Genomics 105 (2015) 150–158

Contents lists available at ScienceDirect

Genomics

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

Amplification-free sequencing of cell-free DNA for prenatal non-invasive diagnosis of chromosomal aberrations

Kasper Karlsson [a], Ellika Sahlin [b], Erik Iwarsson [b], Magnus Westgren [c], Magnus Nordenskjöld [b], Sten Linnarsson [a,*]

[a] Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 1, SE-171 77 Stockholm, Sweden
[b] Clinical Genetics Unit, Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital Solna, 17176 Stockholm, Sweden
[c] Karolinska Institutet, Department of Obstet Gynecol, Division of Clinical Science, Intervention and Technology, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 20 December 2013
Accepted 16 December 2014
Available online 24 December 2014

Keywords:
Amplification-free
Library preparation
Cell-free DNA
Karyotyping
Bias
Fetal aneuploidy

ABSTRACT

Cell-free DNA has been used for fetal rhesus factor and sex determination, fetal aneuploidy screening, cancer diagnostics and monitoring, and other applications. However current methods of using cell free DNA require amplification, which leads to allelic dropout and bias especially when starting with small amounts of DNA. Here we describe an amplification-free method for sequencing of cell-free DNA, even from low levels of starting material. We evaluated this method in the context of prenatal diagnosis of fetal aneuploidy and compared it with a PCR-based library preparation method as well as a recently described method using unique molecular identifiers (UMI). All methods performed well, however coverage was increased by the amplification-free method and GC-induced bias was reduced by both the amplification-free method and the UMI method. Future diagnostic applications including whole genome sequencing of cell-free DNA will benefit from amplification-free sequencing.

1. Introduction

The use of fetal cell-free DNA (cfDNA) to assess fetal sex and Rhesus factor status is already a common practice. The next non-invasive prenatal diagnostics that might enter clinical practice on a large scale is assessment of chromosomal aberrations. This is a long awaited improvement since the current methods (amniocentesis and chorionic villus sampling) carry with them a certain risk for miscarriage. Accurate tests of fetal trisomy 21 status are already available commercially in some parts of the world, including USA and China. However these tests have only guiding value, and are not used as a definite diagnostic tool [1], since tests of trisomy 21 status need to be extremely accurate in order to be accepted diagnostically [2], especially if they are to be used as screening tools. A fundamental limit to the use of cell-free DNA in diagnostics is that it derives from the placenta, and there is a non-negligible risk that the placental genotype differs from the embryonic. In the future, cell-free DNA may be used for the detection of more complex or smaller genetic aberrations, such as microdeletions, amplifications and single-nucleotide mutations. These future applications place even stronger demands on the quantitative accuracy, whole-genome coverage and minimal error rate of cell-free DNA sequencing methods.

The main problem in detecting fetal genetic aberrations from maternal cell-free DNA is that the fetal content of cfDNA is low compared to the maternal background. 10% has been reported as a median value [3], and current methods have a cut off rate of minimum 4% fetal content for performing the analysis [4]. Fetal content of cfDNA has also been shown to correlate negatively with maternal weight [4], so one of the reasons for low fetal cfDNA content is simply dilution.

Although PCR-based protocols for DNA sequencing work fairly well for cell-free DNA, they include a step of PCR amplification which introduces a bias [5] due to unequal amplification of chromosomal regions and alleles [6]. As a result, some chromosomal regions may be lost (primarily those that are AT rich) and secondary structures, GC content, fragment length and local repeats may all prevent detection of specific parts of the genome such as single-nucleotide variants, short indels and larger microdeletion syndromes. Another problem related with PCR-based protocols is that the number of duplicate reads increases, which increases costs and lowers mapping rate [7]. An ideal sample preparation protocol for prenatal noninvasive diagnosis should avoid introducing bias, cover the entire genome uniformly, and be applicable to the small amounts of cell-free DNA that can be obtained by routine blood sampling.

We adapted a protocol [8,9] developed for sequencing of archaic DNA to sequencing of cell free DNA for non-invasive prenatal karyotyping. The protocol was modified in several ways to allow for amplification-free library preparation.

* Corresponding author.
E-mail address: Sten.linnarsson@ki.se (S. Linnarsson).

http://dx.doi.org/10.1016/j.ygeno.2014.12.005
0888-7543/

Another way to nearly eliminate PCR bias is to use unique molecular identifiers [10–14]. This method, where each molecule is made unique either by adding a degenerate barcode or by diluting the sample before amplification, corrects for most types of bias, including sequencing errors and quantitative bias. In the context of cfDNA shotgun sequencing we [14] have shown that unique molecular identifiers can nearly eliminate amplification bias; however those previous results used unrealistically high levels of fetal DNA, and thus did not explore the performance of UMIs when fetal DNA is highly diluted.

2. Results

2.1. An amplification-free library preparation protocol for cell-free DNA

Cell-free DNA has unique properties: it consists of partially single-stranded, short DNA fragments, typically in the range of 160–200 bp [15]. We adapted a previously published protocol [8,9] originally developed for ancient DNA, which has similar properties. This protocol is based on efficient single-stranded adapter ligation, and thus works on any single-stranded DNA. We optimized the protocol to eliminate the PCR amplification step, and thereby reduce bias. Several important changes to the protocol were introduced to reduce input DNA requirements and eliminate PCR: (1) P1 and P2 adaptors are introduced directly by ligation and not in the PCR step. (2) The P2 adaptor contains a four base pair long degenerate overhang, which blocks ligation of the P1 adaptor directly to the P2 adaptor that would otherwise constitute a large part of the final library (especially when using a low input amount). (3) The stringency wash temperature was also set to 35 °C for the same reason. (4) The yield of the samples will typically be well below the threshold of sequencing according to standard Illumina protocols. A custom denaturation protocol was developed, that makes more efficient use of the available sample while still resulting in fully populated sequencing flowcells.

Fig. 1 shows an outline of the method. The different steps are discussed in more detail in the Supplementary methods.

