Carolina Malcher Amorim de Carvalho Silva

Teste pré-natal não invasivo para detecção de doenças genéticas utilizando sequenciamento de nova geração

Noninvasive prenatal test for detection of genetic diseases using next-generation sequencing

São Paulo

2017

Carolina Malcher Amorim de Carvalho Silva

Teste pré-natal não invasivo para detecção de doenças genéticas utilizando sequenciamento de nova geração

Noninvasive prenatal test for detection of genetic diseases using next-generation sequencing

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutora em Ciências, na Área de Biologia/Genética.

Orientadora: Maria Rita dos Santos e Passos Bueno

São Paulo

2017

Ficha Catalográfica

Silva, Carolina Malcher Amorim de Carvalho.
Teste pré-natal não invasivo para detecção de doenças genéticas utilizando sequenciamento de nova geração / Carolina Malcher Amorim de Carvalho Silva ; orientadora Maria Rita dos Santos Passos Bueno -- São Paulo, 2017.

91 f.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva.

1. Teste pré-natal não invasivo. 2. Sequenciamento de nova geração. 3. Fração fetal. 4. Trissomia. 5. Mutação deponto. I. Passos Bueno, Maria Rita dos Santos, orient. II. Título.

Comissão Julgadora:

_____	_____
Prof(a). Dr(a).	Prof(a). Dr(a).
_____	_____
Prof(a). Dr(a).	Prof(a). Dr(a).

_____

Prof[a]. Dr[a]. Maria Rita dos Santos e Passos Bueno

Orientadora

Dedicatória

À minha família

Agradecimentos

Gostaria de agradecer primeiramente à Rita, por ter me acolhido desde 2011 e me ensinado muito sobre ciência. Por confiar em mim sempre, e apoiar minhas idéias e escolhas.

A todos do Lab200, de quem vou sentir muita saudade. Espero encontrar na frente locais com um ambiente de trabalho tão gostoso e colaborativo como esse. Obrigada a todos que me apoiaram e me divertiram nesses anos todos. Aprendi muito com cada um. <3

Ao Robs, Luciano, Luquinhas e Gerson pelos cafés, amizade e ajuda com a correção da tese. Em especial às meninas: Ágatha, Clarice, Duda, Camila e Priscila. A vida teria sido bem mais chata sem vocês e nossas noites de vinho.

De forma geral pelo pessoal do Genoma, pela ajuda e incentivo desde o início do projeto. Em especial ao Docinho, pois sem ele o projeto teria sido bem mais sofrido.

Ao Iwijn e sua equipe, que me acolherem e me ensinaram muito na Cornell University, onde os 2 meses que passei foram muito enriquecedores.

À toda minha família, que sempre me apoiou, ao longo dessa trajetória incrível de pós-graduação, com a qual aprendi tanto. À minha vó Maysa, que deixou tanta saudade.

Em especial ao meu (finalmente!) marido, Léo, que ao longo desses anos sempre esteve ao meu lado me incentivando, apoiando e divertindo.

À todos os colaboradores e todas as pacientes participantes da pesquisa.

Este trabalho contou com o apoio financeiro da FAPESP, do CNPq e do Ministério da Ciência e Tecnologia do Brasil.

Notas

Esta tese de doutorado compreende um trabalho desenvolvido durante os anos de 2013 a 2017 no Laboratório de Genética do Desenvolvimento do Centro de Pesquisas sobre o Genoma Humano e Células-Tronco (CEGH-CEL/IB-USP), sob orientação da prof. Dra. Maria Rita dos Santos e Passos Bueno.

O projeto que resultou na presente tese foi cadastrado na Plataforma Brasil e contou com o parecer consubstanciado do Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo, número 481.663/2013. Todos os participantes da pesquisa assinaram os Termo de Consentimento Livre e Esclarecido (TCLE) (Anexo 1), enquanto todas as participantes grávidas também assinaram o TCLE presente no Anexo 2.

A tese foi redigida no modelo de artigos e capítulos, no idioma inglês. Trabalhos em em co-autoria e não relacionados ao tema principal da tese encontram-se em Apêndices, ao final da tese.

List of Abbreviations

AF	Allele fraction	**NGS**	Next-generation sequencing
cfDNA	Cell-free DNA	**NIPT**	Noninvasive prenatal test
cfdDNA	Donor cell-free DNA	**OI**	Osteogenesis imperfecta
cffDNA	Fetal cell-free DNA	**SD**	Skeletal dysplasia
ctDNA	Circulating tumor DNA	**SNP**	Single nucleotide polymorphism
DH	Diagnosis hypothesis	**TD**	Thanatophoric Dysplasia
FF	Fetal fraction	**T18**	Trisomy of chromosome 18
GTD	Genomic transplant dynamics	**T21**	Trisomy of chromosome 21
HUG-CELL	Human Genome and Stem Cell Research Center	**VUS**	Variant of unknown significance
MAF	Minor allele fraction	**WGS**	Whole genome sequencing

Table of Contents

Chapter 1

General Introduction

Cell-free DNA and its applications

Cell-free DNA (cfDNA) is naturally found fragmented within the blood at a very low concentration (1000-10000 genomes/ml plasma), originated from apoptosis and/or tissue necrosis. Its half-life is approximately 16 minutes, which makes it a very interesting potential biomarker for dynamic monitoring of diseases associated with "exogenous" DNA (i.e., derived from other sources, such as fetus, cancer or transplanted organ) (LO et al., 1998b, 1997, 1999; STROUN et al., 1989; JAHR et al., 2001; SUZUKI et al., 2008).

Presence of "exogenous" cfDNA was first described for cancer (LEON et al., 1977; STROUN et al., 1989), and later described for fetal DNA (LO et al., 1997), and for transplanted organs (LO et al., 1998a). Use of cfDNA for analysis of non-solid biological tissue is called liquid biopsy, and it has been increasingly used as a diagnostic and monitoring tool for diseases, having the advantage of being non-invasive, as described below.

- ***Cancer Detection***

In general, cancer possesses a strong genetic component (STRATTON; CAMPBELL; FUTREAL, 2009; GARRAWAY; LANDER, 2013; TOMASETTI; LI; VOGELSTEIN, 2017; VOGELSTEIN et al., 2013). The gold standard for cancer monitoring is biopsy analysis; however, other than being invasive, expensive and not

always available, it does not contemplate inter and intratumoral heterogeneity, and it is specific to the moment of punction, which might take too long to translate into clinical action.

The discovery of circulating tumor DNA (ctDNA) (LEON et al., 1977; STROUN et al., 1989) opened possibilities to perform real-time disease monitoring through the diverse disease stages, and can be used not only for early detection of the disease, but also for prognosis and disease treatment, such as predicting response to treatments (CROWLEY et al., 2013; HUANG et al., 2012; ROTHÉ et al., 2014). It also addresses the limitations of biopsy when it comes to assessing inter and intratumoral diversity because ctDNA fragments are derived from tumors across the organism. The ctDNA fraction varies between 0.01 and 93%, and it is correlated with tumor burden, early relapse and treatment response (CROWLEY et al., 2013; DIEHL et al., 2005; JAHR et al., 2001; THIERRY et al., 2010) (Figure 1).

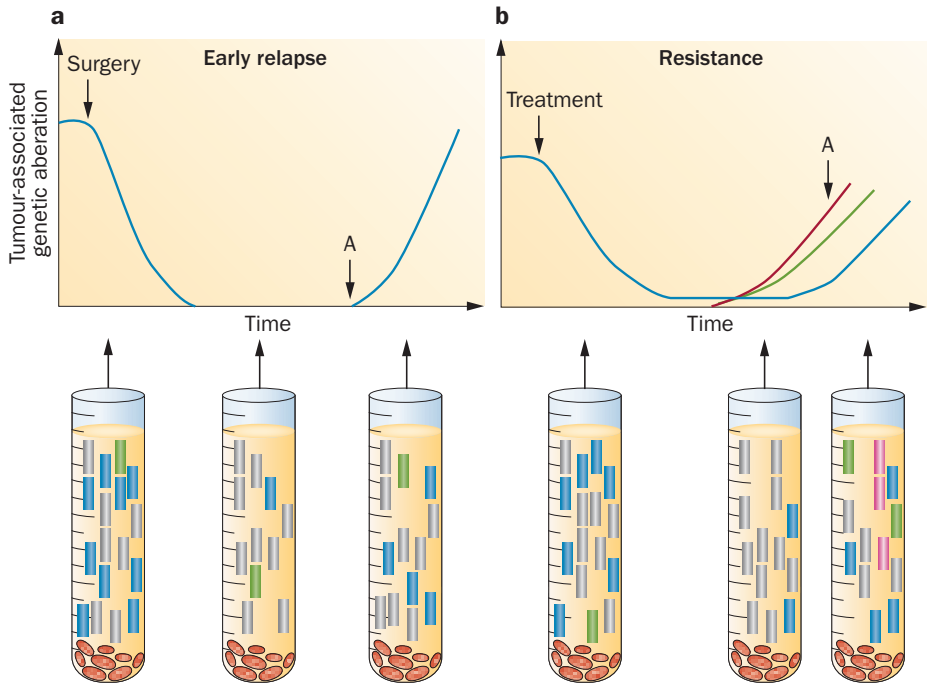

Figure 1: Monitoring tumor-specific aberrations to detect A) relapse after surgery; B) resistance to treatment. 'A' represents the clinically detectable time of relapse. Blue represents an early occurring mutation present in the tumor and reflects overall tumor burden. Red and green represent mutations associated with development of the tumor, as resistant clones. Adapted from CROWLEY *et al.*, 2013.

Detection of tumor-specific mutations in ctDNA helps performing a real-time diagnosis and monitoring of the disease. Although mutation profile varies among patients, cancer types and disease stages (HAN; WANG; SUN, 2017; HUANG et al., 2012; SCHWARZENBACH; HOON; PANTEL, 2011), there are some genes in which mutations are already well characterized for some types of cancer, for instance, *BRAF* in melanomas and *EGFR* for several types of cancer, such as colorectal, breast, thyroid and lung (DAVIES et al., 2002; DOUILLARD et al., 2014; HAN; WANG; SUN, 2017; KRISHNAMURTHY et al., 2017; UCHIDA et al., 2015). The advent of next-generation sequencing (NGS) has permitted simultaneous investigation of such genes of interest and the identification of rare mutations in complex DNA mixtures, due to increased sensitivity of this method (FORSHEW et al., 2012; KINDE et al., 2011; LEARY et al., 2010). Although it does not cover all possible scenarios, tracking mutations in such genes can be used as a screening strategy for early detection of certain types of cancer. Another use for this approach is to provide clues to the etiology of the disease when there is a cancerous mass present, but direct biopsy is too complicated due to technical issues or patient frailty.

It has been show that liquid and tumor biopsies have a concordance of >70%, and liquid biopsies have the additional benefit of detecting genetic information of tumors from metastatic sites of the body, making it a great potential biomarker (DOUILLARD et al., 2014; KRISHNAMURTHY et al., 2017; PATEL; TSUI, 2015; ROTHÉ et al., 2014). Although validation is still required for most cases of cancer, liquid biopsy has already been used for some cancers, such as lung cancer, which recently had the first FDA-approved liquid biopsy (cobas EGFR Mutation Test v2 – Roche Molecular Systems, Inc.). It relies on screening for *EGFR* (epidermal growth factor receptor) mutations to indicate therapeutic decisions regarding the use of EGFR tirosyne kinase inhibitors (TKIs) (KRISHNAMURTHY et al., 2017; UCHIDA et al., 2015). This is particularly important for lung cancer due to the difficulty in obtaining biopsies for some cases.

- ***Transplanted organ follow up (Allograft Rejection Detection)***

Incidence of acute allograft rejection varies depending on the organ type, being as high as 55% in intestinal transplant (BECK et al., 2015; COLVIN-ADAMS et al., 2015; GIELIS et al., 2015; VALAPOUR et al., 2015). The gold standard for transplant rejection diagnosis is allograft biopsy, an invasive procedure that has pitfalls, such as: variability of grading scale due to different observers, and, for small biopsies, chance of missing the acute rejection due to the inhomogeneous nature of graft damage (ARCASOY et al., 2011; MARBOE et al., 2005; SARAIVA et al., 2011). Therefore, a noninvasive method has been pursued to eliminate discomfort and to overcome these pitfalls.

The discovery of donor cfDNA (cfdDNA) in the receptor's blood (LO et al., 1998a) was an important step for the development of a noninvasive method for diagnosis of transplant rejection, since during rejection there is an increase of organ cell death, which will translate into an increased cfdDNA in the receptor's blood. This method basically consists on monitoring the levels of cfdDNA as a rejection signature (BLOOM et al., 2017; DE VLAMINCK et al., 2014; SCHÜT et al., 2017; SIGDEL et al., 2013; SNYDER et al., 2011; ZHANG et al., 1999). This monitoring can be performed with the use of Y-chromosome sequences, however it works only in 25% of cases, in which the receptor is a woman and the donor is a man. More recently, a universal method (regardless of gender) called GTD (Genomic Transplant Dynamics) was described, and it uses single nucleotide polymorphisms (SNPs) to quantify cfdDNA and therefore detect the onset of rejection (BECK et al., 2015; BLOOM et al., 2017; DE VLAMINCK et al., 2014; GIELIS et al., 2015; SNYDER et al., 2011) (Figure 2).

Although there are other methods for SNP detection (such as digital droplet PCR), the use of NGS (targeting a panel of informative SNPs) is advantageous because it has been shown to detect not only organ rejection, but also transplant-related infections simultaneously, through detection of non-human sequence reads used for identification of infectious agents (BECK et al., 2013; DE VLAMINCK et al., 2014, 2015; HIDESTRAND et al., 2014).

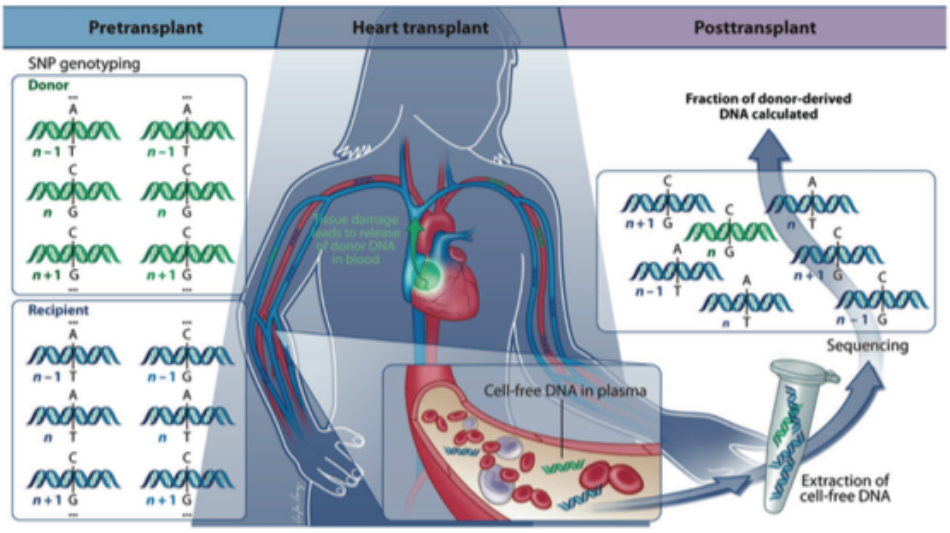

Figure 2: GTD principle: genotyping recipient and donor before the transplant. Sequencing of plasma performed after transplant to detect single bases alleles that were distinct between donor and recipient, allowing discrimination of donor and recipient derived sequences (positions n, but not n-1 and n+1). Adapted from DE VLAMINCK *et al.*, 2014.

The GTD method has shown that the basal cfdDNA fraction is variable and organ-specific, and rejection detection can be achieved months before diagnosis by biopsy (up to 5 months in heart transplant), therefore preventing graft damage (BECK et al., 2013; DE VLAMINCK et al., 2014; SCHÜT et al., 2017). An improvement of the GTD method has been recently accomplished, obviating the need of genotyping the donor or the receptor: it uses a panel of informative SNPS, taking into account different degrees of relatedness between recipient and donor (GRSKOVIC et al., 2016).

Recently, the first commercially available allograft rejection monitoring using cdDNA and NGS was launched in the US (AlloSure – CareDx, Inc.) (WOODWARD et al., 2015).

- ***Prenatal Genetic Disease Diagnosis***

Chromosomal anomalies are responsible for 50% of spontaneous abortions, about 6% of congenital anomalies and 5.6% to 11.5% of perinatal deaths (MOORE;

PERSAUD, 1998; RAI; REGAN, 2006; SANSEVERINO, 2006). Aneuploidies constitute the most common chromosomal anomalies, and the most frequent are autosomal aneuploidies of chromosomes 21 (Down Syndrome), 18 (Edwards Syndrome) and 13 (Patau Syndrome) (HYDE; SCHUST, 2015; NUSSBAUM, 2007; SKRZYPEK; HUI, 2017). Therefore, screening for these aneuplodies is routinely offered in prenatal care (BENN et al., 2013).

Conventional prenatal screening methods for aneuploidy and other birth defects include maternal serum screening and ultrasound imaging in the first and/or second trimester. However, these approaches have their own limitations and suffer from high false positive rates (5% on average), leading to unnecessary invasive diagnostic procedures (BIANCHI et al., 2014; RUSSO; BLAKEMORE, 2014; SHAMSHIRSAZ; BENN PETER; EGAN, 2010; SKRZYPEK; HUI, 2017). If the screening tests indicate that a fetus has an increased risk of aneuploidy, the indication is to proceed with invasive tests, like chorionic villus sampling (CVS) or amniocentesis for diagnosis, offered around 10-13 and after 15 weeks of gestational weeks, respectively. Invasive tests entail, besides discomfort for the mother, risks for the pregnancy as miscarriage and fetal abnormalities (AKOLEKAR et al., 2015; CAUGHEY; HOPKINS; NORTON, 2006; MUJEZINOVIC; ALFIREVIC, 2007).

Discovery of cell-free fetal DNA (cffDNA) led to significant advances in prenatal diagnosis (LO et al., 1997). This DNA of fetal origin ranges between 133 and 166bp in size. It is released into the maternal bloodstream through apoptosis of trophoblastic cells (which form the placenta), being rapidly eliminated after birth, which makes it a promising pregnancy-specific biomarker (BISCHOFF; LEWIS; SIMPSON, 2005; FLORI et al., 2004; LICHTENSTEIN et al., 2006; LO et al., 1999, 2010; SMID et al., 2003). Maternal cfDNA is derived mainly from apoptosis and active secretion of cellular DNA (GAHAN; ANKER; STROUN, 2008; VAN DER VAART; PRETORIUS, 2008).

Since its discovery, cffDNA has been widely used for detection of qualitative traits not present in the mother, such as fetal sex determination (chromosome Y detection), Rh factor (only possible for Rh- mothers), and the presence of *de novo* or paternally inherited monogenic disease mutations (BUSTAMANTE-ARAGONÉS et al., 2012; CHITTY et al., 2011; FAAS et al., 1998; LO et al., 1997, 1998c). With the

advent of next-generation sequencing, it has been possible to detect quantitative traits, such as aneuploidies (CHIU et al., 2008; FAN et al., 2008).

Noninvasive prenatal testing (NIPT) using next-generation sequencing

- ***Aneuploidy detection***

Fan and colleagues (2008) demonstrated for the first time the utility of NGS to reliably detect fetal quantitative traits through read count. After several efforts in 2011, noninvasive prenatal testing (NIPT) became clinically available in the United States and China, changing the prenatal care field (CHIU et al., 2010; EHRICH et al., 2011; FAN et al., 2008; FAN; QUAKE, 2010; LO et al., 2010; PALOMAKI et al., 2011; SEHNERT et al., 2011; SPARKS et al., 2012a, 2012b; ZIMMERMANN et al., 2012).

NIPT tests have high sensitivity (true positive rate) and specificity (true negative rate) values, especially for chromosome 21 (GIL et al., 2015; KOUMBARIS et al., 2016). Bianchi and colleagues (2014) demonstrated that, for detection of trisomy of chromosome 21 (T21) and chromosome 18 (T18) in general obstetric population, NIPT shows lower false positive rates than standard screening (false positive rates: 0.3% vs. 3.6% for T21, $P<0.001$, and 0.2% vs. 0.6% for T18, $P=0.03$) (BIANCHI et al., 2014).

Two approaches have been developed and implemented for NIPT commercially: SNP-based NIPT and read count-based NIPT (NGS and targeted NGS).

- *SNP-based NIPT*

This approach selectively sequences SNPs with high heterozygosity level and compares cfDNA SNPs with parental genotypes. Thousands of SNPs are used for each chromosome of interest, and only informative *loci* are used to determine aneuploidy. The aneuploidy risk is calculated through SNP distribution, taking into

account different hypotheses of the fetus being monosomic, disomic or trisomic. It then compares the expected and observed allele distributions to determine the most likely scenario (LIAO *et al.*, 2011a; ZIMMERMANN *et al.*, 2012; RYAN *et al.*, 2016).

- *Read count-based NIPT*

Next-generation sequencing can be performed as whole genome sequencing (WGS), or as targeted sequencing, which sequences only some parts of the genome (exome and/or a panel of genes of interest).

For trisomy detection, both approaches rely on detecting an increase of the read count originated from the trisomic chromosome. Therefore, women carrying an trisomic fetus will have a higher proportion of read count on maternal plasma DNA (pool of maternal and fetal DNA) than women carrying euploid fetus, and this increase will be proportional to the cffDNA fraction (fetal fraction - FF) present in maternal blood, as shown at Figure 3.

