Research Article

Setting Up of First NIPT Service Laboratory in a Thailand Hospital

Bill CH Chang[1], Takol Chareonsirisuthigul[3], Pareena Janchompoo[3], Chutatip Srichunrusami[3], Ekawat Pasomsub[3], Thipparat Nasrisuk[3], Insee Sensorn[3], Panyu Panburana[2] and Wasun Chantratita[3]*

[1]*Yourgene Bioscience, No. 376-5 Fuxing Rd, Shu-Lin District, New Taipei City, Taiwan*

[2]*Department of Obstetrics and Gynecology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand*

[3]*Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand*

*Corresponding author: Wasun Chantratita, Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, Tel: +6622011470; Fax: +6622011324; E-mail: wasun.cha@mahidol.ac.th

Abstract

The clinical relevant of prenatal genetic screening is rapidly increasing in Thailand. This study aims to set up the very first next generation sequencing (NGS) laboratory for NIPT in a Thai hospital using the Ion Torrent platform. The inter-laboratory validation was carried out with Yourgene Bioscience laboratory. The sensitivities and specificities of the test are over 99.9% for Trisomy-13, Trisomy-18, Trisomy-21, and chromosome X aneuploidies. The positive predictive value of the tests are 83.3%, 91.4%, 95.1%, and 74.2% respectively. Nevertheless, this study demonstrates that a reference database drawn from as few as 100 samples is effective for over 500 new samples.

Keywords: Prenatal genetic screening; Next generation sequencing; Chromosome X; DNA extraction; Aneuploidies calculations; Amniotic fluid

Introduction

In recent years, the development and implementation of non-invasive prenatal test (NIPT) services has attracted a great amount of interest worldwide [1-3]. In 2012, a committee opinion article from the American college of obstetricians and gynecologists (ACOG) concluded that "patients at increased risk of aneuploidy can be offered testing with cell free fetal DNA. This technology can be expected to identify approximately 98% of cases of Down syndrome with false-positive rate of less than 0.5%" [4].

With the growing acceptance of this technology, it is important that laboratories meet the highest quality standards in order to provide reliable services. Previously, most samples from Thailand were sent overseas for testing due to the lack of clinical next generation sequencing (NGS) facility for NIPT. The aims of this study were to set up the very first NGS laboratory for NIPT in a Thai hospital and to offer this service locally.

Materials and Methods

The complete workflow includes plasma DNA extraction, DNA library construction, chip loading, DNA sequencing, and data analysis (Figure 1). Plasma DNA is purified from whole blood using the QIAamp DNA Mini Kit (Qiagen Inc., Cat. No. 51304 or 51306), following the manufacturer's protocol. The minimum required concentration of purified plasma DNA is 0.1 ng/μl. The Ion Plus Fragment Library Kit (Thermo Fisher Scientific Inc., Cat. No. 4471252) with Ion Xpress Barcode Adapters (Thermo Fisher Scientific Inc., Cat. No. 4471250) are used to construct DNA libraries from plasma DNA.
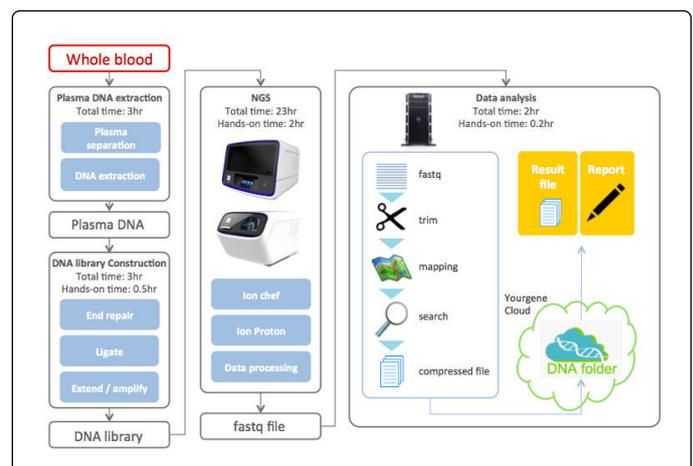

Figure 1: NIPT workflow at center for medical genomics (CMG), Ramathibodi Hospital, Thailand. Plasma DNA was extracted from whole blood drawn from 544 subjects using the QIAamp DNA Mini Kit. End-repair and purify DNA for Ion Torrent 200 bp chemistry were processed before adapter ligation, according to manufacturer protocol. Electrophoresis was performed and the sample fraction around 300 bp was excised using E-Gel SizeSelect Gels. At least 5 cycles of adapter mediated amplification were required to generate quantifiable sequence-ready libraries. The single-end sequencing was performed on the Ion Proton system. Then, all fastq files were aligned to the human reference genome (hg19) and only the sequenced read counts at predefined regions, were uploaded to the Yourgene server (DNA folder) for aneuploidies calculations, specifically chromosomes 13, 18, 21, and X.