2.2. Amplification-free sequencing of clinical prenatal cell-free DNA samples

The amplification-free library preparation was evaluated on clinical samples. 31 samples were sequenced, four were used as controls and 15 had one or more chromosomal aberrations. All autosomal trisomies were correctly classified (Fig. 2). However, surprisingly, for the sex chromosomes, sample number 3 was classified as XY, contradicting the results of QF-PCR, which indicated a XXY genotype: Fig. 3a shows the copy number values per chromosome for sample 3. The sample appears to be a trisomy 18–XY where the increase of reads on chromosome 18 is closely matched by a similar decrease of reads on chromosome X, as would be expected. To understand this discrepancy, we obtained genomic DNA from fetal lung and examined it for aneuploidy both by array comparative genomic hybridization (CGH) and by amplification-free sequencing. In both cases, these tests confirmed a mosaic nature of the genomic DNA, where around 30% of the cells had sex chromosomes XY and 70% had sex chromosomes XXY (Fig. 3b and c). Fetal cfDNA is thought to originate from placental or fetal tissue [1] but unfortunately no placental tissue was saved for this donor. Considering the mosaic nature of the fetus it is plausible that the XY profile observed by amplification-free sequencing of cfDNA accurately reflected the genomic status of the placenta, rather than being a technical error.

All other trisomy samples and the healthy controls were correctly classified by using Z-score statistics (shown in Fig. 2b).

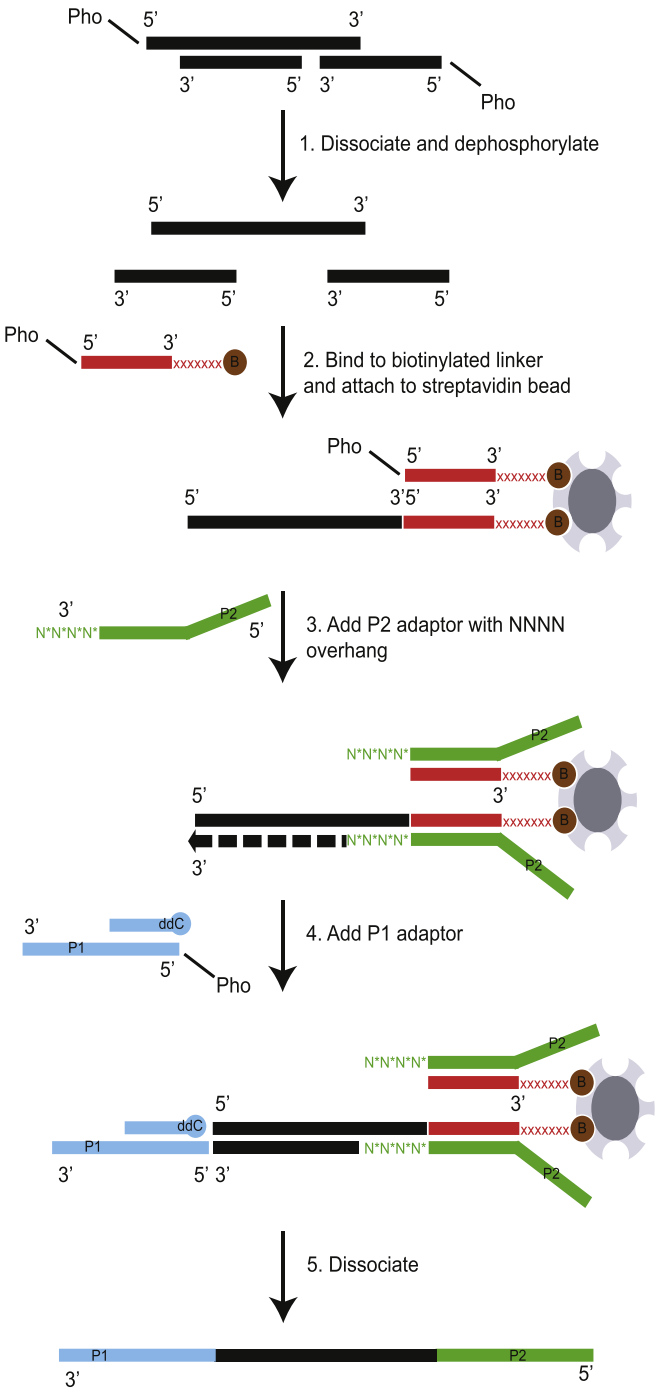
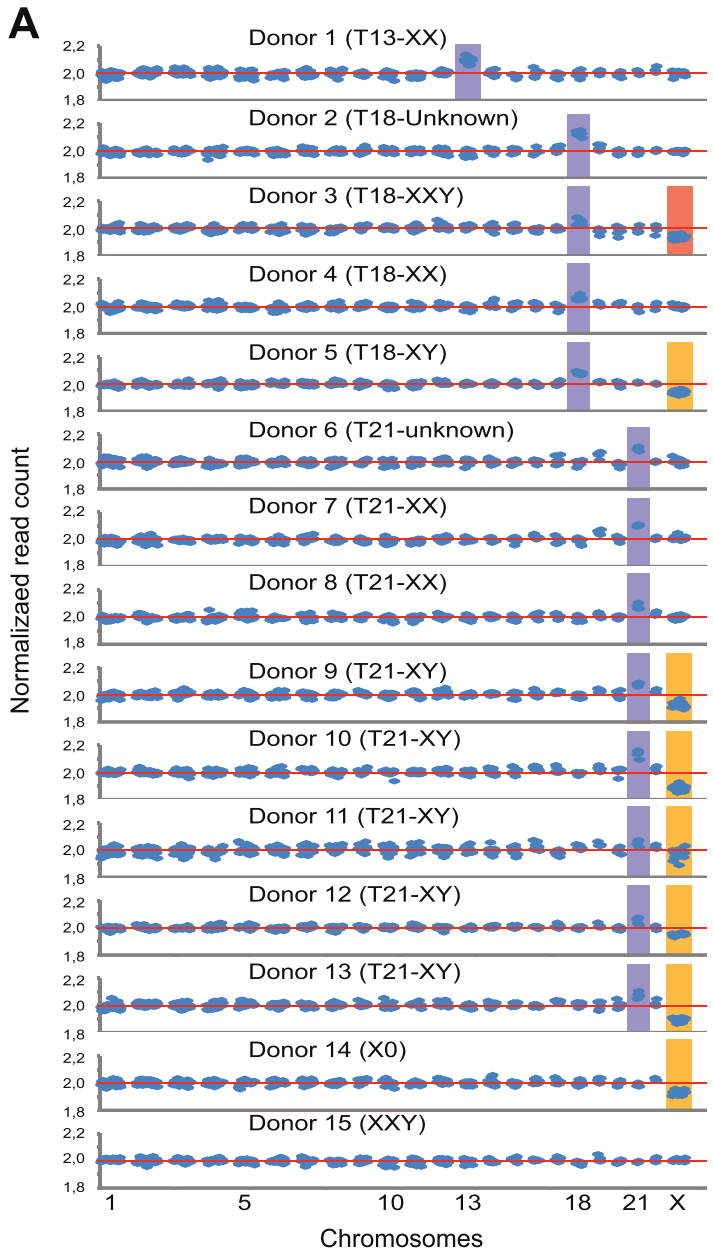