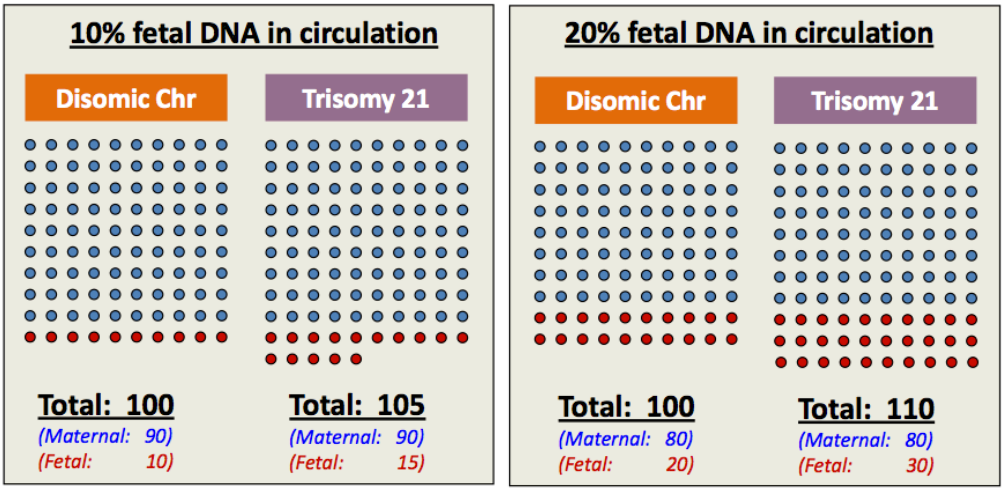

Figure 3: Representation of chromosomal increase in presence of a trisomic fetus for chromosome 21, proportional to fetal fraction. Blue: maternal cfDNA; Red: fetal cfDNA. Adapted from SPARKS; STRUBLE; et al., 2012.

As mentioned above, targeted sequencing is performed for only part of the genome, and the coverage required for detection of trisomies of chromosomes 21, 18 and 13 for targeted sequencing is less than 5% of the coverage required for WGS. Therefore, targeted sequencing is a great strategy to reduce costs and increase coverage, and it was applied for NIPT by Sparks and colleagues (2012) for the first time (ASHOOR et al., 2012a; KOUMBARIS et al., 2016; LIAO et al., 2011; NORTON et al., 2012; SPARKS et al., 2012a, 2012b).

NGS and targeted NGS uses amplification of maternal plasma cfDNA (mix of maternal and fetal cfDNA), and fetal chromosome copy number is determined by comparing the ratio of reads from the chromosome of interest against mean and standard deviation of ratio of a reference disomic dataset, expressed as Z-Score or NCV (normalized chromosome value) (CHIU et al., 2011; SEHNERT et al., 2011; SPARKS et al., 2012a). Z-score is regularly used in the literature and a Z-score higher than 3 is the threshold used to detect trisomy (meaning the result is 3 standard deviations above the mean of the reference dataset, or Z-score < 3 means the result is 99% within the normal expected z-score, i.e. unaffected) (CHIU et al., 2008). The Z-score increases with FF, since samples with higher FF have more fetal cfDNA to be detected (EHRICH et al., 2011; PALOMAKI et al., 2011; SPARKS et al., 2012b). Therefore, estimating FF is a great quality control for NIPT, although not all tests contain this measurement.

Although already offered for patients in Brazil, NIPT technology has not been implemented in our country. The laboratories that commercialize the test outsource the technology or the test itself. Besides decreasing the amount of sequencing required, another advantage of targeted NGS is the ability to detect SNPs, which allow estimation of fetal fraction and detection of monogenic diseases (although all these tests are not performed altogether thus far). Therefore, targeted sequencing is a great option for a clinical scenario.

- ***Monogenic disease detection by NIPT***

Most NIPT initially focused on detection of autosomal dominant traits, which are present only in the fetus and are caused by *de novo* or paternally inherited mutations, such as fetal sex and RH status (BUSTAMANTE-ARAGONES et al., 2008; CHITTY et al., 2011; FAAS et al., 1998; GONZÁLEZ-GONZÁLEZ et al., 2002; LO et al., 1997, 1998c). Identification of fetal sex is useful for instance in cases where couples are at risk of having a child affected by an X-linked disease, as well as conditions where there is possibility of clinical intervention during pregnancy, such avoiding virilisation of affected females in CAH (congenital adrenal hyperplasia) cases by administration of maternal dexamethasone early in pregnancy (RIJNDERS et al., 2001).

Detection of *de novo* or paternally inherited gene disorders is already performed for some diseases, such as skeletal dysplasia (SD) (CHITTY et al., 2011, 2015; DAN et al., 2016; VERHOEF et al., 2016), where detection of the pathogenic mutations confirms fetal diagnostic. On the other hand, mutation detection when the mother carries (or may carry) the mutation, such as autosomal recessive or X-linked diseases is more challenging. For these cases, approaches such as detection of relative haplotypes or mutation dosages have been developed (LAM et al., 2012; LUN et al., 2008b; MA et al., 2014; PARKS et al., 2016). Although most tests for monogenic diseases focus on specific diseases, the feasibility of constructing a fetal genome map using maternal plasma deep sequencing and parental genotype information has already been demonstrated as a proof of concept (FAN et al., 2012; KITZMAN et al., 2012; LO et al., 2010).

There are about 5,000 known monogenic disorders (AMBERGER et al., 2015), and only a few of those are currently being detected through NIPT in clinical settings (BUSTAMANTE-ARAGONÉS et al., 2012; LENCH et al., 2013; SKRZYPEK; HUI, 2017; VERHOEF et al., 2016). Even for those, at least maternal genotype information is required in addition genotype information from plasma sequencing, increasing price and turnaround time for test results. The ideal clinical scenario would be a NIPT relying only on maternal plasma sequencing information.

- ***Cell-free fetal DNA fraction (Fetal fraction)***

Fetal fraction estimation is important for detection of monogenic diseases and aneuploidies, since FF is proportional to the allele fraction (AF – proportion of the mutation in respect with the reference allele) for monogenic diseases, as is proportional with the increase of the reads from the trisomic chromosome for aneuploidy detection (as shown above in Figure 3).

FF increases through pregnancy, corresponding to about 10% of total cfDNA in maternal blood at about the 10^{th} gestational week, varying between patients (LUN et al., 2008a; WANG et al., 2013). There are several factors associated with fetal FF, such as maternal weight (inversely proportional) and maternal ethnicity; however, these factors have not been assessed in the Brazilian population (ASHOOR et al., 2012b; KINNINGS et al., 2015; LO et al., 1998b; POON et al., 2013; ZHOU et al., 2015). FF is an important factor that influences the accuracy of NIPT. The currently accepted FF threshold for NIPT is 4%, and lower FFs may yield false negative results for the test (EHRICH et al., 2011; PALOMAKI et al., 2011; SPARKS et al., 2012a, 2012b; TAKOUDES; HAMAR, 2015). There are several ways of measuring FF, however, most of them are only applicable to pregnancies of male fetuses (real-time PCR or the ratio between the Y chromosome and autosomes' reads) (FAN et al., 2008; LO et al., 1998b) or are performed with the analysis of fetal-specific alleles, requiring genotype information of the mother, at least (CHU et al., 2010; LO et al., 2010).

As explained above, while performing noninvasive prenatal testing using NGS, some groups use targeted sequencing and estimate the FF from the retrieved SNP information, while groups performing low-coverage whole-genome sequencing are not able to retrieve SNP information (LAU et al., 2014; SPARKS et al., 2012b; ZIMMERMANN et al., 2012). For FF estimation without the use of parental genotypes, a maximum likelihood estimation using the binomial distribution is employed to determine the most likely FF based upon measurements from several informative SNP *loci* (JIANG et al., 2012; SPARKS et al., 2012b).

For the clinical scenario, the ideal approach is to be able to estimate FF for any pregnancy, regardless of fetal gender, and independently of the genotype

information of either parent. Any additional laboratory step increases the turnaround time and cost of the exam. It is important to note that this approach can be extrapolated to other fields that use cfDNA as a diagnostic tool, such as cancer detection and transplant rejection (DE VLAMINCK et al., 2014; ROTHÉ et al., 2014).

To our knowledge, there is no NIPT test integrating all analyses (fetal fraction, fetal sex, trisomies and monogenic disease), which would be the ideal clinical scenario, because it diminishes cost and turnaround time, as already discussed above. At the present time we do not have a NIPT test developed in Brazil, since the tests offered in our country outsource the technology or the test itself. Development of such technique is important for the scientific development of our country, as well as for the development of novel tests for the population.

Objectives

Our main objective is to develop an in-house noninvasive prenatal test to detect genetic disorders with the use of next-generation sequencing. We aim to integrate several analyses within one single test, thus reducing laboratory steps.

In this respect, our objective can be divided as follows:

a) To estimate fetal fraction in maternal plasma;
b) To assess correlation between fetal fraction in maternal plasma with gestational age and maternal weight in the Brazilian population;
c) To detect trisomy of chromosomes 21 and 18;
d) To detect monogenic diseases, using skeletal dysplasia as a model;
e) To determine the test sensitivity and specificity.

Chapter 2

Implementation of a comprehensive Brazilian in-house noninvasive prenatal test

Carolina Malcher[1], Guilherme L. Yamamoto[1], Philip Burnham[2], Suzana A. M. Ezquina[1], Naila C. V. Lourenço[1], Sahilla Balkassmi[4]; David S. Marco Antonio[1], Gabriella S. P. Hsia[1], Thomaz Gollop[3], Rita C. Pavanello[1]; Egbert Bakker[4]; Mayana Zatz[1], Débora Bertola[1], Iwijn De Vlaminck[2], Maria Rita Passos-Bueno[1]

[1] Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil

[2] Meinig School of Biomedical Engineering, Cornell University, Ithaca, New York, United States

[3] Faculdade de Medicina de Jundiaí, Jundiaí, São Paulo, Brasil

[4] Human and Clinical Genetics Department, Leiden University Medical Center, Leiden, The Netherlands

ABSTRACT

Objective: To develop and apply a comprehensive noninvasive prenatal test (NIPT) using high-coverage targeted next-generation sequencing to estimate fetal fraction, determine fetal sex, and detect trisomy and monogenic disease, without parental genotype information.

Method: Analysis of 45 pregnancy samples and of 8 mother-child pairs to generate 35 simulated datasets. Fetal fraction (FF) was estimated based on analysis of SNP allele fraction distribution; Z-score calculated for T21, T18 and fetal sex detection; Monogenic disease detection was performed through variant analysis. Model validation was performed using the simulated datasets.

Results: The novel model to estimate fetal fraction is robust and accurate (r^2= 0.994, p-value < 2.2e-16). For samples with FF > 0.04, T21 detection has a sensitivity of 100% (95% CI: 63.06% to 100.00%) and a specificity of 98.53% (95% CI: 92.08% to 99.96%), and T18 detection has sensitivity and specificity of 40% (95% CI: 5.27% to 85.34%) and 98.59% (95% CI: 92.40% to 99.96%), respectively. A blind dataset was used for validation, and 5/5 T21 samples were detected, while for T18 (11Mb partial trisomy) was not possible to detect it. Fetal sex determination was performed with 100% accuracy. We furthermore performed a proof of concept for monogenic disease diagnosis of 5/7 cases of skeletal dysplasia.

Conclusion: It is feasible to perform a comprehensive NIPT using only data from high coverage targeted sequencing, which in addition to detecting trisomies, allow the identification of any pathogenic variant of the candidate genes for monogenic diseases.

RESUMO

Objetivo: Desenvolver e aplicar um teste pré-natal não-invasivo (NIPT) abrangente utilizando sequenciamento de nova geração *targeted* de alta cobertura para estimar fração fetal, determinar sexo fetal, e detectar trissomia e doença monogênica, sem utilizar genótipo parental.

Método: Análise de 45 amostras de grávidas e 8 pares de mãe-filho para gerar 35 datasets simulados. Estimativa de fração fetal (FF) baseada na análise de distribuição de frações alélicas de SNPs; Z-score calculado para detecção de T21, T18 e sexo fetal; Detecção de doença monogênica foi realizada através de análise de variante. Validação do modelo realizada utilizando datasets simulados.

Resultados: O novo modelo para estimar fração fetal é robusto e preciso (r^2= 0.994, p-value < 2.2e-16). Para amostras com FF > 0.04, detecção de T21 tem um sensibilidade de 100% (95% IC: 63.06% a 100.00%) e especificidade de 98.53% (95% IC: 92.08% a 99.96%), e detecção de T18 possui uma sensibilidade e especificidade de 40% (95% IC: 5.27% a 85.34%) e 98.59% (95% IC: 92.40% a 99.96%), respectivamente. Um dataset foi utilizado para validação às cegas, e 5/5 amostras de T21 foram detectadas, enquanto a amostra de T18 (trissomia parcial de 11Mb) não foi possível. Determinação de sexo fetal foi realizada com 100% de precisão. Foi também realizada uma prova de conceito para diagnóstico de doença monogênica em 5/7 casos de Displasia Esquelética.

Conclusão: É possível executar um NIPT abrangente utilizando apenas dados de sequenciamento de alta cobertura *targeted*, o qual além de detectar trissomia, permite a identificação de qualquer variante patogênica de genes candidatos de doenças monogênicas.

INTRODUCTION

Discovery of fetal cell-free DNA (cffDNA) in the maternal bloodstream (LO et al., 1997) has revolutionized prenatal diagnosis. Initially, cell-free (cfDNA) was used for detection of qualitative traits, such as fetal sex (LO et al., 1998; RIJNDERS et al., 2001; WRIGHT et al., 2012) and Rhesus D status (FAAS et al., 1998; FINNING et al., 2002). More recently, next-generation sequencing (NGS) technologies have provided a means for noninvasive detection of fetal aneuploidy with high sensitivity and specificity (FAN *et al.*, 2008; EHRICH *et al.*, 2011; PALOMAKI *et al.*, 2011; SPARKS *et al.*, 2012; GIL *et al.*, 2015; NORTON *et al.*, 2015). Noninvasive prenatal testing (NIPT) using NGS of cfDNA is now being widely used as a screening test for the most common aneuploidies in prenatal setting. Chromosome Y read count has been used for accurately fetal sex determination (CHIU et al., 2011; KOUMBARIS et al., 2016) and fetal fraction (FF) estimation (restricted for male fetuses only) (FAN et al., 2008; HUDECOVA et al., 2014; XU et al., 2016).

High coverage targeted sequencing allows accurate detection of fetal alleles without requiring parental genotyping (LIAO et al., 2011). In addition to aneuploidy detection, this strategy enabled identification of variants associated with monogenic diseases, especially *de novo* variants (CHITTY et al., 2015; LAM et al., 2012; NEW et al., 2014). This method also enabled development of methods for FF estimation using SNPs from sequencing analysis of maternal plasma cfDNA, avoiding the need of parental genotyping and reducing laboratory steps and turnaround time (JIANG *et al.*, 2012; SPARKS *et al.*, 2012; KOUMBARIS *et al.*, 2016). FF estimation is crucial for test accuracy, since insufficient fetal cfDNA may lead to false negative results. Thus, measuring the presence of fetal DNA (independently of fetal sex) in maternal plasma in any test (e.g. trisomy detection) should improve its reliability. The aforementioned analyses are already performed in a clinical setting, however, not within one single test. The development of parameters to perform all these analyses simultaneously using only maternal plasma sequencing data may further reduce cost and turnaround time.

NIPT testing in Brazil is currently offered by private laboratories and is

performed by outsourcing the technology or the test itself. In the present report, we propose the implementation of an in-house NIPT using high-coverage targeted NGS in order to estimate FF, determine fetal sex, detect trisomy and monogenic disease, without the need of parental genotypes. We used skeletal dysplasia (SD) as a monogenic disease model.

METHODS

- ***Subjects and Samples***

Peripheral blood samples were collected from pregnant women and from non-pregnant individuals (mother and child), these last ones to establish a proof of concept of the test. Pregnant women were at least 18 years old, singleton pregnancies and from 10 to 36 gestational weeks. This study was approved by the Research Ethics Committee of Instituto de Biociências (Universidade de São Paulo - Brazil) and informed consent was obtained from the all the patients or legal tutors.

The blood samples were collected in EDTA tubes and plasma processing took place within 6 hours. Blood samples were centrifuged at 1600g for 10 minutes and recentrifuged at 16000g for 10 minutes. Plasma cfDNA extraction of 2-4 ml was performed using the QIAamp Circulating Nucleic Acid kit (Qiagen), following the manufacturer's protocol. cfDNA was first eluted in a total of 150ul and then concentrated to 60ul using SpeedVac.

- ***Proof of concept – Mock Samples***

To establish a proof of concept of the test and validate the bioinformatics pipeline, we generated *in silico* mock samples of pregnancies by mixing the fastq reads from both mothers and children. We mixed the fastq reads with different fractions in order to simulate different "fetal fractions" for each pair, mimicking the

progressive increase in FF during pregnancy from the first to the third trimester. Among these samples, there are pairs with children affected and not affected by Down syndrome.

- ***High Coverage Next-generation Targeted Sequencing of Plasma Samples***

For cfDNA library preparation we used the NEBNext Ultra kit (New England Biolabs) according to the manufacturer's protocol. Libraries were indexed, multiplexed, captured for a gene panel using Nextera Rapid Capture (Illumina) and quantified by real-time quantitative PCR using KAPA Library Quantification kit (KAPA Biosystems). Libraries were then sequenced in MiSeq (Illumina) using MiSeq Reagent kit v3 (2x75 cycles) and also in HiSeq (Illumina) using HiSeq Rapid SBS Kit v2 (2x100 cycles).

We aligned the fastq files using BWA-MEM (LI; DURBIN, 2010), removed duplicated reads using Picard (http://broadinstitute.github.io/picard), realigned based on known local indels using GATK (MCKENNA *et al.*, 2010; DEPRISTO *et al.*, 2011; VAN DER AUWERA *et al.*, 2013), and also removed reads with more than two mismatches using Samtools (LI et al., 2009). We determined the mean coverage of Bam files using Samtools Depth. For FF estimation we performed variant call with all patients using GATK. For the aneuploidy detection, we generated a Depth of Coverage file for each sample using GATK. For monogenic disease detection we called the somatic variants using Mutect (CIBULSKIS et al., 2013). Workflow is outlined on Figure 4.

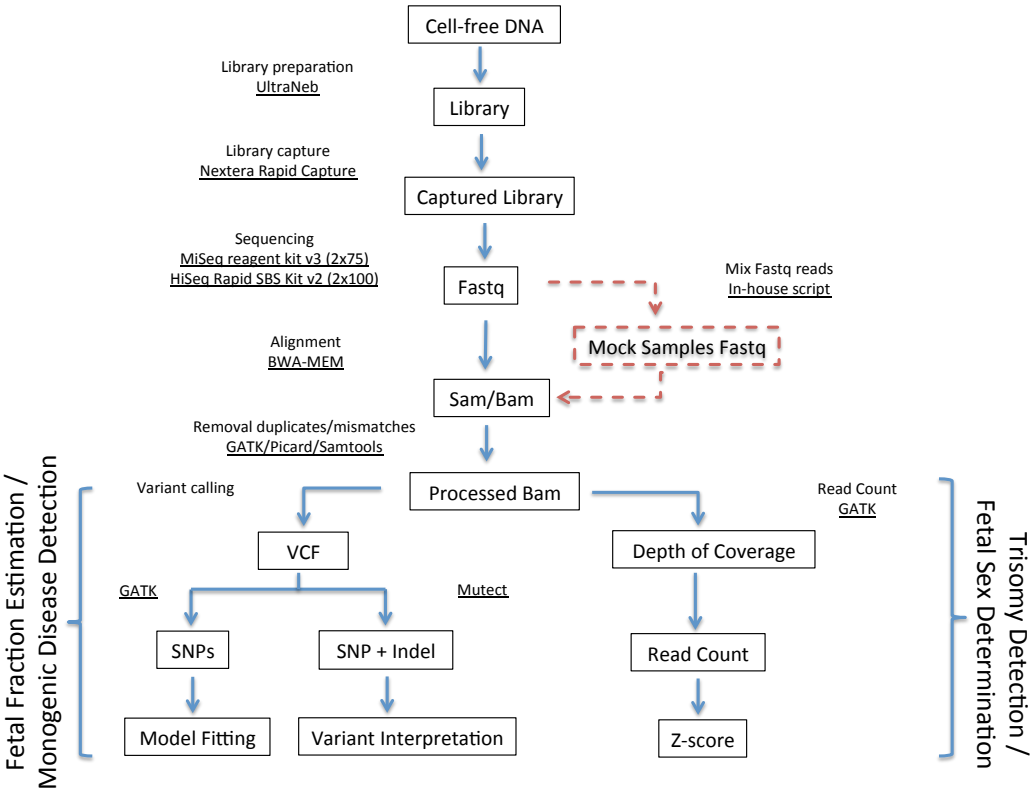

Figure 4: Workflow of the test. Red: performed only in the mother and child fastq files to generate mock samples.

- ***Gene Panel***

For the targeted sequencing we used a panel of genes of clinical interest, used in the routine diagnosis of the HUG-CELL (Human Genome and Stem Cell Research Center). The panel consists of 497 genes of clinical interest (Supporting Table 1) of the following groups of disorders: Hereditary Cancer; Skeletal Dysplasias/Craniofacial diseases; Neuromuscular/Neurodegenerative; Intellectual Deficiency/Autism; Recessive Diseases Screening (http://genoma.ib.usp.br/pt-br).

We used SNPs to estimate FF, and read count to detect aneuploidy. To estimate FF we used 6739 probes distributed in 388 genes across the autosomes (minus chromosomes 18 and 21 – excluded because they are the most common trisomies, in which case can affect minor allele fraction – MAF – estimative), comprising approximately 1.5Mb. For detection of trisomy of chromosome 21 (T21)

we excluded variable regions (frequently observed as false positive CNVs calls in Hidden Markov Model – HMM - analysis of NGS data in more than 100 patients that were sequenced by the same panel for genetic diseases) to minimize variance in the read count analysis. After this correction, we used 240 probes distributed in 19 genes (approximately 33.5 Kb) across chromosome 21 for T21 detection.

- ***Fetal Fraction Estimation / Model Fitting and Evaluation***

Maternal plasma has a mixture of cfDNA from the mother and fetus. For any biallelic SNP there are four possible combinations of maternal/fetal genotypes. *A priori*, we do not know the genotype combination at each interrogated *locus*. Therefore, we used MAF information for the FF estimation (Table 1).