The libraries are quantified with Qubit dsDNA HS (high sensitivity) Assay Kit, Ion Library TaqMan qPCR Mix 2X and Ion Library TaqMan Quantification Assay 20X (Thermo Fisher Scientific Inc.). The Ion Torrent Proton and One Touch 2 systems (Thermo Fisher Scientific

Inc.) have been installed for the workflow. The quantitated libraries are loaded onto Ion Torrent P1 chips and then sequenced with the Ion Torrent Proton system. The sequenced reads are aligned to the human reference genome locally and then, with only the sequenced read counts at predefined regions, are uploaded to the Yourgene server (DNA folder) for aneuploidies calculations, specifically chromosomes 13, 18, 21, and X (Yourgene Bioscience CO., LTD.).

In order to validate the test, the first 100 cell-free DNA samples from Yourgene Bioscience were sent to center for medical genomics (CMG) laboratory and the NGS wet lab was performed. The data was analyzed by Yourgene bioscience pipeline. Yourgene Bioscience developed the analysis algorithm used in this study in 2013. The algorithm minimizes noise signals by accounting for regions of low variability in local frequencies (chromosome wide) and regions of high variability in global frequencies (genome wide) in order to reduce false positives and false negatives.

The details of the algorithm are described in the patent application, US20160026759 (Yourgene Bioscience CO., LTD.). Between 2014 and 2015, a total of 13,234 samples were tested, with patients fully informed that data from the screening test would be used for research and development. There were 77 Trisomy-21 samples, 32 Trisomy-18 samples, 10 Trisomy-13 samples, and 23 chromosome X aneuploidy samples (47XXX, 47XXY, 47XYY, and 45X). For all samples with positive NIPT results, amniocenteses were performed and amniotic fluid analyzed with either multiplex ligation-dependent probe amplification (MLPA), karyotyping or array-comparative genomic hybridization (CGH) techniques.

Peripheral blood samples were obtained from 544 pregnant women at their 11th to 25th weeks of gestation who took antenatal care clinic at Ramathibodi hospital during 2015-2016. Before blood sampling, signed consent forms were obtained from all women and the protocol of the study was approved by the ethical clearance committee on human rights related to researches involving human subjects, faculty of medicine Ramathibodi hospital, Mahidol University.

All personal and sequence information is kept in the local computer to ensure that no private information about the samples is transmitted outside the laboratory. A unique sample identification number is assigned to each individual sample and this identification number, in combination with the final result, is used to generate the report. The option to remove gender information from the analysis is also available if needed.

Results and Discussion

A reference set of euploid samples is needed to train and optimize the algorithm to make euploid/aneuploid calls. We conducted a validation study of 100 samples, in which all the samples have been sequenced and analyzed by both CMG laboratory in Thailand and Yourgene laboratory in Taiwan. The sample consists of 92 normal euploid samples, one Trisomy-5, one Trisomy-13, two Trisomy-21, one 45X, one 47XXX, one 47XXY, and one 47XYY samples. The algorithm used for CMG laboratory is trained using only the data from its sequencing facility. The results produced 100% concordant calls for gender, normal euploid, and aneuploid types.

We wanted to know whether this algorithm trained from only 100 samples would provide a good reference set for future incoming samples. To find out, we applied the algorithm to real clinical samples with confirmed karyotypes. The clinical samples included 544 normal euploid samples, two of Trisomy-13, nine Trisomy-18 samples, fourteen Trisomy-21 samples, and two mosaic Turner syndrome samples. The medians of maternal ages, weights and gestational ages for these samples were 38 years old (range 30-44), 60 kilograms (range 50.7-77.3) and 20 weeks (range11-25) respectively.