Fig. 1. General outline. Step 1: The cell-free DNA is denaturated and dephosphorylated. Step 2: A biotinylated linker is added and ligated to each cfDNA molecule. The linker and cfDNA molecule complex is immobilized to a streptavidin bead. Note that the linker is abundant, and hence a large number of linkers will not be ligated to a cfDNA molecule. Step 3: The first adaptor, which is single stranded and comprise of a random nucleotide overhang, a part that is complementary to the linker, and the Illumina P2 adaptor, is introduced to the sample. The adaptor is extended by a polymerase and creates a double stranded complex. Step 4: The second adaptor is introduced to the sample. The second adaptor is double stranded and the lower strand contains a 5 prime phosphate, and the upper strand contains a 3 prime dideoxy nucleotide. The second adaptor is ligated to the double stranded complex. Note that the second stranded adaptor cannot ligate to an empty, non-ligated linker. Step 5: The strands are dissociated and the free non-bound strand is collected for subsequent sequencing.

2.3. Comparison between the different library preparation methods

To get an estimate of how good a library preparation method is, variance in read density between bins can be measured. For this purpose Coefficient of Variation was used (CV, standard deviation divided by the mean, methods). CV was calculated for the different methods before normalization, after normalization, and after normalization and loess local regression (Fig. 4a–c). For the amplification-free and UMI


A
Donor 1 (T13-XX)
Donor 2 (T18-Unknown)
Donor 3 (T18-XXY)
Donor 4 (T18-XX)
Donor 5 (T18-XY)
Donor 6 (T21-unknown)
Donor 7 (T21-XX)
Donor 8 (T21-XX)
Donor 9 (T21-XY)
Donor 10 (T21-XY)
Donor 11 (T21-XY)
Donor 12 (T21-XY)
Donor 13 (T21-XY)
Donor 14 (X0)
Donor 15 (XXY)
2,2
2,0
1,8
Normalizaed read count
1
5
10
13
18
21
X
Chromosomes
B
Z-score
14
12
10
8
6
4
2
-0
-2
-4
Trisomy
Normal copy number
Z-score = 3
13
18
21
Chromosomes

methods, the variance was reduced compared to the PCR-based method, which was expected due to the lower bias produced by these methods. The amplification-free method still showed some bias compared to the UMI method (Fig. 4d). This may stem from bias in the bridge amplification step during sequencing, which is corrected for by the UMI method but not the amplification-free methods. Other sources of bias may be in the stringency wash step in the amplification-free method. After normalization and loess, all methods approached the theoretically lowest possible CV (i.e. the limit of purely random sampling).

Neither the amplification-free method nor the UMI method outperformed the PCR-based method in terms of CV after normalization, even though the GC bias created by those methods was lower (Fig. 4a and d). We hypothesize that this is because the bias produced by amplification was very reproducible, at least for GC content (Fig. 4d Standard vs standard), and thus could very efficiently be removed by normalization. All three methods performed very close to the theoretical limit, confirming their accuracy in the context of chromosomal copy number detection.

2.4. Reduction in GC bias and increase in genome coverage by amplification-free library preparation

Although all methods were equally accurate in calling chromosomal aberrations, certain applications require full and unbiased coverage of specific genomic regions. If such a region is lost during amplification, it cannot be rescued by normalization. PCR is known to amplify molecules with average GC values better than molecules with extreme GC values. To investigate potential losses as a function of GC content, we plotted the log read count against the GC content of those reads for the amplification-free method and the PCR-based method (Fig. 5a). All reads on chromosome 1 from the PCR-based library preparation method were combined and compared to the combined reads from the amplification-free samples. The samples were trimmed so there would be an equal amount of reads from each: 6.7 million reads on chromosome 1, which is slightly more than 1× coverage. The PCR-based method recovered almost no molecules with a GC content less than 12%, while the amplification-free method recovered low GC content molecules reasonably well. To see how this difference in recovery affected coverage, the difference in coverage between the combined amplification free and PCR-based samples was plotted against the variation of GC content from the average along chromosome 1 (Fig. 5b). In regions where the GC content for a 1 Mbp window was 5% less than average, the amplification-free method showed a 5–10% higher coverage. The coverage statistics are summed up in Table 2.

3. Discussion

We have developed a new library preparation method that allows for amplification-free sequencing of cell-free DNA. There are a number of amplification-free library preparation methods, but to our knowledge none of them work on the small amount of DNA that can be extracted from cell-free DNA. Also these methods are developed for genomic DNA and take double stranded DNA sequences as input material, while our method is developed for prefragmented cell-free DNA and takes single stranded DNA as input material. Examples of amplification-free library preparation methods are Illumina's amplification free protocol that requires 1 μg of input DNA, and Kozarewa et. Al. method [7] that requires 500 ng input DNA. Our amplification-free method can be used on small quantity input DNA – 50 ng can be sufficient for generating around 400 million reads – and have the same benefits as other amplification free protocols: it reduces GC bias, increases coverage, maintains strand information and can be used not only on cfDNA but also genomic DNA, as we showed when we sequenced fetal lung genomic DNA.

