DNA fragments and GC content of those fragments. The method has the limitation of its complexity and a lack of transparency. The algorithm adopted has to be derived separately for each method of DNA analysis. The method proposed here is simple, transparent, and does not require the use of an algorithm. The method is also generalizable in that any chromosome with a similar GC content to the trisomic chromosome of interest could be used as the reference, using chromosome 7 in screening for trisomy 21 and chromosome 4 in screening for trisomy 13. These would need to be validated empirically as we have done here with trisomy 18 screening, before being adopted in practice.

Truncation limits that limit the range of values used to calculate likelihood ratios and hence risk are often used for markers in prenatal screening. However, when the distributions of marker values in affected and unaffected pregnancies are widely separated, as is the case here, this has the effect of ignoring many of the informative values. In such situations not using truncation limits is appropriate provided the modelled estimates are consistent with those based on simple counting (see Table 1). While the precise risk at high MoM values, for example > 1.05 in Appendix Fig. 6, may be uncertain, the probability of being affected becomes extremely high and has little influence on screening performance. Similarly, at low values, for example < 0.98 the pregnancy will almost certainly be unaffected.

In summary, the use of the ratio of DNA fragment counts that map to chromosome 18 to counts that map to chromosome 8 yields a high level of screening performance and avoids the need for more complex GC correction algorithms.

Declaration of Competing Interest

The authors have no interests to declare.

Acknowledgements

We thank Tiesheng Wu for providing IT support and help.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2019.06.015.

References

[1] G.E. Palomaki, C. Deciu, E.M. Kloza, G.M. Lambert-Messerlian, J.E. Haddow, L.M. Neveux, M. Ehrich, D. van den Boom, A.T. Bombard, W.W. Grody, S.F. Nelson, J.A. Canick, DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as down syndrome: an international collaborative study, Genet. Med. 14 (3) (2012) 296–305.
[2] M.M. Gil, V. Accurti, B. Santacruz, M.N. Plana, K.H. Nicolaides, Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis, Ultrasound Obstet. Gynecol. 50 (3) (2017) 302–314.
[3] E.Z. Chen, R.W.K. Chiu, H. Sun, R. Akolekar, K.C. Allen Chan, T.Y. Leung, et al., Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing, PLoS ONE 6 (2011) e21791.
[4] Y. Benjamini, T.P. Speed, Summarizing and correcting the GC content bias in high-throughput sequencing, Nucleic Acids Res. 40 (10) (2012) E72.
[5] A.J. Sehnert, B. Rhees, D. Comstock, E. de Feo, G. Heilek, J. Burke, R.P. Rava, Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free DNA in maternal blood, Clin. Chem. 57 (7) (2011) 1042–1049.
[6] F. Crea, M. Forman, R. Hulme, R.W. Old, D. Ryan, R. Mazey, M.D. Risley, The Iona test: development of an automated cell-free DNA-based screening test for fetal trisomies 21, 18 and 13 that employs the ion proton sequencing platform, Fetal Diag. Ther. 42 (3) (2017) 218–224.
[7] N.J. Wald, H.S. Cuckle, J.W. Densem, K. Nanchalal, P. Royston, T. Chard, J.E. Haddow, G.J. Knight, G.E. Palomaki, J.A. Canick, Maternal serum screening for Down's syndrome in early pregnancy, BMJ 297 (1988) 883–887.
[8] G.M. Savva, K. Walker, J.K. Morris, The maternal age-specific live birth prevalence of trisomies 13 and 18 compared to trisomy 21 (down syndrome), Prenat. Diagn. 30 (2010) 57–64.
[9] J.K. Morris, G.M. Savva, The risk of fetal loss following a prenatal diagnosis of trisomy 13 or trisomy 18, Am. J. Med. Genet. Part A 146A (2008) 827–832.
[10] N.J. Wald, J.P. Bestwick, W.J. Huttly, Improvements in antenatal screening for Down's syndrome, J. Med. Screen. 20 (2013) 7–14.
[11] J.P. Bestwick, W.J. Huttly, N.J. Wald, Detection of trisomy 18 and trisomy 13 using first and second trimester Down's syndrome screening markers, J. Med. Screen. 20 (2013) 57–65 (Corrigendum in J Med Screen 2015;22:52–4).
[12] N.J. Wald, W.J. Huttly, K.W. Murphy, K. Ali, J.P. Bestwick, C.H. Rodeck, Antenatal screening for Down's syndrome using the integrated test at two London hospitals, J. Med. Screen. 16 (2009) 7–10.
[13] Office for National Statistics, Birth Characteristics in England and Wales, 2015 (2014), p. 106.
[14] http://blog.kokocinski.net/index.php/gc-content-of-human-chromosomes?blog=2.