The results again showed 100% concordance in all euploid/ aneuploid calls. It is interesting to note that the algorithm could correctly call all Trisomy-18 samples, despite the fact that no Trisomy-18 samples were used for algorithm training. For all samples with negative NIPT results, cases were followed up until final delivery, and there were no false negative results reported back by any of the clinicians. The sensitivities and specificities of the test are over 99.9% for Trisomy-13, Trisomy-18, Trisomy-21, and chromosome X aneuploidies. The positive predictive value (PPV) of the tests are 83.3%, 91.4%, 95.1%, and 74.2% respectively.

We have also gathered an additional 1000 samples already sequenced at Yourgene to further test the algorithm. The information used is the read counts obtained from the sequencing runs, so no actual sequence or private information was used for these additional samples.

In order to apply the CMG-based algorithm to Yourgene's sequenced data, we must first compare and contrast the data characteristics of the two sites. The simplest way to do this is to examine the raw frequencies of chromosomes 13, 18, and 21. At CMG, the average frequencies for normal euploid samples are: Chr-13: 0.0321; Chr-18: 0.0267; and Chr-21: 0.0132. Whereas at Yourgene, the average frequencies for normal euploid samples are: Chr-13: 0.0342; Chr-18: 0.0277; and Chr-21: 0.0125. There are clear differences between the read frequencies from CMG and Yourgene for all three chromosomes. A normal Chr-21 frequency value from CMG would likely be interpreted as Trisomy-21 at Yourgene. The contributing factors for these observed differences could arise at various stages of the workflow, including DNA extraction, library construction, template generation, or sequencing.

This difference in data from the two sequencing centers also implies that the algorithm parameters cannot be used interchangeably between the two sites. In fact, if we use the algorithm settings for Yourgene and apply them to the CMG data, the normal euploid samples average z-scores become: Chr-13: -9.05; Chr-18: -5.75; Chr-21: 9.79; Chr-X (male): 22.15; and Chr-X (female): 10.16. Whereas, if we apply CMG settings to the CMG data, the average z-scores for the normal euploid samples are 0 as expected. Therefore, we cannot use the CMG-based algorithm to Yourgene data directly.

However, we could transform the data from Yourgene using an unbiased method, so that the data could be fruitfully analyzed using the CMG-based algorithm. In order to apply this data transformation, we analyzed the samples sequenced at both the CMG and Yourgene laboratories, and calculated their z-scores using the CMG-based algorithm. This result established the correlation between the z-scores obtained from both laboratories.

These correlation parameters were then used to transform the Yourgene data so that it can be analyzed using the CMG-based algorithm, and the results are shown in Table 1.

These results show that, once the data is transformed, for all the normal euploid samples the decision remains as euploid, and for all the aneuploid samples the decision remains as aneuploid. This demonstrates that, even though we have only used 100 CMG samples

to train the algorithm it is already effective in providing accurate classifications.

	Z-Chr13	Z-Chr18	Z-Chr21	Z-ChrX
Euploid mean	0	0	0	0
Euploid max	1.98	2.89	2.99	2.85
Euploid min	-2.84	-3.05	-3.55	-2.84
Aneuploid mean	8.15	13.95	11.35	-2.02
Aneuploid max	14.41	20.78	17	17.67
Aneuploid min	5.03	3.12	5.02	-16.24

Table 1: Maximum, mean and minimum z-scores for normal euploid and aneuploid samples. Z-scores are generated based on CMG-based algorithm, and the samples tested are sequenced from Yourgene with a sample size of 1000.

There has been a general belief that a laboratory with more samples sequenced would produce better and more accurate results [5]. However, there have been no detailed studies on the minimum number of samples required to provide clinical services. We tested the various sample sizes, with different samples used for algorithm training and performance testing, and the results are shown in Table 2. The results show that using just 20 samples for algorithm training reduces the number of false positives to zero for all aneuploidies. Finally, we also tested the number of sequence reads required for a sample to maintain accuracy and showed that the accuracy of the test could be maintained with as little as 1.5 million sequence reads.

Theoretically, it is easier to detect a trisomy from a larger chromosome, since an extra chromosome from a larger chromosome would have a higher increase in its read frequencies. However, there are many factors that would influence the amount of increase that ultimately separate the groups of normal euploid and trisomy samples. One of the most important factors is the GC content as PCR and sequencing process is highly influenced by the GC content of a genome [6].