We have compared the amplification-free method with another method developed in our lab, the UMI method, and to a PCR based library preparation method. Table 3 shows a summary of advantages/disadvantages.

We envision that the different methods will be useful in different situations. When there is a need for extremely accurate sequencing and both PCR and sequencing errors need to be corrected for, UMI is the method of choice. An example when this is needed is in targeted sequencing of mutations in leukemia residual disease [13]. The amplification-free method would be the method to use when coverage is important, when partially degraded input material is used, or when there is a need of low bias sequencing combined with low cost and scarce material. Applications where this could be useful include whole genome sequencing and sequencing of formalin fixed paraffin embedded (FFPE) material. The PCR-based method has benefits in terms of time to prepare samples, robustness and ease of use.

4. Materials and methods

4.1. Materials

A clinical evaluation of the amplification-free method was done on 31 samples. Whole blood was collected from pregnant women that had received a positive CUB (combined ultrasound and biochemistry) score. 10 ml blood was collected from each mother and the tube was centrifuged shortly after blood collection. The plasma layer was transferred to fresh tubes and stored in −20 °C until the time for DNA extraction.

Cell-free DNA was extracted from 4 ml plasma using the QIAsymphony DSP Virus/Pathogen Kit on the automatized QIAsymphony RGQ system (Qiagen), using a custom script supplied by the manufacturer.

15 donors carried a fetus with one or more trisomy chromosome, and 16 donors carried healthy fetuses. 4 of the 16 samples with healthy fetus were used for normalization. These samples are called the “amplification-free” samples.

For comparison, cfDNA was extracted from two more healthy donors and were subjected to an amplification based library preparation protocol. These samples are called “PCR-based” samples.

The UMI sample was prepared from genomic DNA from a mother and a male with trisomy 21 as previously described [14], with the difference that it was mixed with 95% maternal DNA and 5% fetal DNA. These samples are called “UMI” samples.

The studies were conducted with permission from the regional ethics committee at Karolinska Institutet and the patients gave written consent to participate. All data files were deposited to the sequence read archive.

4.2. Amplification-free method Library preparation for cell free DNA

4.2.1. Dephosphorylation, heat denaturation and single-stranded adaptor ligation

Libraries were prepared in a PCR free room. 17.1 μl cell free DNA was mixed with 4.8 μl CircLigase II 10× Reaction buffer (Epicenter), 2.4 μl

Fig. 2. Fetal karyotype inferred from amplification free sequencing of maternal cell free DNA. A) X-axis represents chromosomes 1–22 and X, Y-axis represents apparent chromosomal copy numbers. Each dot contains reads from a window with a constant read number of 100,000 in the control sample. The shaded areas represent expected and unexpected deviations in chromosome copy number. The purple shades represent trisomy deviations. The orange shades represent X chromosome deviations due to male pregnancy. The red shade represents erroneous deviation. Only samples with one ore more aneuploidies are shown B) Plot showing z-scores for chromosomes 13, 18 and 21. A cut off of $z = 3$ was used to identify aneuploidy. Red dots represent fetuses with one of the trisomies, blue dots represent fetuses that have a normal karyotype for that chromosome.

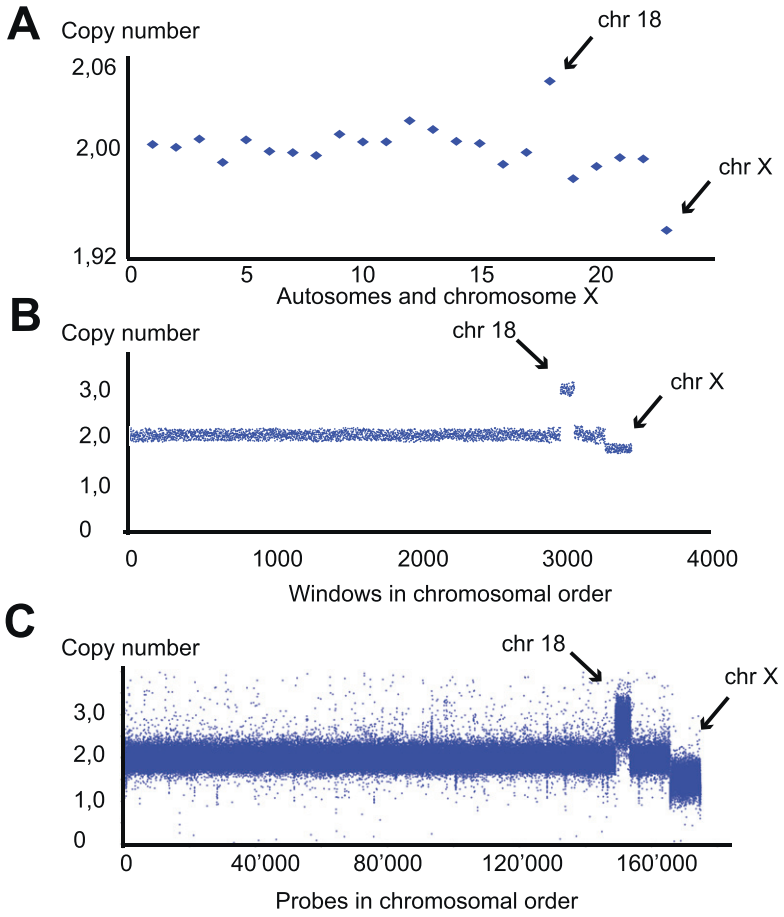


Fig. 3. Confirmation of karyotype of donor 3. Arrows point at chromosomes called aberrant. A) Copy numbers measured by amplification free sequencing of cell free DNA from donor. B) Copy numbers measured by amplification free sequencing of DNA from lung tissue of the fetus from donor 3. C) Copy numbers measured by comparative genomic hybridization of DNA from lung tissue of the fetus from donor 3.