Table 1: MAF given FF for different maternal/fetal genotype combinations in maternal plasma

Maternal-fetal genotypes			Genotype	Minor Allele	Mean of B allele fraction (MAF)
	1	Maternal	AA	B	0
		Fetal	AA		
	2	Maternal	AA	B	FF/2
		Fetal	AB		
	3	Maternal	AB	B	0.5-FF/2
		Fetal	AA		
	4	Maternal	AB	B	0.5
		Fetal	AB		

Based upon MAF information we generated a model in order to fit our data and estimate FF (developed with R, v. 3.2.3). The model takes into account only

biallelic *loci*, incorporates an error rate due to the finite sampling and assumes an *a priori* maternal/fetal genotype combination frequency in maternal plasma (as seen in our own data and described by Jiang and colleagues (2012):

- Homozygous/homozygous: 0.7
- Homozygous/heterozygous: 0.1
- Heterozygous/homozygous: 0.1
- Heterozygous/heterozygous: 0.1

The model generates simulated samples based upon characteristics of the test sample (SNP number and coverage) for a range of different FF values (0 to 0.4, by 0.01). Simulations are then performed for each one of the tested FF values, and the distribution of the MAF values (range: 0.02-0.25, corresponding to the most informative peak - mother homozygous, child heterozygous) is evaluated for the test sample and compared to the same range for each one of the different FF simulated samples. Fitting is then performed based on similarity between simulated and test sample, calculated as the difference between the two vectors of MAF frequency values. Fitting can be performed multiple times and have the mean and standard deviation calculated.

We used simulated samples with different mean coverage values (50X, 100X, 150X, 200X, and 300X), SNP numbers (50, 100, 200, 500, 750, 1000, 1500, 2000, 3000, 4000, 5000) and FF (0 to 0.4, by 0.01) in order to evaluate the model and how these characteristics affect fitting accuracy

We also evaluated the fit procedure accuracy using the mock samples and non-pregnant samples (either mother or child), since we have the expected FF for the mock samples (predicted using fetal-specific alleles), as well as for the non-pregnant samples (expected to be 0). For FF estimation using SNPs we filtered the vcf by:

- Only SNPs on autosomes other than chromosomes 18 and 21
- Excluded LowQual SNPs
- Base coverage>= 100 X

- ***Detection of Trisomy 21 and 18***

For trisomy of chromosomes 21 (T21) and 18 (T18) detection, we used read count generated by GATK Depth of Coverage. For each chromosome we calculated the chromosome proportion, defined as sum of reads on that chromosome divided by total reads of autosomes minus the chromosome of interest.

The reference dataset of each chromosome consisted of pregnant samples (including mock samples) with fetus unaffected by the trisomy. To normalize the reference dataset we calculated the median and standard deviation for the reference dataset, and removed the samples falling outside of 3 median absolute deviations. Importantly, the reference dataset has to contain samples sequenced with the same platform (MiSeq or HiSeq).

We then used a Z-score approach to calculate the genomic representation of the chromosome of interest compared to the reference dataset for each test sample:

$$\text{Z-score}_{\text{test sample}} = (P_{\text{test sample}} - P_{\text{mean reference samples}}) / SD_{\text{mean reference samples}}$$

P = proportion of the chromosome of interest
SD = standard deviation

- ***Fetal Sex Determination***

For the fetal sex determination, we used chromosome Y read count. We counted the reads covering the *SRY* gene using GATK Depth of Coverage and calculated a proportion of chromosome Y, as sum of reads of chromosome Y divided by total sum of autosome reads. The reference dataset consisted of female fetus pregnancies (including mock). It is important to note that the reference dataset has to contain samples sequenced with the same platform (MiSeq or HiSeq). Normalization was applied and Z-score was calculated as for the detection of T21.

- ***Detection of Monogenic Disease / Variant Interpretation***

Skeletal dysplasia is a group of bone and cartilage disorders that affect fetal development *in utero* or postnatally. Prenatal onset SDs are clinically detectable through gestational ultrasound presenting limb defects or reduction. Many of the prenatal onset SDs are autosomal dominant and lethal, but some of them are non-lethal. To have molecular confirmation of the lethality of the fetus prior to birth would help the management of the pregnancy.

The availability of probes for several genes of clinical interest in our panel (including several forms of SDs) allowed us to perform a specific analysis for this disease, aiming to perform a proof of concept analysis in our data for the prenatal detection of monogenic diseases.

For analysis of possibly pathogenic variants, we performed variant call individually using Mutect (CIBULSKIS et al., 2013) and annotated it using Annovar (WANG; LI; HAKONARSON, 2010) and several public databases (ExAC, Exome Variant Server, 1000 Genomes), including our in-house database of 609 Brazilian control exomes. We filtered for rare variants (minor allele frequency < 0.5%) present only in genes related to Dysplasia/Craniofacial disorders (Supporting Table 2). For *de novo* variants in the fetus only we expected to detect the variant MAF as approximately half of FF.

- ***Blind Dataset for validation***

For blind validation of our methodology, we used 8 pregnant samples, comprising controls and fetus affected by T21 or T18, not known previously to test result. Library preparation and sequencing (using HiSeq) was performed as for the rest of the samples used in this work, as described above.

RESULTS

- ***Sample Characterization and Sequencing***

A total of 69 peripheral blood samples were collected, being 45 from pregnant women and 16 from non-pregnant individuals (8 mothers and children pairs). The pregnant women were aged between 20 and 46 (mean: 32.5, standard deviation: 5.86) and 10-36 gestational weeks (mean: 20.4, standard deviation: 9) (Supporting Table 3). Among the 8 non-pregnant pairs used to generate the mock samples, we collected 2 children affected by Down syndrome (T21), 1 affected by Edwards syndrome (T18) and 5 non-affected (Supporting Table 4).

Sequencing of 47/69 samples was performed using MiSeq (33 pregnant women and 14 non-pregnant individuals) yielded an average of 15.2 millions raw reads per sample (ranging from 8,695,612 - 33,517,518). Mean coverage in bam files was 191.65X (39.99X - 294.3X, median: 208.6X).

Sequencing of 14/69 samples using HiSeq (12 pregnant women and 2 non-pregnant individuals) yielded an average of 106,9 millions raw reads per sample (ranging from 35,316,152- 246,692,132). Mean coverage in bam files was 519.92X (201.6X – 928.38X, median: 522.75X). Total average coverage (MiSeq and HiSeq altogether) comprised 267X (Median: 222X).

- ***Mock Samples***

Mock samples generated by fastq file mixture instead of cfDNA mixture of mother and child have the advantage of allowing for files with multiple “fetal fractions” with low cfDNA input and cost, since it only requires sequencing mother and child once. However, there are intrinsic differences of PCR duplicates in the fastq files of both mother and child, which can be a confounding factor when estimating FF directly from the mock sample. In order to have an accurate expected FF for the mock samples, we employed two methods: fetal-specific alleles using maternal

genotype information (LO, Y. M. D. *et al.*, 2010; LIAO *et al.*, 2012) (using only SNP positions with a coverage of at least 100X) and chromosome Y read count for male pregnancies (CHIU et al., 2011; HUDECOVA et al., 2014). The estimated FFs using both approaches (fetal-specific alleles and chromosome Y read count) are strongly correlated, especially for high coverage samples (Pearson correlation r^2–all samples: 0.846, p-value = 2.527e-06; coverage >= 100x: 0.951, p-value = 1.506e-08; coverage >= 150X: 0.997, p-value = 8.449e-07; coverage >= 200X: 0.996, p-value = 0.05119). Since Y read count is only applicable for male pregnancies, we used FF predicted with fetal-specific alleles as our expected FF.

- ***Fetal Fraction Estimation Model Evaluation by Synthetic Datasets***

The fitting of the samples is performed, as explained in the Methodology section, by comparing the MAF values distribution between the test sample and simulated samples for the specific MAF range (0.02–0.25) Figure 5.

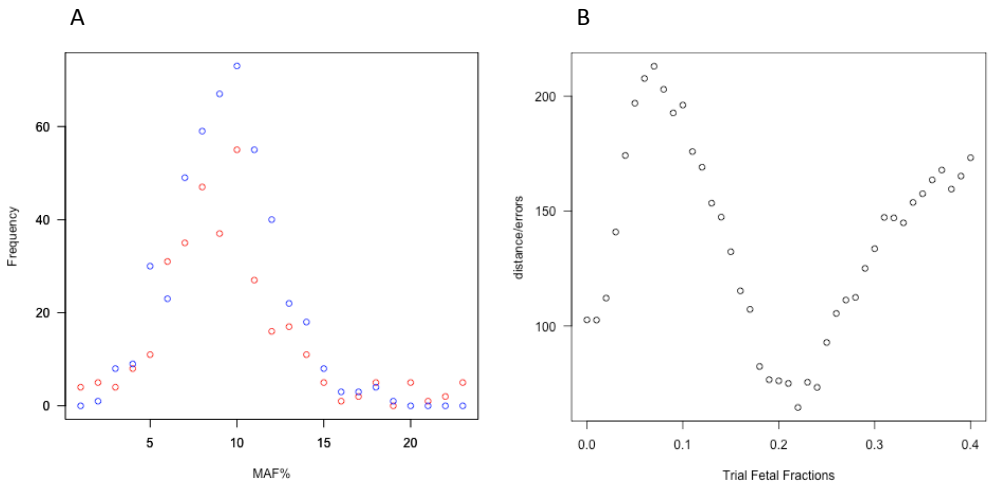

Figure 5: Fit procedure for FF estimation. A) Distribution of MAF values for both test and simulated samples; Red: test sample; Blue: fitted simulated sample (most similar to the test sample). B) Fitting of the test sample to the closest simulated samples.

Using the simulated samples we can see that mean coverage and SNP number affect the fitting using our model, as expected Figure 6.

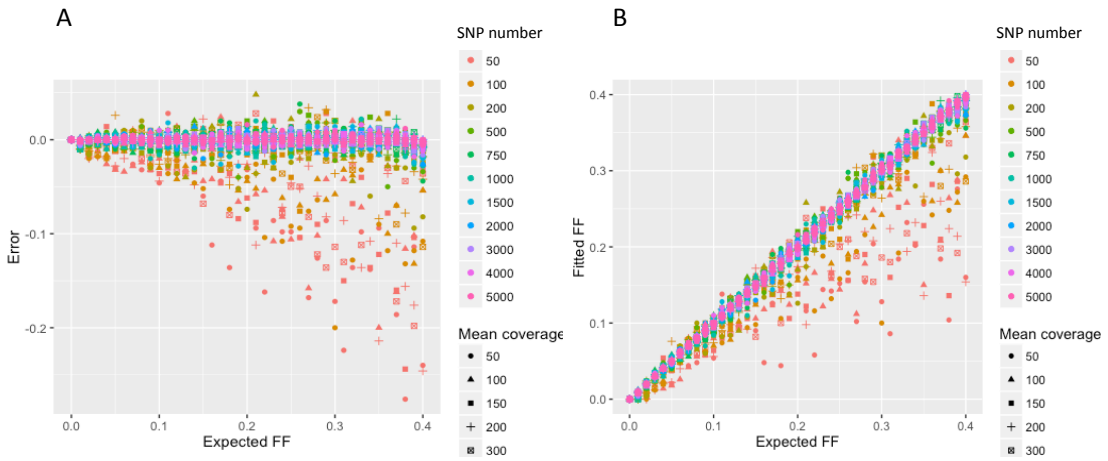

Figure 6: Evaluation of the model according to different FFs. A) Evaluation of the fitting according to different mean coverage and SNP number. B) Error (Fitted-Expected) according to different mean coverage and SNP number. Shape and color incorporate both mean coverage and SNP number values, respectively.

We also tested the model fitting accuracy for mean coverage and SNP number values obtained for our MiSeq sequenced samples (150X and 2000, respectively), which are lower than the samples sequenced using HiSeq. We found high correlation between expected and fitted FF values (Pearson correlation r2 = 0.999, p-value < 2.2e-16), with median degree of deviation of 0.000 (-0.033–0.050), calculated as: (Expected-Fitted)/Expected (Figure 7).

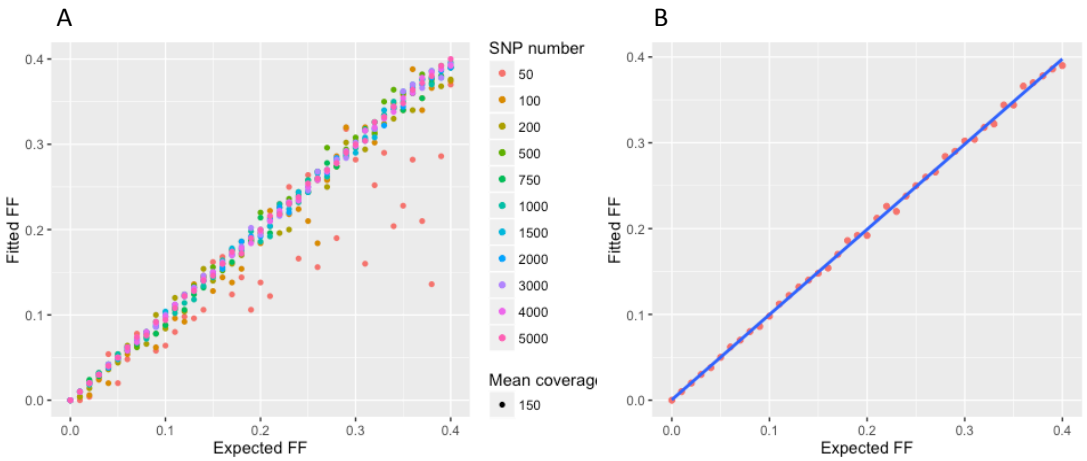

Figure 7: Evaluation of the fitting for mean coverage value = 150X. A) Evaluation of the fitting according to different SNP numbers. B) Evaluation of the fitting for SNP number = 2000. Shape and color incorporate both mean coverage and SNP number values, respectively.

- ***Fetal Fraction Estimation Model Evaluation by Experimental Datasets***

We estimated FF for our mock samples and non-pregnant samples in order to evaluate the model fitting (Supporting Table 5). Correlation between the Expected and Fitted values for these samples is very high (Pearson correlation r2= 0.994, p-value < 2.2e-16) (Figure 8).

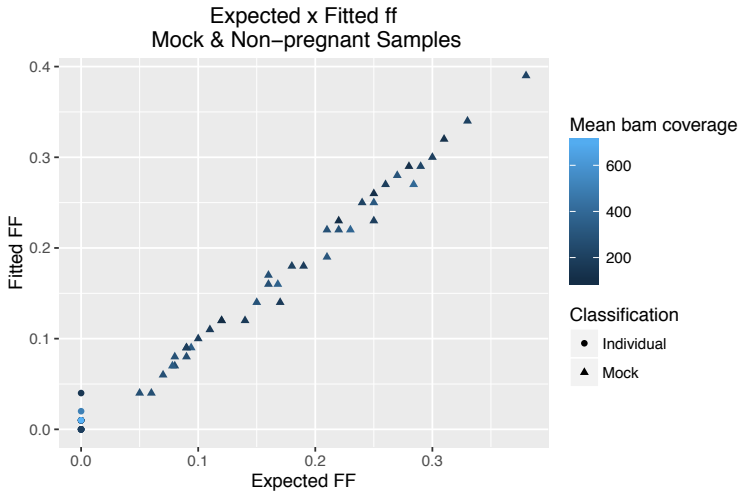

Figure 8: Evaluation of the modeled FF and the mean coverage effect. Individual: Non-pregnant sample. Shape and color incorporate both individual classification and mean coverage value, respectively.

The developed model was then used to estimate FF for all samples (mock, pregnant and non-pregnant). After vcf filtering, our samples had average SNP number of 4162 (11-5529, median: 4423) and 3990 (3514-4327, median: 4018) for MiSeq samples and HiSeq samples respectively (Supporting Table 5).

Mean fitted FF for pregnant samples was 0.12, varying between 0.02-0.30. Correlation analysis showed a strong positive correlation between FF and gestational age (Pearson correlation r^2 = 0.5, p-value = 4.4E-04) (Figure 9). We did not find a significant association between FF and maternal weight (Pearson correlation r^2 = -0.137; p-value =0.38) (Figure 10).

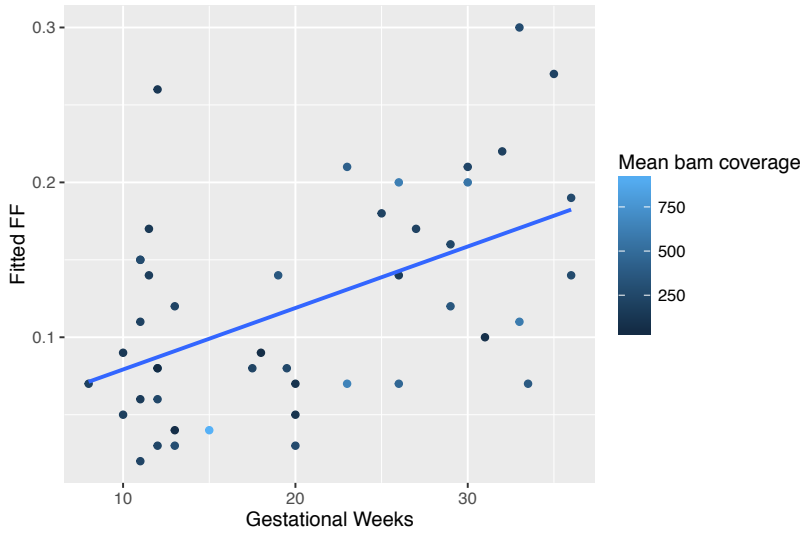

Figure 9: Correlation between gestational week and FF.

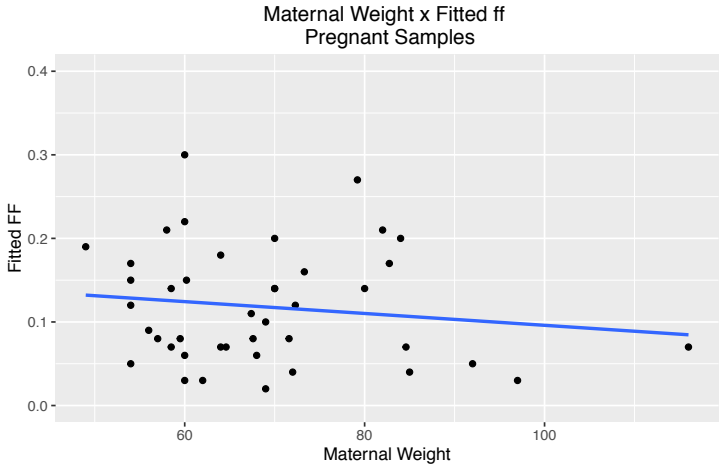

Figure 10: Correlation between FF and maternal weight (kg).

- ***Fetal Sex Determination***

For fetal sex determination, the normalized reference dataset consisted of 33 and 10 samples for MiSeq and HiSeq, respectively. Chromosome Y proportion and Z-score was calculated for each mock and pregnant sample (Supporting Table 6). Male fetus pregnancies have an average proportion of 8.9E-05 (1.27E-05 – 2.38E-04) and Z-score of 185.35 (26.27 – 507.2), while female fetus pregnancies have an

average proportion of 5.05E-07 (0 – 6.05E-06) and Z-score of 0.2 (-1.05 – 12.09). The groups do not overlap and can therefore be easily distinguished. For the 81 samples for which we had confirmation of fetal sex (40/40 mock samples and 41/45 pregnant samples), we observed 100% accuracy (Figure 11), with a sensitivity of 100% (95% CI: 90.51%-100%) and specificity of 100% (95% CI: 91.96%-100%).

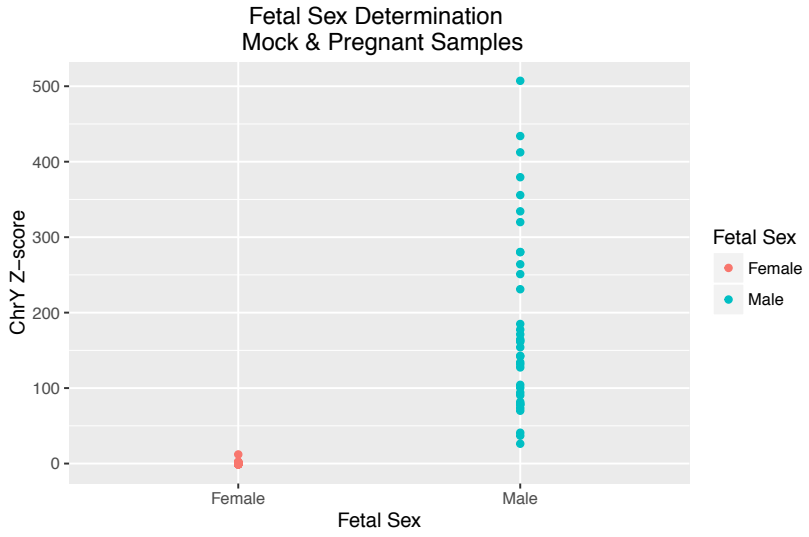

Figure 11: Fetal sex determination. Chromosome Y Z-score according to fetal sex.

- ***Trisomy 18 Detection***

After normalization, the euploid reference dataset for T18 detection consisted of 58 and 11 samples for MiSeq and HiSeq, respectively. We then calculated chromosome 18 Z-score for all 83 samples (5 mock T18, 35 mock not-T18 and 43 non-affected pregnant samples – for 2 we did not had fetus diagnosis confirmation), independently of FF: For the mock T18 samples, 2/5 had a positive Z-score (threshold 3.0) and 3/5 a negative Z-score (false negatives) (Supporting Table 7).