Training Sample Size	Chr13	Chr18	Chr21	ChrX
False Positive				
3	28	15	0	0
5	4	13	0	0
10	2	0	0	0
20	0	0	0	0
100	0	0	0	0
False Negatives				
3	0	0	0	0
5	0	0	0	0
10	0	0	0	0
20	0	0	0	0
100	0	0	0	0

Table 2: Number of false positives and false negatives for various training sample size. Total population size=1000.

In addition, the GC content correlates with gene coding density in each chromosome rather than chromosome size. Furthermore, there has been reported that longer length of sequencing library could contribute to increase GC content effect [7]. And because of this, the combination of sequencing read in each chromosome needs to be normalized differently before analysis, and hence, the characteristics of sensitivities and specificities in each chromosome are different.

Currently, there are three main approaches with high accuracy for NIPT; whole genome shotgun, targeted sequencing and hybridization microarray. In the following years, the epigenetic-based approach for the identification of methylation fetal-specific markers that could potentially be used for fetal chromosome abnormalities screening has been extensively investigated. It has been suggested that ectopic expression of the proteins associated with centrosome biogenesis may lead dysfunction and instability of centrosome, with the consequence of improper mitotic spindle formation and chromosome mis-segregation [3]. Interestingly, recent publication indicates that epigenetic factor is also involved in this process. In cancer cells, it has been shown that histone H3K9 methylase Ehmt2 is essential for the maintenance of centrosome stability and chromosome segregation [8,9].

As the main function of Ehmt2 is establishing of H3K9me2 and maintenance of imprinted DNA methylation it is very likely that it protects the normal structure of centrosome via set up its relative epigenetic markers [10,11]. Therefore, understanding the levels of specific epigenetic modifications might be a novel direction of the non-invasive prenatal test (NIPT) for the diagnosis of aneuploidies disorders. In 2009, Papageorgiou et al. started working on the MeDIP approach for NIPT and identified methylation markers for identifying trisomies with very high accuracy [12]. The only drawback for this approach is finding the right combination of affinity reagents for the methylated regions. Recently, researchers have used bisulfite treatment approach to study methylation profiles, however, this treatment is associated with high degree of DNA degradation and therefore influence the accuracy of NIPT. With this technical limitations, this approach is therefore currently not available in clinical NIPT services.

For future work, we would investigate this further to optimize the test for economic reasons. We would also investigate the factors contributing to variability in the data both within and between laboratories to help quality control on future tests.

Acknowledgement

References

1. Gil MM, Quezada MS, Revello R, Akolekar R, Nicolaides KH (2015) Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound Obstet Gynecol 45: 249-266.

2. Norton ME, Brar H, Weiss J, Karimi A, Laurent LC, et al. (2012) Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. Am J Obstet Gynecol 207: 137.e1-8.

3. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, et al. (2011) DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 13: 913-920.

4. American College of Obstetricians Gynecologists Committee on Genetics (2012) Committee Opinion No. 545: Noninvasive prenatal testing for fetal aneuploidy. Obstet Gynecol 120: 1532-1534.

5. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, et al. (2012) Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 30: 1033-1036.

6. Kotsopoulou I, Tsoplou P, Mavrommatis K, Kroupis C (2015) Non-invasive prenatal testing (NIPT): limitations on the way to become diagnosis. Diagnosis 2: 141-158.

7. Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res 40: 72.

8. Vitre BD, Cleveland DW (2012) Centrosomes, chromosome instability (CIN) and aneuploidy. Curr Opin Cell Biol 24: 809-815.

9. Kondo Y, Shen L, Ahmed S, Boumber Y, Sekido Y, et al. (2008) Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. PLoS One 3: 2037.

10. Shinkai Y, Tachibana M (2011) H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev 25: 781-788.

11. Zhang T, Termanis A, Ozkan B, Bao XX, Culley J, et al. (2016) G9a/GLP complex maintains imprinted DNA methylation in embryonic stemcells. Cell Rep 15: 77-85.

12. Papageorgiou EA, Fiegler H, Rakyan V, Beck S, , et al. (2009) Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies. Am J Pathol 174: 1609-1618.