50 mM MnCl2 (Epicenter) and 0.3 μl RNAse Cocktail (Life Technologies) in a 200 μl PCR strip (Starlab) and incubated at 37 °C for 30 min in a thermocycler. 0.6 μl (1 U) FastAP (Thermo Scientific) was added and the mixture was incubated at 37 °C for 5 min and then at 95 °C for 2 min. The reaction mix was placed immediately in on ice. 0.6 μl (1 U) FastAP was added one more time and the mixture incubated at 37 °C for 10 min, then 75 °C for 10 min and finally at 95 °C for 2 min. The reaction mixture was immediately placed in on ice. 19.2 μl 50% PEG-4000 was added together with the oligo 1 μl "Linker 1" with a concentration of 10 μM. The reaction was mixed by vortexing and 2.4 μl CircLigase II (Epicenter) was added to the tube and the reaction was incubated at 60 °C for 1 h. The samples were put on ice and 2.4 μl 1% Tween-20 was added.

4.2.2. Immobilization of ligation products on streptavidin beads

20 μl MyOne Streptavidin C1 (Life Technologies) beads were washed twice with 200 μl 1 × BWT + SDS (100 mM NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.05% Tween). The beads were resuspended in 150 μl 1 × BWT + SDS. The reaction mix was heated to 95 °C for 2 min and then immediately put on ice. Then the reaction mix was centrifuged and the beads were added to it. The mix with the beads was rotated for 20 min at room temperature, then the supernatant was removed, and the beads were washed once with 200 μl 0.1 × BWT + SDS (100 mM NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.05% Tween-20, 0.5% SDS) and once with 200 μl 0.1 × BWT (100 mM NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.05% Tween).

4.2.3. Primer annealing and extension

The supernatant was removed and the beads were resuspended in 17 μl reaction mixture containing 12.75 μl water, 2 μl Thermopol buffer 10× (New England Biolabs), 1.25 μl of each dNTP (10 mM) and 1 μl of "Fill-in block primer" 100 μM, and incubated at 65 °C in a thermal cycler. The mixture was immediately chilled on ice, and then 3 μl (24 U) Bst DNA polymerase, large fragment was added. The reaction was incubated by ramping the temperature from 15 °C to 37 °C at a speed of 1 °C/min and then holding at 37 °C for 5 min. The beads were kept in suspension by mixing or vortexing at least one time during the incubation. The supernatant was discarded, and the beads were washed with 200 μl 0.1 × BWT + SDS. The beads were resuspended in 200 μl stringency-wash (0.1× SSC buffer (Sigma), 0.1% SDS), and incubated at 35 °C for 3 min in a thermal cycler, then chilled down to 4 °C. The stringency wash and incubation were repeated two more times, and then the mixture was washed with 200 μl 0.1 × BWT.

4.2.4. Removal of 3 prime overhang

The beads were resuspended in 19 μl reaction mix containing 15.75 μl water, 2 μl Tango buffer 10× (Thermo Scientific), 0.25 μl Tween-20 1% and 1 μl each dNTP 10 mM. 1 μl (5 U) *T4* DNA polymerase (Thermo Scientific) was added and the reaction was incubated at 25 °C for 15 min in a thermal cycler. The supernatant was discarded and the beads were washed with 200 μl 0.1 × BWT + SDS, resuspended in 200 μl stringency-wash, incubated at 35 °C for 3 min in a thermal cycler and then washed with 200 μl 0.1 × BWT.