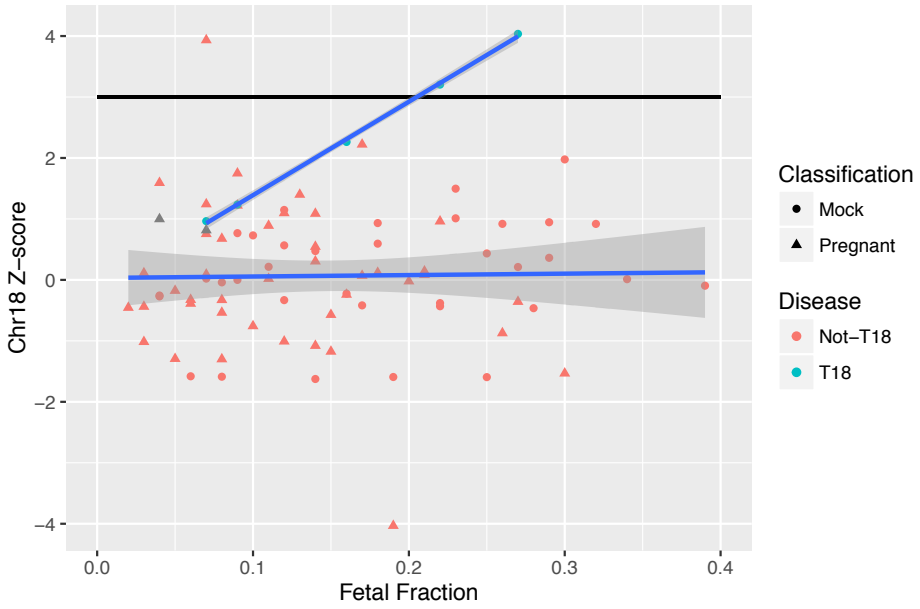

Figure 12: T18 detection. Chromosome 18 Z-score as function of FF. Color and shape incorporate both disease status and type of sample, respectively.

Using a threshold of FF>0.04 as inclusion criteria, we have 76 samples (5 mock T21, 33 mock not-T21 and 38 pregnant samples) with a sensitivity of 40% (95% CI: 5.27% to 85.34%) and a specificity of 98.59% (95% CI: 92.40% to 99.96%). The 43 non-affected pregnant samples had only one false positive (Z-score = 3.93, FF = 0.07) (Figure 12).

- ***Trisomy 21 Detection***

After normalization, the euploid reference dataset for T21 detection consisted of 54 and 16 samples for MiSeq and HiSeq, respectively. We then calculated chromosome 21 Z-score for all 83 samples (10 mock T21, 30 mock not-T21 and 43 non-affected pregnant samples – for 2 we did not had fetus diagnosis confirmation), independently of FF: For the mock T21 samples, 8/10 had a positive Z-score (threshold 3.0) and 2/10 a negative Z-score (false negatives) (Supporting Table 7). The 43 non-affected pregnant samples had only one false positive (Z-score = 6.29, FF = 0.19) (Figure 13).

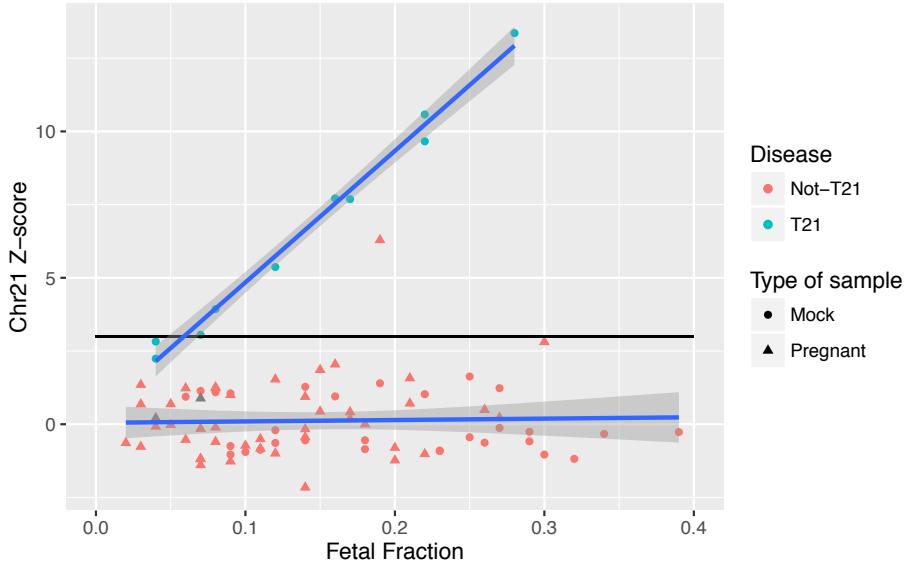

Figure 13: T21 detection. Chromosome 21 Z-score as function of FF. Color and shape incorporate both disease status and type of sample, respectively

Using a threshold of FF>0.04 as inclusion criteria, we have 76 samples (8 mock T21, 30 mock not-T21 and 38 pregnant samples) and a sensitivity of 100% (95% CI: 63.06% to 100.00%) and a specificity of 98.53% (95% CI: 92.08% to 99.96%).

We observed a positive correlation between FF and the Z-score values for the T21 affected samples (Pearson correlation r^2= 0.994, p-value = 6.013e-09), while this is not observed for non-affected samples (Pearson correlation r^2= -0.033, p-value= 0.7821).

- ***Blind Dataset for validation***

We also performed sequencing (using HiSeq) and analysis of 8 additional samples, comprising affected (T21 and T18) and unaffected pregnant samples. The sequencing of these samples yielded an average of 16.5 million raw reads per sample (ranging from 10,338,602 – 23,619,856). Mean coverage in bam files was 56.8X (32.7X – 83.5X, median: 57.6X).

We observed an accuracy of 100% of these samples regarding fetal sex determination and T21 detection (Table 2). For T18, we were not able to detect the affected sample (FD1500107), which has partial trisomy of chromosome 18.

Table 2: Test results for blind dataset.

Sample	Trisomy *	Test Result T21 (Z-score)	Test Result T18 (Z-score)	Fetus Gender	Test Result Fetal Sex (Z-score)
FD1500110	T21	T21 (12.9)	Not-T18 (1.39)	Male	Male (55.7)
FD1500068	T21	T21 (8.5)	Not-T18 (1.2)	Male	Female (-1.05)
FD1500092	T21	T21 (7.8)	Not-T18 (0.68)	Male	Male (51.4)
FD1500073	T21	T21 (6.8)	Not-T18 (2.12)	Male	Female (-1.06)
FD1500098	T21	T21 (4.4)	Not-T18 (2)	Female	Male (101.9)
FD.15.00142	Normal	Not-T21 (0.2)	Not-T18 (-0.11)	Female	Male (98.8)
FD.15.00141	Normal	Not-T21 (0.17)	Not-T18 (1.81)	Female	Female (-1.05)
FD1500107	T18 (11Mb)	Not-T21 (-0.55)	Not-T18 (1.91)	Male	Male (27.4)

*Normal: Not affected by T21 or T18.

- ***Monogenic Disease Detection – Skeletal Dysplasia***

Seven of the pregnant samples had prenatal ultrasound findings suggesting SD. After vcf analysis we detected known pathogenic variants in 5/7 samples (Table 3).

Table 3: Pregnant samples with fetus with SD.

Patient	FF*	GW**	Gene	Pathogenic Mutation	AF***	Confirmed on child
F10177-2	0.22	32	*FLNB*	NM_001457:c.605T>C (p.M202T)	0.11	✔
F10775-1	0.12	29	*FGFR3*	NM_000142:c.742C>T (p.R248C)	0.07	✔
F10609-1	0.2	30	*FGFR3*	NM_000142:c.742C>T (p.R248C)	0.08	✔
F11247-1	0.13	26	*FGFR3*	NM_000142:c.742C>T (p.R248C)	0.034	✔
F11077-1	0.07	23	*FGFR3*	NM_000142: c.1108G>T(p.G370C)	0.031	✗
F10951-1	0.07	33	✗	✗	✗	✗
F10774-1	0.11	33	✗	✗	✗	✗

*FF: fetal fraction
**GW: gestational weeks
***AF: allele fraction

We detected a pathogenic variant in sample F10177-2, located at the *FLNB* gene (NM_001457:c.605T>C:p.M202T) and associated with a lethal form of SD (Ateleosteogenesis type 1/Boomerang dysplasia) (DANIEL et al., 2012).

We identified pathogenic mutations in the *FGFR3* gene in 4 patients. *FGFR3* mutations are associated with Thanatophoric Dysplasia (TD), an autosomal dominant disorder, which was the initial diagnostic hypothesis (DH) for the three samples we detected the mutation NM_000142:c.742C>T:p.R248C, the most common mutation associated with TD type I (TAVORMINA et al., 1995; WILCOX et al., 1998). The

other sample harboring an *FGFR3* mutation (F11077-1) had an initial DH of Campomelic dysplasia, with unspecific ultrasound findings (short bent bones, brachycephaly and narrow thorax). This patient (F11077-1) has a rare mutation in *FGFR3*, associated with TD type I as well (ROUSSEAU et al., 1996).

We were unable to detect pathogenic mutations in 2 samples (F10951-1 and F10774-1), which had an initial DH of osteogenesis imperfecta (OI). We did not have the child's genomic DNA to verify if it was a methodological reason (unable to detect it noninvasively) or if the mutation is not present in our panel. Although we did not find a pathogenic mutation for these cases, we did detect a VUS (variant of unknown significance) for patient F10774-1 (NM_001235.3:c.580C>A:p.R194S) located at *SERPINH1*, a gene already associated with a recessive form of OI (BONAFE et al., 2015).

The MAF of these variants is about half of FF, as expected for *de novo* variants associated with autosomal dominant disorders. No other variant within the expected MAF was classified as pathogenic or probably pathogenic using ACMG criteria (RICHARDS et al., 2015). Results were confirmed by sequencing the child genomic DNA after birth, when available.

DISCUSSION

We have developed a NIPT for genetic diseases using NGS that incorporates the following analysis: FF estimation (using only maternal plasma sequencing data), fetal sex determination, trisomy detection and monogenic disease detection. A key strength of this study is the incorporation of all analyses in one single test, which was performed with the same gene panel used for the regular clinical genomic diagnosis in our center. Such strategy requires only minimal modifications if the panel is up to date. To our knowledge, this is the first work to create an integrated test, and it has the advantage to allow samples with different diagnostic purposes within the same laboratory workflow. While high coverage exome sequencing is not yet financially feasible for prenatal testing, this approach opens up the possibility to test hundreds

of monogenic diseases using NGS by targeting all coding sequences instead of solely relying on investigation of mutational hot spots.

Fetal fraction determination aids in avoiding false negative results and better detecting point mutations. Therefore, we developed a model to estimate FF using only plasma sequencing data regardless of fetal sex. Other groups use SNPs from targeted sequencing data to predict FF through a statistical binomial mixture model, relying on several mother-child genotype combinations to correctly predict FF (JIANG *et al.*, 2012; SPARKS *et al.*, 2012; KOUMBARIS *et al.*, 2016). We showed that it is possible to perform FF estimation using simpler statistics (MAF values vector comparison in R) using only the most informative genotype combination (mother homozygous, child heterozygous).

Since FF is an important factor in NIPT accuracy, and has been correlated with different maternal traits in other populations, we investigated its correlation with factors such as gestational age and maternal weight in Brazilian pregnant women. We found a positive correlation between FF and gestational age, in accordance with other reports (LO, Y. M. *et al.*, 1998; ZIMMERMANN *et al.*, 2012; HUDECOVA *et al.*, 2014; RAVA *et al.*, 2014; ZHOU *et al.*, 2015; XU *et al.*, 2016). On the other hand, we did not find a significant correlation between FF and maternal weight as reported by others (ASHOOR *et al.*, 2012; WANG, E. *et al.*, 2013; HUDECOVA *et al.*, 2014). This lack of correlation, however, may be attributable to small sample size.

In this work, we established a threshold of 0.04 of FF for T21 detection, as reported in the literature, for better accuracy. Literature data indicate high sensitivity and specificity for T21 detection using NGS, varying between 94.4% - 100% and 97.95% - 100%, respectively (CHIU et al., 2011; GIL et al., 2015), and our test sensitivity and specificity values lie between these ranges, showing that we have a high sensitivity and specificity for T21 detection.

The 2 false negative samples for T21 detection have fitted FFs of 0.04, which is the detection limit of the literature (EHRICH et al., 2011; NORTON et al., 2012; PALOMAKI et al., 2012; SPARKS et al., 2012), so they were expected to present low Z-scores. As for T18 detection, the 3 false negative samples have fitted FFs varying between 0.07 and 0.16, which is very high compared to the detection limit of the

literature (4%) (EHRICH et al., 2011; NORTON et al., 2012; PALOMAKI et al., 2012; SPARKS et al., 2012)

It has been shown that sequencing of chromosome 18 has a larger coefficient of variation than chromosome 21, which leads to a less precise measurement of genomic representation of the chromosome 18 (BENJAMINI; SPEED, 2012; CHEN et al., 2011; CHIU et al., 2011; FAN et al., 2008). Although it is known that T18 detection results are less accurate than those for T21, there is room for us to improve T18 detection, since our sensitivity is below what is demonstrated in the literature (CHEN et al., 2011; GIL et al., 2015; KOUMBARIS et al., 2016). One possible way to overcome this issue is by removing the segments of chromosome 18 that show an intrinsic high variability of read count, instead of using the entire chromosome read count.

We also performed the test in a blind dataset for validation, resulting in a 100% accuracy for fetal sex determination and T21. We were unable to detect the one T18 sample in the blind dataset. This sample has a partial trisomy of chromosome 18 (11Mb), and this is possibly the reason for the lack of detection, since we used total chromosome read count detection, and this partial trisomy accounts for about 22.5% of the chromosome 18. An improvement that can be made is to use segments of the chromosome instead of the total read count.

The false positive results for both T18 and T21 can be due to several factors: confined placental mosaicism, fetal mosaicism, vanishing twin, or even maternal malignancies (OSBORNE *et al.*, 2013; GRATI *et al.*, 2014; BIANCHI *et al.*, 2015). We have a high positive correlation for the T21-affected samples, and the false positive sample did not fall into the correlation line, as reported by others (HUDECOVA et al., 2014). It is important to note that one of the pregnant women (F10117-1) was referred to us with a positive diagnosis for T21 from a different clinical service. Our test was negative for T21 and the fetus was indeed unaffected, as confirmed postnatally.

As shown previously, it is possible to perform fetal sex determination using targeted NGS (CHIU et al., 2011; KOUMBARIS et al., 2016). In this work, we demonstrate that it is possible to determine fetal sex with 100% accuracy using only one probe on chromosome Y, instead of multiple probes (KOUMBARIS et al., 2016).

Monogenic disease test for SD was performed for seven SD cases, with a detection rate of 71% (5/7), demonstrating our test ability to incorporate detection of monogenic diseases, especially *de novo* or paternally inherited variants. Our noninvasive detection rate of SD is similar to the postnatal detection rate of SD (75/125 = 60%), using child genomic DNA (unpublished data from our center).

Tanatophoric, Achondroplasia and Osteogenesis Imperfecta are among the most common types of SD (MILKS; HILL; HOSSEINZADEH, 2017; UNGER, 1993). For patient F11077-1, which had an initial DH of Campomelic Dysplasia, we confirmed the diagnosis as TD type I. Despite the existence of clinical overlap, this differential diagnosis is important, because TD is lethal while Campomelic Dysplasia is not always lethal. This differential diagnosis is also important for the medical team in the postpartum management as well as for psychological preparation of the family.

In the 2 patients for whom pathogenic mutations were not detected (F10951-1 and F10774-1), the DH was OI. For these patients, we cannot discard the hypothesis of the pathogenic mutation being in a gene absent from our panel, since the number of genes associated with SD has grown at a fast pace in recent years, especially due to the advent of NGS (BONAFE et al., 2015). Another possible explanation is that the mutation could lay in intronic region, or it is a deletion, which could be missed with the currently available tools. For the patient F10774-1 we detected a VUS in a gene already associated with a recessive form of OI. It is possible that this patient has a recessive form of the disease, and the lack of identification of the second mutation might be a limitation of identifying mutations present in the mother. In that case, the patient has a FF of 0.11. We would expect the mutation present in the mother to have an allele fraction (AF) increase of 0.055, so instead of having AF of 0.5 (if only present in the mother), would be 0.55 (if present both in the mother and in the fetus), and that small difference might not be identified.

Chitty and colleagues (CHITTY et al., 2015) have recently demonstrated the effectiveness of NIPT to detect *FGFR3*-related SDs. However, they targeted hotspots in the gene, which has a pitfall of possibly losing pathogenic variants. Dan and colleagues (DAN et al., 2016) show the feasibility of using targeted sequencing for 16 genes. However, they used maternal and paternal genomic DNA sequencing for variant detection. Comparatively, our test may be more advantageous because we

are covering the entire coding sequence of hundreds of genes associated with monogenic disorders (therefore covering all exonic variants and many differential diagnoses), and also because we are able to perform the detection using only plasma sequencing, lowering costs and turnaround time for the test.

It is important to note that for the 5 SD cases in which we detected the pathogenic mutation, the AF is about half of the FF, which is expected for autosomal dominant disorders. This also demonstrates that our FF estimation model is accurate and helpful for the detection of the pathogenic mutation, since we can target the mutation within the expected AF according to the disease inheritance model.

We showed the relevance of using targeted sequencing to develop an integrated NIPT (using only maternal plasma), combining all analyses (fetal fraction estimation, fetal sex determination, trisomy and monogenic disease detection). Further reduction of sequencing costs will enable an even higher coverage, improving the ability to detect autosomal recessive or X-linked mutations more accurately, where the mother is heterozygous for the variant.

CONCLUSION

To our knowledge, we are the first group in Brazil to develop an in-house, non-invasive prenatal test using NGS. At the present time, NIPT is available for patients in Brazil, but it is performed through outsourcing technology or the test is performed abroad. In this work we demonstrated that, by using a feasible amount of targeted sequencing and relatively simple statistics, it is indeed possible to perform NIPT for several fetal diseases using only plasma sequencing data.

REFERENCES

ASHOOR, G. et al. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: effect of maternal and fetal factors. **Fetal diagnosis and therapy**, v. 31, n. 4, p. 237–43, jan. 2012.

BENJAMINI, Y.; SPEED, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. **Nucleic Acids Research**, v. 40, n. 10, p. e72–e72, maio 2012.

BIANCHI, D. W. et al. Noninvasive Prenatal Testing and Incidental Detection of Occult Maternal Malignancies. **JAMA**, v. 314, n. 2, p. 162, 14 jul. 2015.

BONAFE, L. et al. Nosology and classification of genetic skeletal disorders: 2015 revision. **American Journal of Medical Genetics, Part A**, v. 167, n. 12, p. 2869–2892, 2015.

CHEN, E. Z. et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. **PloS one**, v. 6, n. 7, p. e21791, jan. 2011.

CHITTY, L. S. et al. Non-invasive prenatal diagnosis of achondroplasia and thanatophoric dysplasia: next-generation sequencing allows for a safer, more accurate, and comprehensive approach. **Prenatal Diagnosis**, v. 35, n. 7, p. 656–662, jul. 2015.

CHIU, R. W. K. et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. **BMJ**, v. 342, n. jan11 1, p. c7401–c7401, 11 jan. 2011.

CIBULSKIS, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. **Nature Biotechnology**, v. 31, n. 3, p. 213–219, 10 fev. 2013.

DAN, S. et al. Non-invasive prenatal diagnosis of lethal skeletal dysplasia by targeted capture sequencing of maternal plasma. **PLoS ONE**, v. 11, n. 7, p. 1–16, 2016.

DANIEL, P. B. et al. Disease-associated mutations in the actin-binding domain of filamin B cause cytoplasmic focal accumulations correlating with disease severity. **Human Mutation**, v. 33, n. 4, p. 665–673, abr. 2012.

DEPRISTO, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. **Nature genetics**, v. 43, n. 5, p. 491–8, 2011.

EHRICH, M. et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. **American Journal of Obstetrics and Gynecology**, v. 204, n. 3, p. 205.e1-205.e11, mar. 2011.

FAAS, B. H. et al. Detection of fetal RhD - specific sequences in maternal plasma. **Lancet**, v. 352, p. 1196, 1998.

FAN, H. C. et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. **Proceedings of the National Academy of Sciences**, v.

105, n. 42, p. 16266–16271, 21 out. 2008.

FINNING, K. M. et al. Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RHD genotyping service. **Transfusion**, v. 42, n. 8, p. 1079–85, 2002.

GIL, M. M. et al. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. **Ultrasound in Obstetrics & Gynecology**, v. 45, n. 3, p. 249–266, mar. 2015.

GRATI, F. R. et al. Fetoplacental mosaicism: potential implications for false-positive and false-negative noninvasive prenatal screening results. **Genetics in Medicine**, v. 16, n. 8, p. 620–624, 13 ago. 2014.

HUDECOVA, I. et al. Maternal Plasma Fetal DNA Fractions in Pregnancies with Low and High Risks for Fetal Chromosomal Aneuploidies. **PLoS ONE**, v. 9, n. 2, p. e88484, 28 fev. 2014.

JIANG, P. et al. FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. **Bioinformatics**, v. 28, n. 22, p. 2883–2890, 15 nov. 2012.

KOUMBARIS, G. et al. Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal Sex. **Clinical Chemistry**, v. 62, n. 6, p. 848–855, 1 jun. 2016.

LAM, K.-W. G. et al. Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to -Thalassemia. **Clinical Chemistry**, v. 58, n. 10, p. 1467–1475, 1 out. 2012.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 2010.

LIAO, G. J. W. et al. Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles. **Clinical Chemistry**, v. 57, n. 1, p. 92–101, 1 jan. 2011.

LIAO, G. J. W. et al. Noninvasive Prenatal Diagnosis of Fetal Trisomy 21 by Allelic Ratio Analysis Using Targeted Massively Parallel Sequencing of Maternal Plasma DNA. **PLoS ONE**, v. 7, n. 5, p. e38154, 29 maio 2012.

LO, Y. M. et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. **American journal of human genetics**, v. 62, n. 4, p. 768–75, abr. 1998.