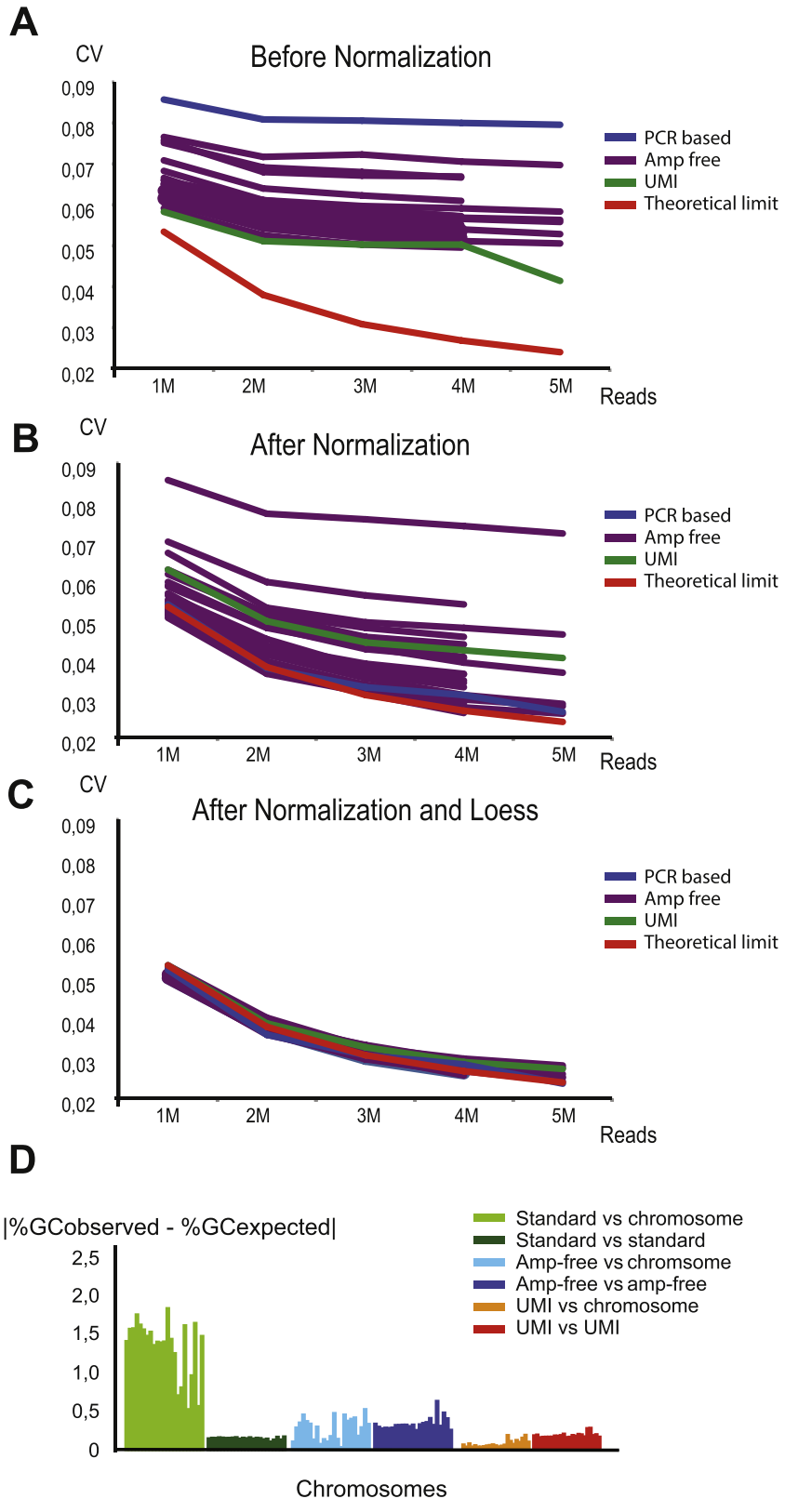


Fig. 4. No difference between methods after normalization and GC correction. For each sample 1 to 5 million reads were randomly extracted. X-axis represents increasing number of reads, Y-axis represents the CV value calculated from the variation of the number of reads per 1 million base pair bins within each sample. The blue line denotes the standard sample, the green denotes the UMI sample, the purple lines denote the amplification free samples and the red line denotes the theoretical best CV value based on pure sampling error. Note that only amplification free samples containing 4 million or more mapped reads are included. A) CV value before normalization; B) CV value after normalization; C) CV value after normalization and Loess regression. D) Difference in observed versus expected GC content by chromosome. Light colors, sample compared to ideal genome sequence. Dark colors, two samples compared to each other. Green, standard sample preparation. Blue, amplification-free sample preparation. Red, UMI sample preparation.

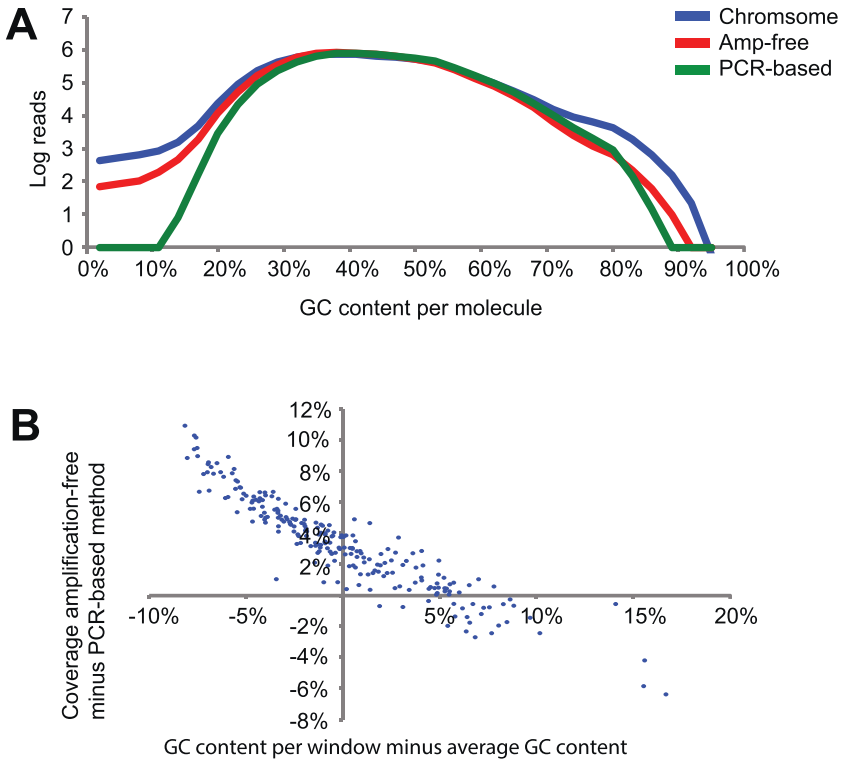


Fig. 5. Amplification-free method recovers molecules with low GC content. A) Read coverage as a function of GC content. The Y-axis represents normalized log reads for chromosome 1. The X-axis represents bins of GC content. First the normalized log reads were calculated in bins of 1 percental unit in size and then the average value of three bins were used for the plot. Blue line, ideal chromosomal DNA. Red line, amplification-free library preparation. Green line, PCR-based library preparation. The X-axis denotes the chromosomal GC content per window subtracted with the average GC content of chromosome 1. The Y-axis denotes the fraction of bases covered per window for the amplification free method subtracted by the fraction of bases covered by the PCR based method.

4.2.5. Blunt end adaptor ligation

The beads were resuspended in 14 μl of a reaction mix containing 9.5 μl water, 2 μl T4 DNA ligase buffer 10× (Thermo Scientific), 2 μl PEG-4000 50%, and 0.5 μl Tween-20 1%. 2 μl Blunt-end index-Adaptor 100 μM was added and the reaction was mixed thoroughly by vortexing. 4 μl (5 U) T4 DNA ligase (Thermo Scientific) was added and the mixture was incubated at 25 °C for 1 h in a thermal cycler. The beads were kept suspended by gently mixing at every 20 min during the incubation. The supernatant was discarded and the beads were washed with 200 μl 0.1 × BWT + SDS, resuspended in 200 μl stringency-wash, incubated at 35 °C for 3 min in a thermal cycler and then washed with 200 μl 0.1 × BWT.

4.2.6. Elution

A stock solution of 0.1 M NaOH and another stock solution containing 0.1 M HCl diluted in a special hybridization buffer (Standard Illumina hyb buffer HT1 and 5 mM Tris pH 7,6) were made. 10 μl 0.1 M OH and 10 μl 0.1 M HCl were mixed and the pH was checked with a pH-stick. The amount of HCl stock solution was varied until a satisfying pH (7–8) was reached. The beads were resuspended in a mix of 10 μl 0.1 M OH and 10 μl EBT. The beads were mixed by pipetting and the reaction was incubated 5 min at room temperature. 20 μl of denatured template was transfered to a tube on ice containing X μl of the HCl stock solution, where X was the amount decided from the pH testing. 20-X μl cold Illumina HT1 buffer with 5 mM Tris-HCl pH 7.6 were added to the mixture. pH was checked again and should be around 7–8.

The concentration of each sample was checked with KAPA library Quant (KAPA Biosystems) and pooled. Note that the samples at this point are denatured and diluted and thereby already ready to put on the Illumina Sequencing machine. The concentration of the samples varied, the average was 165 pM (+-394).

4.3. UMI sample library preparation

The UMI sample was prepared from genomic female DNA where 5% genomic DNA was added from a boy with trisomy 21. As a control only the female genomic DNA was used. The samples were prepared with enzymatic fragmentation, adaptor ligation and PCR as previously described [14]. The PCR reactions were run 17 cycles with Phusion HF. The samples were aliquoted before PCR, aiming to obtain ~10 million UMIs. The actual number of UMIs was 8.6 (151 million reads) for the

Table 1
Sequencing information.

Sample	Mapped reads (million)	Unique positions (million)	Mapped reads/unique positions
Standard	41.7	39.1	1.07
Standard control	41.5	37.9	1.10
UMI	256	13.1	19.5
UMI control	151	8.6	17.6
Amplification free (average)	4.5	4.5	1.01
Amplification free (range)	1.7–10.4	1.7–10.3	1.00–1.01

Table 2
Comparison of coverage between the amplification free method and the standard method.

	Ampfree all samples	Standard and standard control
Reads	6.8 M	6.8 M
Positions	6.5 M	6.1 M
Number of bases covered	151 M	144 M
Bases covered per read	22.3	21.3
Total bases covered of chromosome 1	60.5%	57.7%

All information in Table 2 comes from chromosome 1 only.

Table 3
Comparison of benefits using different library preparation methods.

	Ampfree	UMI	Standard method
Near perfect quantification	Yes	Yes	Yes
Uniform genome coverage	Yes	No	No
Correction of PCR bias	Yes	Yes	No
Correction of sequencing errors	No	Yes	No
Strand information	Yes	No	No
Relative reads required	1×	10×	1×

control sample and 13.1 million (256 million reads) for the sample with spiked in trisomy DNA.

The 5 prime genomic position of each mapped read was used as the UMI. Table 1 shows the number of reads and UMIs for the different samples. The libraries were intentionally made from a small aliquot and sequenced until all molecules were observed multiple times to improve precision and accuracy of the UMI method.

4.4. PCR-based library preparation

To compare the new methods with a normal library preparation method one sample and one control were prepared in the same way as the UMI sample (using the NEBnext kit for Illumina library preparation) but without taking an aliquot before PCR. The DNA used was cell free DNA from a mother with a male fetus. As a control cfDNA from a mother with a female fetus was used.

4.5. Amplification-free sequencing data analysis

The samples were sequenced with 43 to 45 bp length, but all samples were trimmed down to 43 bp length. Since the mapping should be done with no mismatch and only uniquely mapping reads the alignment could be speeded up by extracting all possible sequences with 43 base pairs from the hg19 genome. These were mapped back to the same genome with bowtie allowing for only uniquely matching sequences. The resulting sequences then constituted a database against which the sample reads were matched. In the database the GC content for each possible read was retained and extended to in total 167 base pairs, which has been shown to be close to the average length of a cell free DNA fragment [15,16].

Of the 31 amplification-free samples sequenced four samples with healthy female fetus were chosen for normalization. Windows with a constant number of 10,000 unique reads and variable number of base pairs were created from one of the normalizing samples, and then average number of unique reads from all the four normalizing samples were counted within each window. For each sample the number of reads and the average GC content were counted for each window. The number of reads per window was centered around one, based on the average number of reads per window on all chromosomes with the exception of sex chromosomes, potential trisomy chromosomes 13, 18 and 21, and chromosome 16 and 19 that sometimes showed unregular number of reads. Loess was used to correct for residual GC bias and was calculated based on the number of reads and the average sample GC content per window, and implemented through the R package “loess” with a smoothing parameter of 0.3. To exclude outliers the 10% windows with lowest and highest number of normalized reads were removed.

4.6. Comparison between methods

All samples were mapped against the database created earlier with only uniquely matching fragments. The genome was divided into one million base pair bins, and each sample's reads were counted within each bin. Coefficient of Variation was used for comparison and was calculated on all bins in the genome except for bins in the chromosomes with a potential aberration (chr 13, 18, 21, X and Y). Chromosomes 16 and 19 were also excluded since the amplification-free method showed deviating results for these chromosomes, especially chromosome 19. CV was calculated on unique reads, after normalization and after normalization and loess (Fig. 4a–c)

4.7. Statistical analysis

Z-score has been used in a number of publications to determine trisomy status for chromosomes with massive parallel sequencing data for cell free DNA [17]. Since the sample size was small and the trisomy status was known beforehand, the mean and standard deviation for the population for the chromosomes of interest (13, 18 and 21) were taken from all samples known not to be a trisomy for that particular chromosome. For our samples a Z-score of 3 could distinguish between the true trisomies and the healthy controls (Fig. 2b). Z-score for the X chromosome was not calculated due to limited number of female fetuses.