LO, Y. M. D. et al. Early report Presence of fetal DNA in maternal plasma and serum. **The Lancet**, v. 350, p. 485–487, 1997.

LO, Y. M. D. et al. Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. **Science Translational Medicine**, v. 2, n. 61, p. 61ra91-61ra91, 8 dez. 2010.

MCKENNA, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, 1 set. 2010.

MILKS, K. S.; HILL, L. M.; HOSSEINZADEH, K. Evaluating skeletal dysplasias on prenatal ultrasound: an emphasis on predicting lethality. **Pediatric Radiology**, v. 47, n. 2, p. 134–145, 1 fev. 2017.

NEW, M. I. et al. Noninvasive Prenatal Diagnosis of Congenital Adrenal Hyperplasia Using Cell-Free Fetal DNA in Maternal Plasma. **The Journal of Clinical Endocrinology & Metabolism**, v. 99, n. 6, p. E1022–E1030, jun. 2014.

NORTON, M. E. et al. Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. **American Journal of Obstetrics and Gynecology**, v. 207, n. 2, p. 137.e1-137.e8, ago. 2012.

NORTON, M. E. et al. Cell-free DNA Analysis for Noninvasive Examination of Trisomy. **The New England journal of medicine**, v. 372, n. 17, p. 1589–97, 2015.

OSBORNE, C. M. et al. Discordant noninvasive prenatal testing results in a patient subsequently diagnosed with metastatic disease. **Prenatal Diagnosis**, v. 33, n. 6, p. 609–611, jun. 2013.

PALOMAKI, G. E. et al. DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study. **Genetics in Medicine**, v. 13, n. 11, p. 913–920, nov. 2011.

PALOMAKI, G. E. et al. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. **Genetics in Medicine**, v. 14, n. 3, p. 296–305, 2 mar. 2012.

RAVA, R. P. et al. Circulating Fetal Cell-Free DNA Fractions Differ in Autosomal Aneuploidies and Monosomy X. **Clinical Chemistry**, v. 60, n. 1, p. 243–250, 1 jan. 2014.

RICHARDS, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **Genetics in Medicine**, v. 17, n. 5, p. 405–423, 2015.

RIJNDERS, R. J. et al. Fetal sex determination from maternal plasma in pregnancies at risk for congenital adrenal hyperplasia. **Obstetrics and gynecology**, v. 98, n. 3, p. 374–8, 2001.

ROUSSEAU, F. et al. Missense FGFR3 mutations create cysteine residues in thanatophoric dwarfism type I (TD1). **Human Molecular Genetics**, v. 5, n. 4, p. 509–512, 1996.

SPARKS, A. B. et al. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. **American Journal of Obstetrics and Gynecology**, v. 206, n. 4, p. 319.e1-319.e9, abr. 2012.

TAVORMINA, P. L. et al. Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. **Nature Genetics**, v. 9, n. 3, p. 321–328, mar. 1995.

UNGER, Sheila; SCHERER, Gerd; SUPERTI-FURGA, Andrea. Campomelic Dysplasia. In: PAGON, Roberta A et al (Orgs.). **GeneReviews(®)**. Seattle (WA): University of Washington, Seattle, 1993.

VAN DER AUWERA, G. A. et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: **Current Protocols in Bioinformatics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013. v. 11p. 11.10.1-11.10.33.

WANG, E. et al. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. **Prenatal Diagnosis**, v. 33, n. 7, p. 1–5, jul. 2013.

WANG, K.; LI, M.; HAKONARSON, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic acids research**, v. 38, n. 16, p. e164, 2010.

WILCOX, W. R. et al. Molecular, radiologic, and histopathologic correlations in thanatophoric dysplasia. **American Journal of Medical Genetics**, v. 78, n. 3, p. 274–281, 1998.

WRIGHT, C. F. et al. Non-invasive prenatal diagnostic test accuracy for fetal sex using cell-free DNA a review and meta-analysis. **BMC Research Notes**, v. 5, n. 1, p. 476, jan. 2012.

XU, X.-P. et al. A Method to Quantify Cell-Free Fetal DNA Fraction in Maternal Plasma Using Next Generation Sequencing: Its Application in Non-Invasive Prenatal Chromosomal Aneuploidy Detection. **PLOS ONE**, v. 11, n. 1, p. e0146997, 14 jan. 2016.

ZHOU, Y. et al. Effects of Maternal and Fetal Characteristics on Cell-Free Fetal DNA Fraction in Maternal Plasma. **Reproductive Sciences**, v. 22, n. 11, p. 1–7, 1 nov. 2015.

ZIMMERMANN, B. et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. **Prenatal Diagnosis**, v. 32, n. 13, p. 1233–1241, dez. 2012.

SUPPORTING INFORMATION

Supporting Table 1: Genes present in the clinical panel.

Genes					
AARS	*COL3A1*	*FKTN*	*LZTR1*	*PCCB*	*SHANK3*
ABCD1	*COL5A1*	*FLNA*	*MAP2K1*	*PCDH15*	*SHH*
ACADM	*COL5A2*	*FLNB*	*MATN3*	*PDGFRA*	*SHOC2*
ACADVL	*COL6A1*	*FLNC*	*MBD2*	*PDHA1*	*SHOX*
ACAT1	*COL6A2*	*FMR1*	*MBD5*	*PDHX*	*SIX3*
ACP5	*COL6A3*	*FOLR1*	*MC2R*	*PDSS1*	*SLC17A6*
ACTA1	*COL9A1*	*FOXE1*	*MCCC1*	*PDSS2*	*SLC19A3*
ACTB	*COL9A2*	*FOXG1*	*MCCC2*	*PFKL*	*SLC22A5*
ACTG1	*COL9A3*	*FOXP2*	*MECP2*	*PHF8*	*SLC25A13*
ADA	*COMP*	*FTCD*	*MED25*	*PHGDH*	*SLC25A15*
ADAMTS18	*COQ2*	*FUS*	*MEGF10*	*PLCB4*	*SLC26A2*
ADAMTSL2	*COQ9*	*FXN*	*MEN1*	*PLEC*	*SLC26A4*
ADCK3	*CP*	*GADD45G*	*MESP2*	*PLOD1*	*SLC2A1*
AFG3L2	*CPS1*	*GALNS*	*MFN2*	*PMP22*	*SLC5A5*
ALDH3A2	*CREBBP*	*GALNT1*	*MID1*	*PMS1*	*SMAD4*
ALDH7A1	*CRISPLD2*	*GALNT12*	*MIR1205*	*PMS2*	*SMARCAL1*
ALPL	*CRTAP*	*GALT*	*MIR1206*	*POGZ*	*SNAP29*
ALX1	*CRYAA*	*GAMT*	*MIR1207*	*POLR1C*	*SNRPN*
ALX3	*CRYAB*	*GARS*	*MIR1208*	*POLR1D*	*SOD1*
ALX4	*CTDP1*	*GATM*	*MIR140*	*POMGNT1*	*SOS1*
AMN	*CTNNB1*	*GCDH*	*MIR200B*	*POMT1*	*SOS2*
ANO5	*CTSK*	*GCH1*	*MLH1*	*POMT2*	*SOST*
APC	*CUBN*	*GDAP1*	*MMAA*	*POR*	*SOX9*

APTX	*CUL7*	*GDF5*	*MMAB*	*POTED*	*SP7*
ARG1	*CYP21A2*	*GDF6*	*MMACHC*	*PPIB*	*SPAST*
ARHGAP29	*CYP27A1*	*GIF*	*MMADHC*	*PQBP1*	*SPG11*
ARX	*CYP7B1*	*GJA1*	*MMP13*	*PRKAR1A*	*SPG7*
ASL	*DAG1*	*GJB1*	*MMP9*	*PRPS1*	*SPR*
ASPA	*DBT*	*GJB2*	*MOCS1*	*PRSS1*	*SRY*
ASS1	*DCLRE1C*	*GJB3*	*MOCS2*	*PRX*	*STK11*
ATL1	*DDR2*	*GJB6*	*MPZ*	*PSAT1*	*SUMO1*
ATM	*DES*	*GLB1*	*MRAP*	*PSPH*	*TARDBP*
ATP7A	*DFNB31*	*GLI2*	*MRE11A*	*PTCH1*	*TBX1*
ATP7B	*DHCR24*	*GLI3*	*MSH2*	*PTEN*	*TBX22*
ATR	*DHCR7*	*GNAI3*	*MSH3*	*PTH1R*	*TCAP*
AXIN2	*DLAT*	*GNAS*	*MSH6*	*PTPN11*	*TCF12*
B4GALT6	*DLL3*	*GRHL3*	*MSTN*	*PTS*	*TCF4*
BAP1	*DMD*	*GRIK1*	*MSX1*	*PVRL1*	*TCOF1*
BARD1	*DNAJB6*	*HADHA*	*MSX2*	*PVT1*	*TERT*
BCKDHA	*DNM2*	*HADHB*	*MTHFR*	*PYGM*	*TFAP2A*
BCKDHB	*DUOX2*	*HBB*	*MTM1*	*QDPR*	*TG*
BIN1	*DYM*	*HERC2*	*MTMR2*	*RAB23*	*TGFBR1*
BMP4	*DYNC2H1*	*HES7*	*MTR*	*RAB7A*	*TGFBR2*
BMPR1A	*DYRK1A*	*HLCS*	*MTRR*	*RABEP2*	*TGIF1*
BRAF	*DYSF*	*HMGCL*	*MUT*	*RAD50*	*TH*
BRCA1	*EDN1*	*HNRPDL*	*MUTYH*	*RAD51*	*TMC1*
BRCA2	*EFNB1*	*HRAS*	*MYC*	*RAD51C*	*TMIE*
BRIP1	*EGR2*	*HSPB1*	*MYO6*	*RAD51D*	*TMPRSS15*
BTD	*EIF2AK3*	*HSPB8*	*MYO7A*	*RAD51L3*	*TMPRSS3*
C21orf29	*EIF4A3*	*IDUA*	*MYOT*	*RAF1*	*TNFRSF11A*
CACNA1S	*ELN*	*IFT80*	*NAGS*	*RAG1*	*TP53*

CANT1	*EMD*	*IL11RA*	*NBN*	*RAG2*	*TP63*
CAPN3	*EPCAM*	*IL2RG*	*NDN*	*RAI1*	*TPM2*
CAV3	*EPHA3*	*IL7R*	*NDRG1*	*RARA*	*TPM3*
CBS	*ERF*	*IRF6*	*NDUFV2*	*RBBP8*	*TPO*
CCDC26	*ERG*	*ITGA7*	*NEFL*	*RECQL4*	*TRAPPC2*
CD3D	*ETHE1*	*IVD*	*NF1*	*REEP1*	*TRIM32*
CD3E	*EVC*	*JAG1*	*NIPA1*	*RET*	*TRIP11*
CDH1	*EVC2*	*JAK3*	*NIPBL*	*RGMA*	*TRPC6*
CDH2	*EXO1*	*JAM2*	*NKX3-2*	*RHD*	*TRPS1*
CDH23	*EXT1*	*KATNAL2*	*NOG*	*RIT1*	*TRPV4*
CDH7	*FAF1*	*KBTBD13*	*NOL8*	*RMRP*	*TSC1*
CDH8	*FAH*	*KCNE1*	*NOTCH3*	*ROR2*	*TSC2*
CDK4	*FANCA*	*KIAA1267*	*NPC1*	*RPS27L*	*TSHB*
CDKL5	*FANCC*	*KIF22*	*NPC2*	*RTTN*	*TSHR*
CDKN2A	*FANCD2*	*KIF5A*	*NPR2*	*RUNX1*	*TSHZ1*
CFL1	*FANCE*	*KRAS*	*NRAS*	*RUNX2*	*TSPEAR*
CFTR	*FANCF*	*L1CAM*	*NSD1*	*RYR1*	*TWIST1*
CHEK2	*FANCG*	*LAMA2*	*NT5C3*	*SALL1*	*UBE3A*
CHST14	*FBLN5*	*LAMP2*	*NTNG1*	*SATB2*	*USH1C*
CHST3	*FBN1*	*LARGE*	*OFD1*	*SBDS*	*USH1G*
CLCN1	*FECH*	*LEMD3*	*OGN*	*SBF2*	*USH2A*
CLDN14	*FGD4*	*LEPRE1*	*OTC*	*SCN4A*	*USH2A*
CLPTM1L	*FGF8*	*LFNG*	*OTOF*	*SCO2*	*USHBP1*
CLRN1	*FGFR1*	*LIFR*	*OXCT1*	*SEPN1*	*USP25*
CNTNAP2	*FGFR2*	*LIPI*	*PABPN1*	*SERPINH1*	*VAPB*
COL10A1	*FGFR3*	*LITAF*	*PAFAH1B1*	*SETBP1*	*VHL*
COL11A1	*FHL1*	*LMAN1*	*PAH*	*SETX*	*WNT3*
COL11A2	*FIG4*	*LMBRD1*	*PALB2*	*SGCA*	*XRCC3*

COL18A1	*FILIP1L*	*LMNA*	*PAPSS2*	*SGCB*	*ZAP70*
COL1A1	*FKBP10*	*LRP5*	*PAX8*	*SGCD*	*ZIC2*
COL1A2	*FKBP14*	*LRSAM1*	*PCBD1*	*SGCG*	
COL2A1	*FKRP*	*LTBP4*	*PCCA*	*SH3BP2*	

Supporting Table 2: Genes for Dysplasia/Craniofacial disorders present in the clinical panel.

Genes					
ACP5	*COL9A3*	*FIG4*	*LFNG*	*PTCH1*	*TBX22*
ADAMTS18	*COMP*	*FKBP10*	*LIFR*	*PTH1R*	*TBX6*
ADAMTSL2	*CREB3L1*	*FKBP14*	*LMNA*	*PVRL1*	*TCF12*
ALPL	*CRTAP*	*FLNA*	*LRP5*	*RAB23*	*TCOF1*
ALX1	*CTSK*	*FLNB*	*MATN3*	*RECQL4*	*TFAP2A*
ALX3	*DDR2*	*GALNS*	*MESP2*	*RMRP*	*TGFBR1*
ALX4	*DLL3*	*GDF5*	*MMP13*	*ROR2*	*TGFBR2*
ANO5	*DYM*	*GDF6*	*MMP9*	*RUNX2*	*TGIF1*
BMP1	*DYNC2H1*	*GJA1*	*MSX1*	*SALL1*	*TMEM38B*
CANT1	*EDN1*	*GLB1*	*MSX2*	*SATB2*	*TNFRSF11A*
CHST14	*EFNB1*	*GLI2*	*NKX3-2*	*SBDS*	*TP63*
CHST3	*EIF2AK3*	*GLI3*	*NOG*	*SERPINF1*	*TRAPPC2*
COL10A1	*EIF4A3*	*GNAI3*	*NPR2*	*SERPINH1*	*TRIP11*
COL11A1	*ELN*	*GNAS*	*OFD1*	*SH3BP2*	*TRPS1*
COL11A2	*ERF*	*GRHL3*	*PAPSS2*	*SHH*	*TRPV4*
COL18A1	*EVC*	*HES7*	*PLCB4*	*SHOX*	*TSHZ1*
COL1A1	*EVC2*	*HUWE1*	*PLOD1*	*SIX3*	*TWIST1*
COL1A2	*EXT1*	*IFITM5*	*PLOD2*	*SLC26A2*	*WNT1*
COL2A1	*FBLN5*	*IFT80*	*PLS3*	*SMARCAL1*	*WNT3*
COL3A1	*FBN1*	*IL11RA*	*POLR1C*	*SOST*	*ZIC1*
COL5A1	*FGF8*	*IRF6*	*POLR1D*	*SOX9*	*ZIC2*
COL5A2	*FGFR1*	*KIF22*	*POR*	*SP7*	
COL9A1	*FGFR2*	*LEMD3*	*PPIB*	*SPARC*	
COL9A2	*FGFR3*	*LEPRE1*	*PRKAR1A*	*TBX1*	

Supporting Table 3: Summary of pregnant samples.

Sample*	Sequencing	Fetal Sex**	Maternal Age (years)	Gestational Weeks	Mean Bam Coverage (X)
P2182	MiSeq	Male	35	27	208.6
F9940-1	MiSeq	Male	37	11.5	141.2
F9914-1	MiSeq	Female	35	13	219.1
F9892-1	MiSeq	Male	35	12	244.2
F9891-1	MiSeq	Male	32	20	269.7
F9614-1	MiSeq	Female	32	12	39.99
F9531-1	MiSeq	Male	36	20	97.65
F9530-1	MiSeq	Female	37	11	226.9
F9269-1	MiSeq	Male	46	10	186.9
F9198-1	MiSeq	Female	34	17.5	222
F9176-1	MiSeq	Female	30	11.5	173.6
F9151-1	MiSeq	Female	35	31	80.04
F9051-1	MiSeq	Female	35	11	271.6
F8998-1	MiSeq	Male	21	26	161.1
F8850-1	MiSeq	Female	44	11	203.9
F8775-1	MiSeq	NA	37	18	62.71
F8762-1	MiSeq	Female	38	10	147.2
F8660-1	MiSeq	Female	35	13	294.3
F8659-1	MiSeq	Female	36	12	233.1
F8642-1	MiSeq	Female	39	11	162.1
F8626-1	MiSeq	NA	27	12	132.2
F8614-1	MiSeq	Male	29	13	66.12
F10396-1	MiSeq	Male	40	36	267.9
F10351-1	MiSeq	Female	32	33	283.3

F10177-2	MiSeq	Female	20	32	194.2
F10157-1	MiSeq	Female	39	29	242.4
F10138-1	MiSeq	Male	33	25	210.2
F10117-1	MiSeq	Female	25	30	233.6
F10113-1	MiSeq	Male	34	35	212.5
F10063-1	MiSeq	Female	30	19.5	253.2
C26729	MiSeq	Male	28	20	105.6
C26139	MiSeq	Female	35	11	250.2
C18177	MiSeq	Female	25	36	269.9
P2183	HiSeq	Male	33	19	357.4
F11247.1	HiSeq	Female	30	26	620.6
F11091.1	HiSeq	NA	35	15	928.38
F11077.1	HiSeq	Female	20	23	641.7
F10951-1	HiSeq	Female	23	33.5	374.7
F10855-1	HiSeq	Male	36	23	418.3
F10795-1	HiSeq	Female	31	12	541.4
F10775-1	HiSeq	Male	27	29	356.5
F10774-1	HiSeq	Male	22	33	592
F10764-1	HiSeq	Female	26	26	463.8
F10609-1	HiSeq	NA	34	30	556.3
C24778	HiSeq	Male	33	8	201.6

*Sample name_fetalXX, where XX indicates the original fastq proportion admixture (in percentage).
**NA: Information not available.

Supporting Table 4: Summary of non-pregnant individuals.

Sample	Relationship	Sequencing	Disease*	Gender	Mean Bam Coverage (X)
C20598	Child	MiSeq	Normal	Male	209.7
C20672	Mother	MiSeq	Normal	Female	145.7
C23944	Child	MiSeq	Normal	Male	92.85
C23945	Mother	MiSeq	Normal	Female	84.55
C25147	Child	MiSeq	Normal	Male	215
C25148	Mother	MiSeq	Normal	Female	173
C26449	Child	MiSeq	Normal	Male	223.4
C26450	Mother	MiSeq	Normal	Female	265.3
F10114-1	Child	MiSeq	T21	Female	184.5
F10114-2	Mother	MiSeq	Normal	Female	261.4
F9219-1	Child	MiSeq	Normal	Female	186.8
F9219-2	Mother	MiSeq	Normal	Female	156.6
F9966-1	Child	MiSeq	T21	Female	198.6
F9966-2	Mother	MiSeq	Normal	Female	243
F10801-1	Child	HiSeq	T18	Female	722.1
F10801-2	Mother	HiSeq	Normal	Female	504.1

*Normal: not affected by T21 or T18.

Supporting Table 5: Sample information.