4.8. QF-PCR analysis of aberrant sample

The sample was analyzed with quantitative fluorescence–polymerase chain reaction (QF-PCR) for detection of aneuploidies involving chromosomes 13, 18, 21, X and Y. DNA was extracted directly from fetal lung tissue, and the extraction was performed using the InstaGene Matrix protocol (Bio-Rad), and analyzed using a QF-PCR panel for the investigation of fragment length of specific short tandem repeats (STRs) as previously described [18]. Four specific STRs on chromosome 13 (D13S305, D13S634, D13S628, D13S742), five on chromosome 18 (D18S386, D18S391, D18S499, D18S535, D18S978), four on chromosome 21 (D21S11, D21S1270, D21S1411, D21S1435), seven on the X (AMELX, DXS981, DXS996, DXS1187, DXS1283E, P39, XHPRT), three on the Y chromosome (AMELY, DYS448, SRY), and one STR present on both the X and Y chromosomes (X22) were included in the analysis. The analysis results were analyzed using GeneMapper Software 5 (Applied Biosystems).

4.9. Array analysis of aberrant sample

One sample showed discordant result between the amplification-free method and the QF-PCR analysis. This sample was further analyzed with a 180 K oligonucleotide array with evenly distributed whole-genome coverage (Oxford Gene Technology). The DNA sample from fetal lung tissue and a pooled reference DNA from ten healthy males (Promega) were fluorescently labeled with Cy3 and Cy5 (Enzo Life Sciences), respectively. The microarray glass slides were scanned in a 3 μm resolution microarray scanner. The scanned image files were processed and checked for quality control (QC) metrics using Feature Extraction software version 10.7.3.1 (Agilent Technologies), in accordance with the guidelines from the manufacturer. Analysis of copy number variants (CNVs) was performed using the CytoSure Interpret Software, version 4.1.9 (Oxford Gene Technology) with data aligned to the human reference sequence GRCh37/hg19.

4.10. Comparison of genome coverage between the amplification-free method and the PCR-based method

Since the samples for the amplification-free method were sequenced with a shallow read dept (around 5 million molecules), the samples were pooled for the comparison of genome coverage. Similarly the PCR-based sample and control were pooled for this analysis. Chromosome 1 was used as a model chromosome and all genome coverage calculations were done on this chromosome. 6.7 million molecules mapped to chromosome 1 for the PCR-based sample, and therefore 6.7 million molecules were used from the pooled amplification-free samples for a fair comparison.

Acknowledgments

We thank the Staff at Women's Health Dept Obstet Gynecol, Karolinska University Hospital Huddinge for their invaluable help in recruiting parents. We also thank the agencies who funded the research: European Research Council (BRAINCELL, 261063), Swedish Research Council (STARGET), Breast Cancer Theme Center (BRECT) and the KI AstraZeneca Research Collaboration.

Conflict of interest

Kasper Karlsson and Sten Linnarsson have filed a patent application for the amplification free library preparation method.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2014.12.005.

References

[1] L. Hui, D.W. Bianchi, Recent advances in the prenatal interrogation of the human fetal genome, Trends Genet. 29 (2012) 84–91.
[2] M. Allyse, et al., Cell-free fetal DNA testing for fetal aneuploidy and beyond: clinical integration challenges in the US context, Hum. Reprod. 27 (11) (2012) 3123–3131.
[3] A. Bustamante-Aragones, et al., Non-invasive prenatal diagnosis of single-gene disorders from maternal blood, Gene 504 (1) (2012) 144–149.
[4] G. Ashoor, et al., Fetal fraction in maternal plasma cell-free DNA at 11–13 weeks' gestation: relation to maternal and fetal characteristics, Ultrasound Obstet. Gynecol. 31 (2012) 237–243.
[5] J. Dabney, M. Meyer, Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries, Biotechniques 52 (2) (2012) 87–94.
[6] S. Hahn, A.M. Garvin, E. Dinaro, W. Holzgreve, Allele drop-out can occur in alíeles differing by a single nucleotide and is not alleviated by preamplification or minor template increments, Genet. Test. 2 (Number 4) (1998).
[7] I. Kozarewa, et al., Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes, Nat. Methods 6 (4) (2009) 291–295.
[8] M. Meyer, A High-coverage Genome Sequence from an Archaic Denisovan Individual, 2012.
[9] M.T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA, Nat. Protoc. 8 (4) (2013) 737–748.
[10] G.K. Fu, et al., Counting individual DNA molecules by the stochastic attachment of diverse labels, Proc. Natl. Acad. Sci. U. S. A. 108 (22) (2011) 9026–9031.
[11] I. Kinde, et al., Detection and quantification of rare mutations with massively parallel sequencing, Proc. Natl. Acad. Sci. U. S. A. 108 (23) (2011) 9530–9535.
[12] K. Shiroguchi, et al., Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes, Proc. Natl. Acad. Sci. U. S. A. 109 (4) (2012) 1347–1352.
[13] M.W. Schmitt, et al., Detection of ultra-rare mutations by next-generation sequencing, Proc. Natl. Acad. Sci. U. S. A. 109 (36) (2012) 14508–14513.
[14] T. Kivioja, et al., Counting absolute numbers of molecules using unique molecular identifiers, Nat. Methods 9 (1) (2012) 72–74.
[15] H.C. Fan, et al., Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood, Proc. Natl. Acad. Sci. U. S. A. 105 (42) (2008) 16266–16271.
[16] S. Breitbach, et al., Direct quantification of cell-free, circulating DNA from unpurified plasma, PLoS One 9 (3) (2014) e87838.
[17] R.W. Chiu, Y.M. Lo, Noninvasive prenatal diagnosis empowered by high-throughput sequencing, Prenat. Diagn. 32 (4) (2012) 401–406.
[18] K. Mann, C. Donaghue, S.P. Fox, Z. Docherty, C.M. Ogilvie, Strategies for the rapid prenatal diagnosis of chromosome aneuploidy, Eur. J. Hum. Genet. 12 (2004) 907–915.