Sample*	Sequencing	Classification	Mean Bam Coverage (X)	Mean SNP number	Expected FF**	Fitted FF	Standard Deviation Fitted FF
F9966-2	MiSeq	Individual	243	4589	0	0	0
F9966-1	MiSeq	Individual	198.6	3866	0	0	0
F9219-2	MiSeq	Individual	156.6	4997	0	0	0
F9219-1	MiSeq	Individual	186.8	5413	0	0	0
F10114-2	MiSeq	Individual	261.4	4445	0	0	0
F10114-1	MiSeq	Individual	184.5	3867	0	0	0
C26450	MiSeq	Individual	265.3	4578	0	0	0
C26449	MiSeq	Individual	223.4	4180	0	0.01	0
C25148	MiSeq	Individual	173	3902	0	0.01	0.006
C25147	MiSeq	Individual	215	4364	0	0.01	0.006
C23945	MiSeq	Individual	84.55	1873	0	0.01	0
C23944	MiSeq	Individual	92.85	2573	0	0.01	0
C20672	MiSeq	Individual	145.7	3437	0	0.04	0.01
C20598	MiSeq	Individual	209.7	4204	0	0.01	0
F10801-2	HiSeq	Individual	504.1	4171	0	0.02	0
F10801-1	HiSeq	Individual	722.1	4105	0	0.01	0
F9966_fetal5	MiSeq	Mock	257.8	4652	0.09	0.08	0
F9966_fetal4	MiSeq	Mock	254.6	4635	0.08	0.07	0.006
F9966_fetal20	MiSeq	Mock	296.6	4791	0.27	0.28	0
F9966_fetal15	MiSeq	Mock	285.4	4754	0.22	0.22	0.006
F9966_fetal10	MiSeq	Mock	272.5	4722	0.16	0.16	0
F9219_fetal5	MiSeq	Mock	166.1	5227	0.11	0.11	0.006
F9219_fetal4	MiSeq	Mock	163.4	5197	0.09	0.09	0.01
F9219_fetal20	MiSeq	Mock	195.8	5529	0.3	0.3	0.006
F9219_fetal15	MiSeq	Mock	187.6	5459	0.25	0.23	0.015
F9219_fetal10	MiSeq	Mock	177.8	5370	0.19	0.18	0.006
F10114_fetal5	MiSeq	Mock	271	4503	0.06	0.04	0.012
F10114_fetal4	MiSeq	Mock	269.8	4494	0.05	0.04	0
F10114_fetal20	MiSeq	Mock	281.8	4566	0.21	0.22	0.01
F10114_fetal15	MiSeq	Mock	279.3	4557	0.16	0.17	0.006

F10114_fetal10	MiSeq	Mock	275.8	4526	0.12	0.12	0
C26449_fetal5	MiSeq	Mock	281.2	4654	0.08	0.08	0.006
C26449_fetal4	MiSeq	Mock	278.5	4640	0.07	0.06	0.01
C26449_fetal20	MiSeq	Mock	314.7	4801	0.25	0.25	0
C26449_fetal15	MiSeq	Mock	305	4775	0.21	0.19	0.006
C26449_fetal10	MiSeq	Mock	293.8	4723	0.15	0.14	0.006
C25147_fetal5	MiSeq	Mock	187.5	4096	0.1	0.1	0.006
C25147_fetal4	MiSeq	Mock	184.8	4063	0.09	0.09	0.006
C25147_fetal20	MiSeq	Mock	220.7	4407	0.29	0.29	0.01
C25147_fetal15	MiSeq	Mock	211	4325	0.24	0.25	0
C25147_fetal10	MiSeq	Mock	200	4220	0.18	0.18	0.006
C23944_fetal5	MiSeq	Mock	106	3226	0.22	0.23	0.012
C23944_fetal4	MiSeq	Mock	93.79	2575	0.12	0.12	0.006
C23944_fetal20	MiSeq	Mock	111.3	3613	0.28	0.29	0
C23944_fetal15	MiSeq	Mock	107.6	3408	0.25	0.26	0.006
C23944_fetal10	MiSeq	Mock	120.7	3940	0.31	0.32	0.006
C20598_fetal5	MiSeq	Mock	166.4	3841	0.17	0.14	0.01
C20598_fetal4	MiSeq	Mock	162.4	3766	0.14	0.12	0.006
C20598_fetal20	MiSeq	Mock	216.3	4439	0.38	0.39	0.015
C20598_fetal15	MiSeq	Mock	201.7	4289	0.33	0.34	0.021
C20598_fetal10	MiSeq	Mock	185.2	4101	0.26	0.27	0.006
F10801_fetal5	HiSeq	Mock	335.3	3737	0.094	0.09	0
F10801_fetal4	HiSeq	Mock	330.4	3730	0.078	0.07	0.006
F10801_fetal20	HiSeq	Mock	396.6	3817	0.284	0.27	0.011
F10801_fetal15	HiSeq	Mock	378.3	3790	0.23	0.22	0.011
F10801_fetal10	HiSeq	Mock	357.9	3764	0.168	0.16	0
P2182	MiSeq	Pregnant	208.6	4968	NA	0.17	0.006
F9940-1	MiSeq	Pregnant	141.2	3757	NA	0.17	0.01

F9914-1	MiSeq	Pregnant	219.1	4813	NA	0.12	0.006
F9892-1	MiSeq	Pregnant	244.2	5239	NA	0.06	0.012
F9891-1	MiSeq	Pregnant	269.7	4620	NA	0.03	0.006
F9614-1	MiSeq	Pregnant	39.99	11	NA	0.08	0.04
F9531-1	MiSeq	Pregnant	97.65	2999	NA	0.05	0.006
F9530-1	MiSeq	Pregnant	226.9	4213	NA	0.02	0.006
F9269-1	MiSeq	Pregnant	186.9	3862	NA	0.05	0.021
F9198-1	MiSeq	Pregnant	222	4404	NA	0.08	0.006
F9176-1	MiSeq	Pregnant	173.6	3875	NA	0.14	0
F9151-1	MiSeq	Pregnant	80.04	1906	NA	0.1	0.015
F9051-1	MiSeq	Pregnant	271.6	4584	NA	0.15	0.006
F8998-1	MiSeq	Pregnant	161.1	3707	NA	0.14	0.006
F8850-1	MiSeq	Pregnant	203.9	4222	NA	0.11	0
F8775-1	MiSeq	Pregnant	62.71	234	NA	0.09	0.045
F8762-1	MiSeq	Pregnant	147.2	3373	NA	0.09	0.012
F8660-1	MiSeq	Pregnant	294.3	4647	NA	0.03	0.006
F8659-1	MiSeq	Pregnant	233.1	4375	NA	0.03	0
F8642-1	MiSeq	Pregnant	162.1	5355	NA	0.06	0.01
F8626-1	MiSeq	Pregnant	132.2	4590	NA	0.26	0.006
F8614-1	MiSeq	Pregnant	66.12	362	NA	0.04	0.01
F10396-1	MiSeq	Pregnant	267.9	4656	NA	0.19	0.162
F10351-1	MiSeq	Pregnant	283.3	5166	NA	0.3	0.012
F10177-2	MiSeq	Pregnant	194.2	4814	NA	0.22	0.01
F10157-1	MiSeq	Pregnant	242.4	5066	NA	0.16	0.01
F10138-1	MiSeq	Pregnant	210.2	4188	NA	0.18	0.006
F10117-1	MiSeq	Pregnant	233.6	4356	NA	0.21	0
F10113-1	MiSeq	Pregnant	212.5	4898	NA	0.27	0.006
F10063-1	MiSeq	Pregnant	253.2	5258	NA	0.08	0
C26729	MiSeq	Pregnant	105.6	2176	NA	0.07	0
C26139	MiSeq	Pregnant	250.2	5085	NA	0.15	0.01
C18177	MiSeq	Pregnant	269.9	4367	NA	0.14	0
P2183	HiSeq	Pregnant	357.4	3974	NA	0.14	0.006
F11247.1	HiSeq	Pregnant	620.6	4188	NA	0.13	0.006
F11091.1	HiSeq	Pregnant	928.38	4327	NA	0.04	0.01
F11077.1	HiSeq	Pregnant	641.7	4201	NA	0.07	0.006
F10951-1	HiSeq	Pregnant	374.7	4070	NA	0.07	0
F10855-1	HiSeq	Pregnant	418.3	4018	NA	0.21	0
F10795-1	HiSeq	Pregnant	541.4	4038	NA	0.08	0
F10775-1	HiSeq	Pregnant	356.5	4009	NA	0.12	0
F10774-1	HiSeq	Pregnant	592	4235	NA	0.11	0.006
F10764-1	HiSeq	Pregnant	463.8	4002	NA	0.07	0
F10609-1	HiSeq	Pregnant	556.3	4134	NA	0.2	0.006
C24778	HiSeq	Pregnant	201.6	3514	NA	0.07	0

*Sample name_fetalXX, where XX indicates the original fastq proportion admixture (in percentage).
**Calculated by fetal-specific alleles using maternal genotype information.

Supporting Table 6: Chromosome Y Z-score and proportion values for fetal sex determination.

Sample*	Sequencing	Classification	Fetal Sex	Mean Bam Coverage (X)	Fitted FF	ChrY prop	ChrY Z-score
C20598_fetal20	MiSeq	Mock	Male	216.3	0.39	2.38E-04	507.21
C20598_fetal15	MiSeq	Mock	Male	201.7	0.34	2.04E-04	433.99
C25147_fetal20	MiSeq	Mock	Male	220.7	0.29	1.94E-04	412.39
F10396-1	MiSeq	Pregnant	Male	267.9	0.19	1.78E-04	379.45
C25147_fetal15	MiSeq	Mock	Male	211	0.25	1.67E-04	355.76
C26449_fetal20	MiSeq	Mock	Male	314.7	0.25	1.57E-04	334.25
C20598_fetal10	MiSeq	Mock	Male	185.2	0.27	1.50E-04	319.95
F10113-1	MiSeq	Pregnant	Male	212.5	0.27	1.32E-04	280.30
C23944_fetal10	MiSeq	Mock	Male	120.7	0.32	1.32E-04	280.18
C26449_fetal15	MiSeq	Mock	Male	305	0.19	1.24E-04	264.09
C25147_fetal10	MiSeq	Mock	Male	200	0.18	1.18E-04	251.04
F10138-1	MiSeq	Pregnant	Male	210.2	0.18	1.09E-04	231.01
C26449_fetal10	MiSeq	Mock	Male	293.8	0.14	8.71E-05	184.98
C23944_fetal5	MiSeq	Mock	Male	106	0.23	8.36E-05	177.51
C23944_fetal20	MiSeq	Mock	Male	111.3	0.29	8.06E-05	171.11
P2182	MiSeq	Pregnant	Male	208.6	0.17	7.75E-05	164.50
F10855-1	HiSeq	Pregnant	Male	418.3	0.21	9.97E-05	162.78
C20598_fetal5	MiSeq	Mock	Male	166.4	0.14	7.60E-05	161.30
C23944_fetal15	MiSeq	Mock	Male	107.6	0.26	7.26E-05	154.05
C20598_fetal4	MiSeq	Mock	Male	162.4	0.12	6.73E-05	142.74
C25147_fetal5	MiSeq	Mock	Male	187.5	0.1	6.70E-05	142.10
F9940-1	MiSeq	Pregnant	Male	141.2	0.17	6.33E-05	134.21
F8998-1	MiSeq	Pregnant	Male	161.1	0.14	6.20E-05	131.44
F8626-1	MiSeq	Pregnant	NA	132.2	0.26	6.06E-05	128.45
C25147_fetal4	MiSeq	Mock	Male	184.8	0.09	6.01E-	127.38

						05	
C26729	MiSeq	Pregnant	Male	105.6	0.07	4.93E-05	104.35
C26449_fetal5	MiSeq	Mock	Male	281.2	0.08	4.75E-05	100.51
F9891-1	MiSeq	Pregnant	Male	269.7	0.03	4.45E-05	94.11
P2183	HiSeq	Pregnant	Male	357.4	0.14	5.57E-05	90.48
C26449_fetal4	MiSeq	Mock	Male	278.5	0.06	3.87E-05	81.74
F10775-1	HiSeq	Pregnant	Male	356.5	0.12	4.92E-05	79.80
C23944_fetal4	MiSeq	Mock	Male	93.79	0.12	3.70E-05	78.11
F10774-1	HiSeq	Pregnant	Male	592	0.11	4.80E-05	77.82
F9269-1	MiSeq	Pregnant	Male	186.9	0.05	3.48E-05	73.42
F9892-1	MiSeq	Pregnant	Male	244.2	0.06	3.32E-05	70.01
F8775-1	MiSeq	Pregnant	NA	62.71	0.09	2.53E-05	53.15
C24778	HiSeq	Pregnant	Male	201.6	0.07	2.55E-05	40.85
F11091.1	HiSeq	Pregnant	NA	928.38	0.04	2.47E-05	39.53
F9531-1-Maria	MiSeq	Pregnant	Male	97.65	0.05	1.77E-05	36.94
F8614-1	MiSeq	Pregnant	Male	66.12	0.04	1.27E-05	26.28
F8659-1	MiSeq	Pregnant	Female	233.1	0.03	6.05E-06	12.09
F10951-1	HiSeq	Pregnant	Female	374.7	0.07	2.09E-06	2.38
F9198-1	MiSeq	Pregnant	Female	222	0.08	1.37E-06	2.11
F10114_fetal5	MiSeq	Mock	Female	271	0.04	1.14E-06	1.62
F10114_fetal4	MiSeq	Mock	Female	269.8	0.04	1.14E-06	1.62
F10114_fetal10	MiSeq	Mock	Female	275.8	0.12	1.12E-06	1.58
F10114_fetal15	MiSeq	Mock	Female	279.3	0.17	1.10E-06	1.53
F10114_fetal20	MiSeq	Mock	Female	281.8	0.22	1.09E-06	1.51
F10177-2	MiSeq	Pregnant	Female	194.2	0.22	8.26E-07	0.95
F9176-1	MiSeq	Pregnant	Female	173.6	0.14	7.83E-07	0.86
F10609-1	HiSeq	Pregnant	ukn	556.3	0.2	1.13E-06	0.80
F11077.1	HiSeq	Pregnant	Female	641.7	0.07	1.11E-06	0.77

F9914-1	MiSeq	Pregnant	Female	219.1	0.12	7.11E-07	0.70
C26139	MiSeq	Pregnant	Female	250.2	0.15	6.33E-07	0.54
F10063-1	MiSeq	Pregnant	Female	253.2	0.08	6.29E-07	0.53
F10117-1	MiSeq	Pregnant	Female	233.6	0.21	6.19E-07	0.51
F9530-1	MiSeq	Pregnant	Female	226.9	0.02	5.78E-07	0.42
F8660-1	MiSeq	Pregnant	Female	294.3	0.03	5.41E-07	0.34
F10351-1	MiSeq	Pregnant	Female	283.3	0.3	2.83E-07	-0.21
F10764-1	HiSeq	Pregnant	Female	463.8	0.07	4.23E-07	-0.36
F10795-1	HiSeq	Pregnant	Female	541.4	0.08	2.47E-07	-0.65
F11247.1	HiSeq	Pregnant	Female	620.6	0.13	2.42E-07	-0.66
F9966_fetal5	MiSeq	Mock	Female	257.8	0.08	0.00E+00	-0.81
F9966_fetal4	MiSeq	Mock	Female	254.6	0.07	0.00E+00	-0.81
F9966_fetal20	MiSeq	Mock	Female	296.6	0.28	0.00E+00	-0.81
F9966_fetal15	MiSeq	Mock	Female	285.4	0.22	0.00E+00	-0.81
F9966_fetal10	MiSeq	Mock	Female	272.5	0.16	0.00E+00	-0.81
F9614-1	MiSeq	Pregnant	Female	39.99	0.08	0.00E+00	-0.81
F9219_fetal5	MiSeq	Mock	Female	166.1	0.11	0.00E+00	-0.81
F9219_fetal4	MiSeq	Mock	Female	163.4	0.09	0.00E+00	-0.81
F9219_fetal20	MiSeq	Mock	Female	195.8	0.3	0.00E+00	-0.81
F9219_fetal15	MiSeq	Mock	Female	187.6	0.23	0.00E+00	-0.81
F9219_fetal10	MiSeq	Mock	Female	177.8	0.18	0.00E+00	-0.81
F9151-1	MiSeq	Pregnant	Female	80.04	0.1	0.00E+00	-0.81
F9051-1	MiSeq	Pregnant	Female	271.6	0.15	0.00E+00	-0.81
F8850-1	MiSeq	Pregnant	Female	203.9	0.11	0.00E+00	-0.81
F8762-1	MiSeq	Pregnant	Female	147.2	0.09	0.00E+00	-0.81
F8642-1	MiSeq	Pregnant	Female	162.1	0.06	0.00E+00	-0.81
F10157-1	MiSeq	Pregnant	Female	242.4	0.16	0.00E+00	-0.81
C18177	MiSeq	Pregnant	Female	269.9	0.14	0.00E	-0.81

						+00	
F10801_fetal5	HiSeq	Mock	Female	335.3	0.09	0.00E+00	-1.06
F10801_fetal4	HiSeq	Mock	Female	330.4	0.07	0.00E+00	-1.06
F10801_fetal20	HiSeq	Mock	Female	396.6	0.27	0.00E+00	-1.06
F10801_fetal15	HiSeq	Mock	Female	378.3	0.22	0.00E+00	-1.06
F10801_fetal10	HiSeq	Mock	Female	357.9	0.16	0.00E+00	-1.06

*Sample name_fetalXX, where XX indicates the original fastq proportion admixture (in percentage)

Supporting Table 7: Z-score of chromosomes 18 and 21 for detection of trisomy.

Sample*	Sequencing	Classification	Disease**	Mean Bam Coverage (X)	Fitted FF	Chr18 Z-score	Chr21 Z-score
F9966_fetal20	MiSeq	Mock	T21	296.6	0.28	-0.46	13.36
F9966_fetal15	MiSeq	Mock	T21	285.4	0.22	-0.38	10.58
F10114_fetal20	MiSeq	Mock	T21	281.8	0.22	-0.43	9.66
F9966_fetal10	MiSeq	Mock	T21	272.5	0.16	-0.23	7.71
F10114_fetal15	MiSeq	Mock	T21	279.3	0.17	-0.42	7.69
F10114_fetal10	MiSeq	Mock	T21	275.8	0.12	-0.33	5.37
F9966_fetal5	MiSeq	Mock	T21	257.8	0.08	-0.04	3.93
F9966_fetal4	MiSeq	Mock	T21	254.6	0.07	0.02	3.05
F10114_fetal5	MiSeq	Mock	T21	271	0.04	-0.27	2.82
F10114_fetal4	MiSeq	Mock	T21	269.8	0.04	-0.26	2.24
F10801_fetal20	HiSeq	Mock	T18	396.6	0.27	4.04	1.23
F10801_fetal4	HiSeq	Mock	T18	330.4	0.07	0.96	1.14
F10801_fetal5	HiSeq	Mock	T18	335.3	0.09	1.24	1.05
F10801_fetal15	HiSeq	Mock	T18	378.3	0.22	3.20	1.03
F10801_fetal10	HiSeq	Mock	T18	357.9	0.16	2.26	0.95
F10351-1	MiSeq	Pregnant	Normal	283.3	0.3	-1.53	2.81
F10157-1	MiSeq	Pregnant	Normal	242.4	0.16	-0.24	2.05
C26139	MiSeq	Pregnant	Normal	250.2	0.15	-1.17	1.86
C26449_fetal20	MiSeq	Mock	Normal	314.7	0.25	-1.59	1.63
F10855-1	HiSeq	Pregnant	Normal	418.3	0.21	0.09	1.57
C26449_fetal15	MiSeq	Mock	Normal	305	0.19	-1.59	1.40
F8660-1	MiSeq	Pregnant	Normal	294.3	0.03	-1.02	1.35
C26449_fetal10	MiSeq	Mock	Normal	293.8	0.14	-1.63	1.28
F10795-1	HiSeq	Pregnant	Normal	541.4	0.08	-1.30	1.26
F10063-1	MiSeq	Pregnant	Normal	253.2	0.08	-0.54	1.24
F9892-1	MiSeq	Pregnant	Normal	244.2	0.06	-0.33	1.23
C26449_fetal5	MiSeq	Mock	Normal	281.2	0.08	-1.59	1.09
F8762-1	MiSeq	Pregnant	Normal	147.2	0.09	1.22	1.00
C26449_fetal4	MiSeq	Mock	Normal	278.5	0.06	-1.58	0.94
C18177	MiSeq	Pregnant	Normal	269.9	0.14	-1.08	0.94
C24778	HiSeq	Pregnant	Normal	201.6	0.07	0.82	0.88
F10117-1	MiSeq	Pregnant	Normal	233.6	0.21	0.14	0.71
F9531-1	MiSeq	Pregnant	Normal	97.65	0.05	-1.29	0.69
F8659-1	MiSeq	Pregnant	Normal	233.1	0.03	0.11	0.68
F9051-1	MiSeq	Pregnant	Normal	271.6	0.15	-0.57	0.43
P2182	MiSeq	Pregnant	Normal	208.6	0.17	0.07	0.41
F10113-1	MiSeq	Pregnant	Normal	212.5	0.27	-0.36	0.23
F9940-1	MiSeq	Pregnant	Normal	141.2	0.17	2.22	0.18
F10138-1	MiSeq	Pregnant	Normal	210.2	0.18	0.12	0.00
F9269-1	MiSeq	Pregnant	Normal	186.9	0.05	-0.18	-0.02
F8614-1	MiSeq	Pregnant	Normal	66.12	0.04	1.59	-0.08

F9198-1	MiSeq	Pregnant	Normal	222	0.08	-0.33	-0.11
C20598_fetal10	MiSeq	Mock	Normal	185.2	0.27	0.21	-0.13
F10764-1	HiSeq	Pregnant	Normal	463.8	0.07	0.09	-0.15
P2183	HiSeq	Pregnant	Normal	357.4	0.14	0.54	-0.17
C23944_fetal4	MiSeq	Mock	Normal	93.79	0.12	1.15	-0.20
C25147_fetal20	MiSeq	Mock	Normal	220.7	0.29	0.36	-0.26
C20598_fetal20	MiSeq	Mock	Normal	216.3	0.39	-0.10	-0.27
C20598_fetal15	MiSeq	Mock	Normal	201.7	0.34	0.01	-0.33
F9176-1	MiSeq	Pregnant	Normal	173.6	0.14	0.30	-0.44
C25147_fetal15	MiSeq	Mock	Normal	211	0.25	0.43	-0.44
F10774-1	HiSeq	Pregnant	Normal	592	0.11	0.89	-0.50
F8642-1	MiSeq	Pregnant	Normal	162.1	0.06	-0.39	-0.53
C25147_fetal10	MiSeq	Mock	Normal	200	0.18	0.59	-0.55
C20598_fetal5	MiSeq	Mock	Normal	166.4	0.14	0.47	-0.55
C23944_fetal20	MiSeq	Mock	Normal	111.3	0.29	0.95	-0.58
F9614-1	MiSeq	Pregnant	Normal	39.99	0.08	0.68	-0.60
C23944_fetal15	MiSeq	Mock	Normal	107.6	0.26	0.92	-0.63
C20598_fetal4	MiSeq	Mock	Normal	162.4	0.12	0.57	-0.64
F9530-1	MiSeq	Pregnant	Normal	226.9	0.02	-0.46	-0.64
F9151-1	MiSeq	Pregnant	Normal	80.04	0.1	-0.76	-0.73
F9219_fetal4	MiSeq	Mock	Normal	163.4	0.09	0.00	-0.74
F9891-1	MiSeq	Pregnant	Normal	269.7	0.03	-0.44	-0.77
F11247.1	HiSeq	Pregnant	Normal	620.6	0.13	1.40	-0.80
F8850-1	MiSeq	Pregnant	Normal	203.9	0.11	0.02	-0.83
F9219_fetal10	MiSeq	Mock	Normal	177.8	0.18	0.93	-0.85
F9219_fetal5	MiSeq	Mock	Normal	166.1	0.11	0.21	-0.87
C23944_fetal5	MiSeq	Mock	Normal	106	0.23	1.01	-0.90
F9219_fetal15	MiSeq	Mock	Normal	187.6	0.23	1.50	-0.92
C25147_fetal5	MiSeq	Mock	Normal	187.5	0.1	0.73	-0.95
F10775-1	HiSeq	Pregnant	Normal	356.5	0.12	1.10	-1.00
F10177-2	MiSeq	Pregnant	Normal	194.2	0.22	0.96	-1.01
C25147_fetal4	MiSeq	Mock	Normal	184.8	0.09	0.77	-1.03
F9219_fetal20	MiSeq	Mock	Normal	195.8	0.3	1.98	-1.04
C23944_fetal10	MiSeq	Mock	Normal	120.7	0.32	0.92	-1.18
C26729	MiSeq	Pregnant	Normal	105.6	0.07	3.93	-1.18
F11077.1	HiSeq	Pregnant	Normal	641.7	0.07	1.24	-1.19
F10609-1	HiSeq	Pregnant	Normal	556.3	0.2	-0.02	-1.23
F10951-1	HiSeq	Pregnant	Normal	374.7	0.07	0.75	-1.39
F8998-1	MiSeq	Pregnant	Normal	161.1	0.14	1.09	-2.16
F10396-1	MiSeq	Pregnant	NA	267.9	0.19	-4.03	6.29
F9914-1	MiSeq	Pregnant	NA	219.1	0.12	-1.01	1.53
F8626-1	MiSeq	Pregnant	NA	132.2	0.26	-0.87	0.49
F11091.1	HiSeq	Pregnant	NA	928.38	0.04	1.00	0.21
F8775-1	MiSeq	Pregnant	NA	62.71	0.09	1.75	-1.27

*Sample name_fetalXX, where XX indicates the original fastq proportion admixture (in percentage).
**Normal: not affected by T21 or T18; NA: Information not available.

Chapter 3

General Discussion and Conclusions

Implementing these novel techniques in Brazil is of great importance, so we can be competitive at the research world scenario as well as open the opportunity to develop better cost-effective tests to be offered to our population. We are located at the Human Genome and Stem Cell Research Center (HUG-CELL), a reference center of genetic diseases in Brazil at Universidade de São Paulo. An important feature of this work is that it is a scalable test for monogenic diseases. This means that once the gene panel is ready and the workflow established, it is possible to test virtually any disease there are genes on the panel for, and with the advent of new technologies and analyses, we have a great potential of developing new noninvasive tests (without need of any parental genotype) for other diseases with variants difficult to be detected currently, such as for recessive disorders (indicated by slight increase in AF, as discussed previously).

Also, the ability to incorporate new techniques enables translating them into new fields, increasing the impact of our research within the country, for instance permitting us to apply the knowledge obtained in this work in areas such as cancer and transplant rejection monitoring, already performed abroad. In addition, it diminishes our dependence on sending genetic tests abroad, which mean we have a possible additional source of revenue for our country, and we are able to offer that for our patients at university extension-basis, as well as transfer the technology for companies, like spin-offs from Universities.

The methodology used here can be more easily applied to allograft rejection than for cancer monitoring in the short term for 2 main reasons: 1) detection of allograft rejection is less complex than cancer, since its goal is to detect increase of the cdDNA fraction (which is performed similarly as we did for fetal fraction), while for

cancer the knowledge of the etiology is often needed, and the cancer genome may change during disease progression, due to metastasis and/or therapy resistance acquire; 2) For cancer there is a need for a higher coverage (since cdDNA fraction can be as low as 0.01%), and this amount of sequencing is not financially feasible with the current technology applied by us.

The main novelty of this work was to be able to integrate several analyses (fetal fraction, fetal sex, trisomies and monogenic diseases), without needing parental genotype. In this scenario, having the fetal fraction estimated is an important quality control of the test, since we are able to differentiate true negatives from samples with a low FF (which is a limitating factor for detection). Besides, having the FF information helps to detect monogenic disease variants, since we can detect variants within the expected AF range. However, most of groups that do perform the FF estimation use genotype information from at least the mother, which increases cost and turnaround laboratory time. In this work we describe a new model for estimate fetal fraction using only plasma sequencing data, overcoming these issues.

The development of this estimation model was performed in collaboration with the research group of Dr. Iwijn De Vlaminck, at Cornell University, a reference group at analyzing variants from cfDNA sequencing data. The group of Dr. De Vlaminck utilizes the sequencing data of cfDNA from plasma of organ receptors to detect rejection of the transplanted organ. Setting these kind of collaborations is very important for at least 2 reasons: 1) increasing the quality of the research being performed in Brazil, therefore being competitive 2) having contact with new research lines, being able to translate the knowledge and mindset to our environment.

As mentioned, HUG-CELL is a reference center in our country, and has a multidisciplinary team to attend hundreds of patients and their families. Translational projects like the one described here have the potential to bring better care for our patients and offer exams not available for them before. Thus, the purpose of this study is in line with the objective of the Ministry of Science Technology and Innovation Ministry and the National Council for Scientific and Technological Development (CNPq) designed to promote science, technology and innovation in the country.

RESUMO

Desde 2011 a área de diagnóstico pré-natal sofreu uma revolução com a introdução de teste pré-natal não-invasivo (NIPT) de doenças genéticas, que se baseia na análise de DNA fetal livre de células presente no plasma materno. Apesar de estar disponível no Brasil, nós dependemos em terceirizar a tecnologia desenvolvida no exterior, ou o teste em si. Sendo assim, nosso objetivo foi desenvolver e implementar um teste NIPT abrangente utilizando sequenciamento de nova geração de alta cobertura *targeted* para: 1) estimar fração fetal; 2) determinar sexo fetal; 3) detectar trissomia; 4) detectar doença monogênica.

Nós desenvolvemos um modelo robusto e preciso (r^2= 0.994, p-value < 2.2e-16) para estimativa de fração fetal baseado na distribuição da fração alélica de SNPs. Nós utilizamos Z-score para determinação de sexo fetal (100% de precisão) e detecção de trissomia dos cromossomos 21 (T21) e 18 (T18), atingindo uma sensibilidade de 100% (95% IC: 63.06% - 100.00%) e especificidade de 98.53% (95% IC: 92.08% - 99.96%) para T21, e 40% (95% IC: 5.27% - 85.34%) e 98.59% (95% IC: 92.40% - 99.96%) para T18. Para detecção de doença monogênica (displasia esquelética) nós realizamos análise de variante, com uma taxa de detecção de 71% (5/7). Até onde sabemos, este é o primeiro trabalho a integrar todas as análises em um único teste, e a realizar detecção de doença monogênica sem utilizar genótipo parental.

Nós mostramos neste trabalho que é possível implementar essas técnicas no nosso país, utilizando recursos disponíveis e/ou desenvolvendo colaborações com grupos referência internacionais. Isto demonstra o potencial de desenvolver tecnologias internas, e aplicá-las à outras áreas não-invasivas, como diagnóstico e monitoramento de câncer e rejeição de transplante.

ABSTRACT

Since 2011 the prenatal diagnosis field has undergone a revolution with the introduction of a noninvasive prenatal test (NIPT) for genetic diseases relying on analysis of fetal cell-free DNA present in maternal plasma. Although available in Brazil, we rely on outsourcing the technology developed abroad or the test itself. Therefore, our objective was to develop and implement a comprehensive NIPT using high-coverage targeted next-generation sequencing to: 1) estimate fetal fraction; 2) determine fetal sex; 3) detect trisomy; 4) detect monogenic disease.

We developed a robust and accurate model (r^2= 0.994, p-value < 2.2e-16) for fetal fraction estimation based on distribution of SNP minor allele fraction (MAF). We used Z-score for fetal sex determination (100% accuracy) and trisomy detection of chromosomes 21 (T21) and 18 (T18), achieving a sensitivity of 100% (95% CI: 63.06% - 100.00%) and a specificity of 98.53% (95% CI: 92.08% - 99.96%) for T21, and 40% (95% CI: 5.27% - 85.34%) and 98.59% (95% CI: 92.40% - 99.96%) for T18. For monogenic disease detection (skeletal dysplasia) we performed variant analysis, with 71% (5/7) of detection rate. To our knowledge, this is the first work to integrate all analysis in one single test, and to perform monogenic disease detection without using parental genotype.

We showed in this work that it is possible to implement such techniques in our country, using available resources and/or engaging in collaboration with reference research groups abroad. It shows the potential of developing internal technologies, and applying it to other noninvasive fields, such as cancer and organ rejection diagnostic and monitoring.

REFERENCES

AKOLEKAR, R. et al. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. **Ultrasound in Obstetrics & Gynecology**, v. 45, n. 1, p. 16–26, jan. 2015.

AMBERGER, J. S. et al. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. **Nucleic Acids Research**, v. 43, n. Database issue, p. D789-98, jan. 2015.

ARCASOY, S. M. et al. Pathologic interpretation of transbronchial biopsy for acute rejection of lung allograft is highly variable. **American Journal of Transplantation**, v. 11, n. 2, p. 320–328, 2011.

ASHOOR, G. et al. Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18. **American Journal of Obstetrics and Gynecology**, v. 206, n. 4, p. 322.e1-322.e5, 2012a.

ASHOOR, G. et al. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: effect of maternal and fetal factors. **Fetal diagnosis and therapy**, v. 31, n. 4, p. 237–43, jan. 2012b.

BECK, J. et al. Digital droplet PCR for rapid quantification of donor DNA in the circulation of transplant recipients as a potential universal biomarker of graft injury. **Clinical Chemistry**, v. 59, n. 12, p. 1732–1741, 2013.

BECK, J. et al. Donor-Derived Cell-Free DNA Is a Novel Universal Biomarker for Allograft Rejection in Solid Organ Transplantation. **Transplantation Proceedings**, v. 47, n. 8, p. 2400–2403, out. 2015.

BENN, P. et al. Position statement from the Aneuploidy Screening Committee on behalf of the Board of the International Society for Prenatal Diagnosis. **Prenatal Diagnosis**, v. 33, n. 7, p. 622–629, jul. 2013.

BIANCHI, D. W. et al. DNA sequencing versus standard prenatal aneuploidy screening. **The New England journal of medicine**, v. 370, n. 9, p. 799–808, 27 fev. 2014.

BISCHOFF, F. Z.; LEWIS, D. E.; SIMPSON, J. L. Cell-free fetal DNA in maternal blood: Kinetics, source and structure. **Human Reproduction Update**, v. 11, n. 1, p. 59–67, 2005.

BLOOM, R. D. et al. Cell-Free DNA and Active Rejection in Kidney Allografts. **Journal of the American Society of Nephrology**, p. ASN.2016091034, 2017.

BUSTAMANTE-ARAGONES, A. et al. Prenatal diagnosis of Huntington disease in maternal plasma: direct and indirect study. **European Journal of Neurology**, v. 15, n. 12, p. 1338–1344, dez. 2008.

BUSTAMANTE-ARAGONÉS, A. et al. Non-invasive prenatal diagnosis of single-gene disorders from maternal blood. **Gene**, v. 504, n. 1, p. 144–9, 1 ago. 2012.

CAUGHEY, A. B.; HOPKINS, L. M.; NORTON, M. E. Chorionic Villus Sampling Compared With Amniocentesis and the Difference in the Rate of Pregnancy Loss. **Obstet Gynecol**, v.

108, n. 3, p. 612–616, 2006.

CHITTY, L. S. et al. New aids for the non-invasive prenatal diagnosis of achondroplasia: dysmorphic features, charts of fetal size and molecular confirmation using cell-free fetal DNA in maternal plasma. **Ultrasound in Obstetrics & Gynecology**, v. 37, n. 3, p. 283–289, mar. 2011.

CHITTY, L. S. et al. Non-invasive prenatal diagnosis of achondroplasia and thanatophoric dysplasia: next-generation sequencing allows for a safer, more accurate, and comprehensive approach. **Prenatal Diagnosis**, v. 35, n. 7, p. 656–662, jul. 2015.

CHIU, R. W. K. et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 51, p. 20458–63, 23 dez. 2008.

CHIU, R. W. K. et al. Maternal Plasma DNA Analysis with Massively Parallel Sequencing by Ligation for Noninvasive Prenatal Diagnosis of Trisomy 21. **Clinical Chemistry**, v. 56, n. 3, p. 459–463, 1 mar. 2010.

CHIU, R. W. K. et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. **BMJ**, v. 342, n. jan11 1, p. c7401–c7401, 11 jan. 2011.

CHU, T. et al. Statistical considerations for digital approaches to non-invasive fetal genotyping. **Bioinformatics (Oxford, England)**, v. 26, n. 22, p. 2863–6, 15 nov. 2010.

COLVIN-ADAMS, M. et al. OPTN/SRTR 2013 Annual Data Report: Heart. **American Journal of Transplantation**, v. 15, n. S2, p. 1–28, jan. 2015.

CROWLEY, E. et al. Liquid biopsy: monitoring cancer-genetics in the blood. **Nature reviews. Clinical oncology**, v. 10, n. 8, p. 472–84, 2013.

DAN, S. et al. Non-invasive prenatal diagnosis of lethal skeletal dysplasia by targeted capture sequencing of maternal plasma. **PLoS ONE**, v. 11, n. 7, p. 1–16, 2016.

DAVIES, H. et al. Mutations of the BRAF gene in human cancer. **Nature**, v. 417, n. 6892, p. 949–954, 27 jun. 2002.

DE VLAMINCK, I. et al. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. **Science Translational Medicine**, v. 6, n. 241, p. 241ra77-241ra77, 2014.

DE VLAMINCK, I. et al. Noninvasive monitoring of infection and rejection after lung transplantation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 43, p. 13336–41, 2015.

DIEHL, F. et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. **Proc Natl Acad Sci U S A**, v. 102, n. 45, p. 16368–16373, 2005.

DOUILLARD, J.-Y. et al. Gefitinib Treatment in EGFR Mutated Caucasian NSCLC: Circulating-Free Tumor DNA as a Surrogate for Determination of EGFR Status. **Journal of Thoracic Oncology**, v. 9, n. 9, p. 1345–1353, set. 2014.

EHRICH, M. et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. **American Journal of Obstetrics and Gynecology**, v. 204, n. 3, p. 205.e1-205.e11, mar. 2011.

FAAS, B. H. et al. Detection of fetal RhD - specific sequences in maternal plasma. **Lancet**, v. 352, p. 1196, 1998.

FAN, H. C. et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. **Proceedings of the National Academy of Sciences**, v. 105, n. 42, p. 16266–16271, 21 out. 2008.

FAN, H. C. et al. Non-invasive prenatal measurement of the fetal genome. **Nature**, v. 487, n. 7407, p. 320–4, 19 jul. 2012.

FAN, H. C.; QUAKE, S. R. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. **PloS one**, v. 5, n. 5, p. e10439, jan. 2010.

FLORI, E. et al. Circulating cell-free fetal DNA in maternal serum appears to originate from cyto- and syncytio-trophoblastic cells. Case report. **Human Reproduction**, v. 19, n. 3, p. 723–724, 2004.

FORSHEW, T. et al. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. **Science Translational Medicine**, v. 4, n. 136, p. 136ra68-136ra68, 30 maio 2012.

GAHAN, P. B.; ANKER, P.; STROUN, M. Metabolic DNA as the Origin of Spontaneously Released DNA? **Annals of the New York Academy of Sciences**, v. 1137, n. 1, p. 7–17, ago. 2008.

GARRAWAY, L. A.; LANDER, E. S. Lessons from the Cancer Genome. **Cell**, v. 153, n. 1, p. 17–37, mar. 2013.

GIELIS, E. M. et al. Cell-Free DNA: An Upcoming Biomarker in Transplantation. **American Journal of Transplantation**, v. 15, n. 10, p. 2541–2551, 2015.

GIL, M. M. et al. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. **Ultrasound in Obstetrics & Gynecology**, v. 45, n. 3, p. 249–266, mar. 2015.

GONZÁLEZ-GONZÁLEZ, M. C. et al. Prenatal detection of a cystic fibrosis mutation in fetal DNA from maternal plasma. **Prenatal Diagnosis**, v. 22, n. 10, p. 946–948, 2002.

GRSKOVIC, M. et al. Validation of a Clinical-Grade Assay to Measure Donor-Derived Cell-Free DNA in Solid Organ Transplant Recipients. **The Journal of Molecular Diagnostics**, v. 18, n. 6, p. 890–902, nov. 2016.

HAN, X.; WANG, J.; SUN, Y. Circulating Tumor DNA as Biomarkers for Cancer Detection. **Genomics, Proteomics & Bioinformatics**, v. 15, n. 2, p. 59–72, abr. 2017.

HIDESTRAND, M. et al. Highly Sensitive Noninvasive Cardiac Transplant Rejection Monitoring Using Targeted Quantification of Donor-Specific Cell-Free Deoxyribonucleic Acid. **Journal of the American College of Cardiology**, v. 63, n. 12, p. 1224–1226, abr. 2014.

HUANG, Z. et al. Quantitation of Plasma Circulating DNA Using Quantitative PCR for the Detection of Hepatocellular Carcinoma. **Pathology & Oncology Research**, v. 18, n. 2, p. 271–276, 21 abr. 2012.

HYDE, K. J.; SCHUST, D. J. Genetic Considerations in Recurrent Pregnancy Loss. **Cold Spring Harbor Perspectives in Medicine**, v. 5, n. 3, p. a023119–a023119, 1 mar. 2015.

JAHR, S. et al. DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and

Evidence for Their Origin from Apoptotic and Necrotic Cells DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necr. **Cancer Research**, v. 61, n. 4, p. 1659–1665, 2001.

JIANG, P. et al. FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. **Bioinformatics**, v. 28, n. 22, p. 2883–2890, 15 nov. 2012.

KINDE, I. et al. Detection and quantification of rare mutations with massively parallel sequencing. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 23, p. 9530–5, 2011.

KINNINGS, S. L. et al. Factors affecting levels of circulating cell-free fetal DNA in maternal plasma and their implications for noninvasive prenatal testing. **Prenatal Diagnosis**, p. n/a-n/a, 2015.

KITZMAN, J. O. et al. Noninvasive whole-genome sequencing of a human fetus. **Science translational medicine**, v. 4, n. 137, p. 137ra76, 6 jun. 2012.

KOUMBARIS, G. et al. Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal Sex. **Clinical Chemistry**, v. 62, n. 6, p. 848–855, 1 jun. 2016.

KRISHNAMURTHY, N. et al. Liquid Biopsies for Cancer: Coming to a Patient near You. **Journal of Clinical Medicine**, v. 6, n. 1, p. 3, 2017.

LAM, K.-W. G. et al. Noninvasive Prenatal Diagnosis of Monogenic Diseases by Targeted Massively Parallel Sequencing of Maternal Plasma: Application to -Thalassemia. **Clinical Chemistry**, v. 58, n. 10, p. 1467–1475, 1 out. 2012.

LAU, T. K. et al. Non-invasive prenatal testing for fetal chromosomal abnormalities by low-coverage whole-genome sequencing of maternal plasma DNA: Review of 1982 consecutive cases in a single center. **Ultrasound in Obstetrics and Gynecology**, v. 43, n. 3, p. 254–264, 2014.

LEARY, R. J. et al. Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing. **Science Translational Medicine**, v. 2, n. 20, p. 20ra14-20ra14, 24 fev. 2010.

LENCH, N. et al. The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: Challenges and progress made. **Prenatal Diagnosis**, v. 33, n. 6, p. 555–562, 2013.

LEON, S. A. et al. Free DNA in the serum of cancer patients and the effect of therapy. **Cancer research**, v. 37, n. 3, p. 646–50, mar. 1977.

LIAO, G. J. W. et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. **Clinical Chemistry**, v. 57, n. 1, p. 92–101, 2011.

LICHTENSTEIN, A. V. et al. Circulating Nucleic Acids and Apoptosis. **Annals of the New York Academy of Sciences**, v. 945, n. 1, p. 239–249, 2006.

LO, Y. D. et al. Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. **The Lancet**, v. 351, n. 9112, p. 1329–1330, maio 1998a.

LO, Y. M. et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. **American journal of human genetics**, v.

62, n. 4, p. 768–75, abr. 1998b.

LO, Y. M. D. et al. Early report Presence of fetal DNA in maternal plasma and serum. **The Lancet**, v. 350, p. 485–487, 1997.

LO, Y. M. D. et al. Prenatal Diagnosis of Fetal RhD Status by Molecular Analysis of Maternal Plasma. **New England Journal of Medicine**, v. 339, n. 24, p. 1734–1738, 1998c.

LO, Y. M. D. et al. Rapid Clearance of Fetal DNA from Maternal Plasma. **The American Journal of Human Genetics**, v. 64, n. 1, p. 218–224, jan. 1999.

LO, Y. M. D. et al. Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. **Science Translational Medicine**, v. 2, n. 61, p. 61ra91-61ra91, 8 dez. 2010.

LUN, F. M. F. et al. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. **Clinical Chemistry**, v. 54, n. 10, p. 1664–1672, 2008a.

LUN, F. M. F. et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 50, p. 19920–5, 16 dez. 2008b.

MA, D. et al. Haplotype-based approach for noninvasive prenatal diagnosis of congenital adrenal hyperplasia by maternal plasma DNA sequencing. **Gene**, v. 544, n. 2, p. 252–258, 2014.

MARBOE, C. C. et al. Nodular endocardial infiltrates (Quilty lesions) cause significant variability in diagnosis of ISHLT Grade 2 and 3A rejection in cardiac allograft recipients. **Journal of Heart and Lung Transplantation**, v. 24, n. 7 SUPPL., p. 219–226, 2005.

MOORE, K. L.; PERSAUD, T. V. N.; The Developing Human: Clinically Oriented Embryology, 6ª ed.; Philadelphia: Saunders/ Elsevier, 1998.

MUJEZINOVIC, F.; ALFIREVIC, Z. Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. **Obstetrics and gynecology**, v. 110, n. 3, p. 687–694, 2007.

NORTON, M. E. et al. Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. **American Journal of Obstetrics and Gynecology**, v. 207, n. 2, p. 137.e1-137.e8, ago. 2012.

NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. Thompson & Thompson genetics in medicine. 7a ed. Philadelphia: Saunders/ Elsevier, 2007.

PALOMAKI, G. E. et al. DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study. **Genetics in Medicine**, v. 13, n. 11, p. 913–920, nov. 2011.

PARKS, M. et al. Non-invasive prenatal diagnosis of Duchenne and Becker muscular dystrophies by relative haplotype dosage †. p. 312–320, 2016.

PATEL, K. M.; TSUI, D. W. Y. The translational potential of circulating tumour DNA in oncology. **Clinical Biochemistry**, v. 48, n. 15, p. 957–961, out. 2015.

POON, L. C. Y. et al. Maternal plasma cell-free fetal and maternal DNA at 11-13 weeks' gestation: relation to fetal and maternal characteristics and pregnancy outcomes. **Fetal diagnosis and therapy**, v. 33, n. 4, p. 215–23, jan. 2013.

RAI, R.; REGAN, L. Recurrent miscarriage. **Lancet**, v. 368, p. 601–611, 2006.

RIJNDERS, R. J. et al. Fetal sex determination from maternal plasma in pregnancies at risk for congenital adrenal hyperplasia. **Obstetrics and gynecology**, v. 98, n. 3, p. 374–8, 2001.

ROTHÉ, F. et al. Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. **Annals of Oncology**, v. 25, n. 10, p. 1959–1965, 2014.

RUSSO, M. L.; BLAKEMORE, K. J. A historical and practical review of first trimester aneuploidy screening. **Seminars in Fetal and Neonatal Medicine**, v. 19, n. 3, p. 183–187, 2014.

RYAN, A. et al. Validation of an Enhanced Version of a Single-Nucleotide Polymorphism-Based Noninvasive Prenatal Test for Detection of Fetal Aneuploidies. **Fetal diagnosis and therapy**, 2016.

SANSEVERINO. Investigação genética pré-natal. In: Rotinas em Obstetrícia. 5ª ed. Porto Alegre: Artmed, 2006.

SARAIVA, F. et al. Complications of endomyocardial biopsy in heart transplant patients: A retrospective study of 2117 consecutive procedures. **Transplantation Proceedings**, v. 43, n. 5, p. 1908–1912, 2011.

SCHÜT et al. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: A prospective, observational, multicenter cohort study. p. 1–19, 2017.

SCHWARZENBACH, H.; HOON, D. S. . B.; PANTEL, K. Cell-free nucleic acids as biomarkers in cancer patients. **Nature Reviews Cancer**, v. 11, n. 6, p. 426–437, 12 jun. 2011.

SEHNERT, A. J. et al. Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. **Clinical chemistry**, v. 57, n. 7, p. 1042–9, jul. 2011.

SHAMSHIRSAZ, A. A.; BENN PETER, P.; EGAN, J. F. X. The role of second-trimester serum screening in the post-first-trimester screening era. **Clinics in Laboratory Medicine**, v. 30, n. 3, p. 667–676, 2010.

SIGDEL, T. K. et al. A Rapid Noninvasive Assay for the Detection of Renal Transplant Injury. **Transplantation Journal**, v. 96, n. 1, p. 97–101, jul. 2013.

SKRZYPEK, H.; HUI, L. Noninvasive prenatal testing for fetal aneuploidy and single gene disorders. **Best Practice & Research Clinical Obstetrics & Gynaecology**, p. 1–13, 28 fev. 2017.

SMID, M. et al. No evidence of fetal DNA persistence in maternal plasma after pregnancy. **Human genetics**, v. 112, n. 5–6, p. 617–618, 2003.

SNYDER, T. M. et al. Universal noninvasive detection of solid organ transplant rejection. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 15, p. 6229–34, 2011.

SPARKS, A. B. et al. Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. **Prenatal Diagnosis**, v. 32, n. 1, p. 3–9, jan. 2012a.

SPARKS, A. B. et al. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. **American Journal of Obstetrics and Gynecology**, v. 206, n. 4, p. 319.e1-319.e9, abr. 2012b.

STRATTON, M.; CAMPBELL, P.; FUTREAL, P. The cancer genome. **Nature**, v. 458, n. 7239, p. 719–724, 2009.

STROUN, M. et al. Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients. **Oncology**, v. 46, n. 5, p. 318–322, 26 jun. 1989.

SUZUKI, N. et al. Characterization of circulating DNA in healthy human plasma. **Clinica Chimica Acta**, v. 387, n. 1–2, p. 55–58, 2008.

TAKOUDES, T.; HAMAR, B. Performance of non-invasive prenatal testing when fetal cell-free DNA is absent. **Ultrasound in Obstetrics and Gynecology**, v. 45, n. 1, p. 112, 2015.

THIERRY, A. R. et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. **Nucleic Acids Research**, v. 38, n. 18, p. 6159–6175, 2010.

TOMASETTI, C.; LI, L.; VOGELSTEIN, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. **Science**, v. 355, n. 6331, p. 1330–1334, 24 mar. 2017.

UCHIDA, J. et al. Diagnostic Accuracy of Noninvasive Genotyping of EGFR in Lung Cancer Patients by Deep Sequencing of Plasma Cell-Free DNA. **Clinical Chemistry**, v. 61, n. 9, p. 1191–1196, 1 set. 2015.

VALAPOUR, M. et al. OPTN/SRTR 2013 Annual Data Report: Lung. **American Journal of Transplantation**, v. 15, n. S2, p. 1–28, jan. 2015.

VAN DER VAART, M.; PRETORIUS, P. J. Circulating DNA. Its origin and fluctuation. **Annals of the New York Academy of Sciences**, v. 1137, n. 0, p. 18–26, ago. 2008.

VERHOEF, T. I. et al. Non-invasive prenatal diagnosis (NIPD) for single gene disorders : cost analysis of NIPD and invasive testing pathways. p. 1–7, 2016.

VOGELSTEIN, B. et al. Cancer Genome Lanscapes. **Science**, v. 339, n. 6127, p. 1546–1558, 2013.

WANG, E. et al. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. **Prenatal Diagnosis**, v. 33, n. 7, p. 1–5, jul. 2013.

WOODWARD, R. N. et al. Proficiency testing for and surveillance of heart transplant health using donor-derived cell-free DNA. **Journal of Molecular Diagnostics**, v. 17, n. 6, p. 859, 2015.

ZHANG, J. et al. Presence of donor- and recipient-derived DNA in cell-free urine samples of renal transplantation recipients: Urinary DNA chimerism. **Clinical Chemistry**, v. 45, n. 10, p. 1741–1746, 1999.

ZHOU, Y. et al. Effects of Maternal and Fetal Characteristics on Cell-Free Fetal DNA Fraction in Maternal Plasma. **Reproductive Sciences**, v. 22, n. 11, p. 1–7, 1 nov. 2015.

ZIMMERMANN, B. et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. **Prenatal Diagnosis**, v. 32, n. 13, p. 1233–1241, dez. 2012.

ANNEX 1

Informed consent form signed by all participants

Universidade de São Paulo
Instituto de Biociências

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Biobanco do Centro de Estudos do Genoma Humano
Instituto de Biociências

A pesquisa com material biológico humano é muito importante para conhecer as causas das doenças genéticas e também para desenvolver terapias que, no futuro, possam melhorar o tratamento dessas doenças e trazer benefícios para os pacientes e familiares.

Para esse fim, estamos pedindo a contribuição de uma amostra de ( ) sangue e/ou ( ) outro tecido (*especifique tipo*: _____) para a realização de pesquisas. A decisão de doar é voluntária, e você pode se recusar a doar. Isso em nada vai afetar o seu tratamento médico ou atendimento.

O Biobanco

As amostras doadas e materiais delas extraídos ficarão armazenadas no Biobanco do Centro de Estudos do Genoma Humano - Instituto de Biociências (CEGH-IB), sob a guarda do Comitê Gestor do Biobanco. Esse Biobanco também vai armazenar suas informações pessoais necessárias para a realização das pesquisas. As informações e os seus dados genéticos terão um código para preservar seu anonimato.

Utilização das Amostras em Pesquisas

As amostras do Biobanco somente poderão ser utilizadas em pesquisas que tenham como objetivo compreender os mecanismos das doenças genéticas humanas ou desenvolver terapias que tragam benefícios a portadores de doenças genéticas. Toda pesquisa realizada com as amostras do Biobanco deverá ser previamente aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos do Instituto de Biociências (CEP-IB).

As amostras doadas poderão ser cedidas a pesquisadores de outros centros, desde que os projetos de pesquisa tenham sido aprovados pelo CEP da instituição da qual se originou o pedido, pelo CEP-IBUSP e pelo Comitê Gestor do Biobanco. Nenhuma amostra será vendida ou comercializada.

Confidencialidade de Informação

Junto com a amostra doada, serão coletadas informações a seu respeito e sobre sua família, tais como idade, sexo, histórico de doenças genéticas, entre outras

informações necessárias para a realização das pesquisas. Essas informações serão associadas a um código para você não ser identificado. Toda informação a seu respeito e de sua família fornecida ao pesquisador será mantida em sigilo.

Desconfortos e Benefícios

( ) A doação da amostra de SANGUE não trará nenhum risco à sua saúde, além do desconforto da introdução da agulha para a retirada de sangue.

( ) A doação da amostra de MUCOSA ORAL não trará nenhum risco à sua saúde. A coleta é feita passando uma escovinha no interior da bochecha.

A doação da amostra de _____ (*especifique o tecido*)

( ) virá de material excedente de cirurgia já programada, que de outra forma seria descartado. Nenhum procedimento extra será necessário, e não haverá nenhum risco específico.

( ) virá de material descartado por processos biológicos naturais. Nenhum procedimento extra será necessário, e não haverá nenhum risco específico.

( ) necessitará de um procedimento específico para a coleta, como descrito abaixo:

(*Descreva o procedimento e os possíveis riscos ou desconfortos associados a ele*)

_____

_____

_____

Você não obterá benefício ao doar sua amostra para o banco. Nós esperamos que as pesquisas realizadas aumentem nosso conhecimento sobre as doenças humanas, e que isso um dia traga melhoria no diagnóstico e tratamento dessas doenças.

Retirada do Consentimento

É seu direito retirar o consentimento para utilização das suas amostras no momento em que desejar, sem necessidade de qualquer explicação. Nesse caso, o material restante no banco será destruído, e não será usado em outras pesquisas. A retirada do consentimento não irá interferir no seu atendimento médico.

Resultado das Pesquisas

Os resultados das pesquisas serão publicados em revistas científicas e divulgados em encontros científicos. Nenhuma informação pessoal sua será divulgada.

Pesquisador responsável: _____

Contato: _____

Centro de Estudos do Genoma Humano
Instituto de Biociências, Universidade de São Paulo

Universidade de São Paulo
Instituto de Biociências

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Biobanco do Centro de Estudos do Genoma Humano
Instituto de Biociências

Eu, (*insira nome e profissão*) _____, nascido(a) em ___ / ___ / ___, residente e domiciliado na _____ _____, portador da Cédula de identidade _____, abaixo assinado(a),

Responsável pelo menor _____, nascido em ___ / ___ / ___,

Declaro:

que fui informado pelo responsável abaixo assinado:

- do objetivo da doação;
- do local onde as amostras serão guardadas e condições de sua utilização em pesquisa;
- do procedimento para coleta bem como seus riscos e benefícios;
- da confidencialidade dos meus dados pessoais;
- que posso retirar meu consentimento para utilização das amostras a qualquer momento, sem prejuízo para o meu tratamento;
- entendi as informações e pude fazer todas as perguntas que julguei necessárias.

Concordo:

- de livre e espontânea vontade em doar amostras de _____ para o Biobanco do CEGH-IB;

- que minhas amostras e informações pessoais fiquem armazenadas nesse banco e venham a ser utilizadas em pesquisas futuras que tenham sido aprovadas pelo Comitê de Ética em Pesquisa com Seres Humanos do Instituto de Biociências.

( ) Paciente / ( ) Responsável:

_____

Pesquisador Responsável pela Coleta:

_____

ANNEX 2

Informed consent form signed by all pregnant participants

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(maiores de 18 anos)

ESTUDO: Detecção de doenças genéticas fetais através de teste pré-natal não invasivo utilizando sequenciamento de nova geração

Você está sendo convidada a participar do projeto de pesquisa acima citado. O documento abaixo contém todas as informações necessárias sobre a pesquisa que estamos fazendo. Sua colaboração neste estudo será de muita importância para nós.

Eu, (insira nome e profissão) _____, domiciliado na _____, portador da Cédula de identidade (RG) _____, inscrito no CPF/MF _____, e nascido(a) em _____ / _____ /_____ , abaixo assinado(a), concordo de livre e espontânea vontade em participar do estudo "***Detecção de doenças genéticas fetais através de teste pré-natal não invasivo utilizando sequenciamento de nova geração***", e esclareço que obtive todas as informações.

Estou ciente que:

I) O estudo se faz necessário para que se possa desenvolver e implementar um teste diagnóstico pré-natal não invasivo de doenças genéticas, denominado ***"Detecção de doenças genéticas fetais através de teste pré-natal não invasivo utilizando sequenciamento de nova geração"***;

II) Será(ão) realizada(s) coleta(s) de 15 ml de sangue periférico, o que confere risco mínimo à saúde da mãe;

III) Essa(s) coleta(s) serão feitas apenas para este estudo e em nada influenciará (influenciarão) o meu tratamento; não vai (vão) me curar; não vai (vão) me causar nenhum problema, exceto o pequeno incômodo de dor no momento da coleta (introdução da agulha para retirada do sangue);

IV) Tenho a liberdade de desistir ou de interromper a colaboração neste estudo no momento em que desejar, sem necessidade de qualquer explicação;

V) A desistência não causará nenhum prejuízo à minha saúde ou bem estar físico. Não virá interferir no atendimento ou tratamento médico;

VI) Concordo que os resultados provenientes desse estudo sejam divulgados em publicações científicas, desde que meus dados pessoais sejam preservados;

VII) Concordo que não haverá possibilidade de receber o resultado do estudo durante a vigência da gravidez;

VIII) Caso eu desejar, poderei tomar conhecimento dos resultados, ao final desta pesquisa

( ) Desejo conhecer os resultados desta pesquisa

( ) Não desejo conhecer os resultados desta pesquisa

IX) Autorizo o pesquisador a entrar em contato após o nascimento da criança para validação do resultado encontrado na pesquisa;

X) O material colhido será depositado no Biobanco do Centro de Estudos do Genoma Humano (CEPID-FAPESP-IB), sob a responsabilidade do Instituto de Biociências, e guarda da Prof. Dra. Maria Rita dos Santos e Passos Bueno. Para isso será assinado o TCLE específico já aprovado pelo Comitê de Ética em Pesquisa – Seres Humanos, Instituto de Biociências da USP.

São Paulo, _____ de _____ de 20____

Participante: _____

Pesquisador Responsável pelo Projeto: _____

Carolina Malcher Amorim de Carvalho Silva

Telefone para contato: (11) 3091-9910

Endereço: Rua do Matão, 277, sala 200 - CEP 05508-090
Instituto de Biociências, Universidade de São Paulo

APPENDIX 1

Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate

Published in **Human Mutation** 36:1029–1033, 2015.

This study describes the impact of rare epithelial cadherin (*CDH1*) gene variants to etiology of Nonsyndromic Cleft Lip with or without Cleft Palate.

Our collaboration was important on the implementation and performance of the SKAT association test.

BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate

Luciano Abreu Brito,[1] Guilherme Lopes Yamamoto,[1] Soraia Melo,[2,3] Carolina Malcher,[1] Simone Gomes Ferreira,[1] Joana Figueiredo,[2,3] Lucas Alvizi,[1] Gerson Shigeru Kobayashi,[1] Michel Satya Naslavsky,[1] Nivaldo Alonso,[4] Temis Maria Felix,[5] Mayana Zatz,[1] Raquel Seruca,[2,3,6] and Maria Rita Passos-Bueno[1]*

[1]Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil; [2]IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal; [3]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; [4]Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil; [5]Laboratório de Medicina Genômica, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brasil; [6]Faculty of Medicine, University of Porto, Porto, Portugal

Communicated by Iain McIntosh

Received 2 February 2015; accepted revised manuscript 17 June 2015.
Published online 29 June 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22827

ABSTRACT: Nonsyndromic orofacial cleft (NSOFC) is a complex disease of still unclear genetic etiology. To investigate the contribution of rare epithelial cadherin (*CDH1*) gene variants to NSOFC, we target sequenced 221 probands. Candidate variants were evaluated via in vitro, in silico, or segregation analyses. Three probably pathogenic variants (c.760G>A [p.Asp254Asn], c.1023T>G [p.Tyr341*], and c.2351G>A [p.Arg784 His]) segregated according to autosomal dominant inheritance in four nonsyndromic cleft lip with or without cleft palate (NSCL/P) families (Lod score: 5.8 at θ = 0; 47% penetrance). A fourth possibly pathogenic variant (c.387+5G>A) was also found, but further functional analyses are needed (overall prevalence of *CDH1* candidate variants: 2%; 15.4% among familial cases). *CDH1* mutational burden was higher among probands from familial cases when compared to that of controls ($P = 0.002$). We concluded that *CDH1* contributes to NSCL/P with mainly rare, moderately penetrant variants, and *CDH1* haploinsufficiency is the likely etiological mechanism.

Hum Mutat 36:1029–1033, 2015.

KEY WORDS: *CDH1*; oral clefts; gastric cancer; two-hit model; rare variant

APPENDIX 2

A New Locus for Nonsyndromic Cleft Lip with or without Cleft Palate

Submitted to **Journal of Dental Research** in February 20th, 2017.

This study describes a novel locus associated with Nonsyndromic Cleft Lip with or without Cleft Palate.

Our collaboration was important on the genotype analysis of candidate genes.

Journal of
Dental Research

A New Locus for Nonsyndromic Cleft Lip with or without Cleft Palate

Journal:	*Journal of Dental Research*
Manuscript ID	JDR-17-0170
Manuscript Type:	Research Reports
Date Submitted by the Author:	20-Feb-2017
Complete List of Authors:	Masotti, Cibele; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology; Hospital Sirio-Libanes, Instituto de Ensino e Pesquisa do Hospital Sirio-Libanes Brito, Luciano; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Nica, Alexandra; University of Geneva Medical School, Department of Genetic Medicine and Development; Swiss Institute of Bioinformatics Geneva Ludwig, Kerstin; Institute of Human Genetics Biomedical Center University of Bonn Sigmund-Freud-Strasse 25 D-53127 Bonn, Germany, Department of Genomics; University of Bonn, Institute of Human Genetics Nunes, Kelly; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Savastano, Clarice; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Malcher, Carolina; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Ferreira, Simone; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Kobayashi, Gerson; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Bueno, Daniela; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Alonso, Nivaldo; Faculdade de Madicina da Universidade de Sao Paulo, Department of Plastic Surgery Franco, Diogo; Universidade Federal do Rio de Janeiro, Hospital Universitário Clementino Fraga Filho, Department of Plastic Surgery Rojas-Martinez, Augusto; Universidad Autonoma de Nuevo Leon, Centro de Investigación y Desarrollo en Ciencias de la Salud Dos Santos, Sidney; Universidade Federal do Para, Institute of Biological Sciences Galante, Pedro; Hospital Sirio-Libanes, Instituto de Ensino e Pesquisa do Hospital Sirio-Libanes Meyer, Diogo; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology Hunemeier, Tabita; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology

	Mangold, Elisabeth; University of Bonn, Institute of Human Genetics Dermitzakis, Emmanouil; University of Geneva Medical School, Department of Genetic Medicine and Development; Swiss Institute of Bioinformatics Geneva Passos-Bueno, Maria Rita; Universidade de Sao Paulo, Department of Genetics and Evolutionary Biology
Keywords:	Craniofacial biology/genetics, Molecular genetics, Orofacial cleft(s), Gene expression, Genomics, Bioinformatics
Abstract:	A valuable approach to understand how individual and population genetic differences can predispose to disease is to access the impact of genetic variants on cellular functions (e.g., gene expression) of cell and tissue types related to pathological states. In order to understand the genetic basis of nonsyndromic cleft lip with or without cleft palate (NSCL/P) susceptibility, a complex and prevalent developmental disorder, we searched for genetic variants with regulatory role in a disease-related tissue, the lip muscle (orbicularis oris muscle, OOM) of affected individuals. From 46 OOM samples, which are frequently discarded during corrective surgeries that patients with orofacial clefts routinely undergo, we derived mesenchymal stem cells, and correlated the individual genetic differences with gene expression. Through this strategy, we detected significant cis-eQTLs, i.e. DNA variants affecting gene expression, and selected a few candidates to conduct an association study in a large Brazilian cohort (624 patients and 668 controls). This resulted in the discovery of a novel susceptibility locus for NSCL/P, rs1063588, the best eQTL for the *MRPL53* gene, for which the association was mostly driven by the Native American ancestry component of our sample. *MRPL53* (2p13.1) encodes a 39S protein subunit of mitochondrial ribosomes and interacts with *MYC*, a transcription factor required for normal facial morphogenesis. Our study not only illustrates the importance of sampling admixed populations, but also the relevance of measuring the functional effects of genetic variants over cellular phenotypes to dissect the complexity of human organism phenotypes.

ERRATA

Na tese de doutorado intitulada “Teste pré-natal não invasivo para detecção de doenças genéticas utilizando sequenciamento de nova geração“, na Tabela 2 da página 41, onde se lê:

Sample	Trisomy*	Test Result T21 (Z-score)	Test Result T18 (Z-score)	Fetus Gender	Test Result Fetal Sex (Z-score)
FD1500110	T21	T21 (12.9)	Not-T18 (1.39)	Male	Male (55.7)
FD1500068	T21	T21 (8.5)	Not-T18 (1.2)	Male	Female (-1.05)
FD1500092	T21	T21 (7.8)	Not-T18 (0.68)	Male	Male (51.4)
FD1500073	T21	T21 (6.8)	Not-T18 (2.12)	Male	Female (-1.06)
FD1500098	T21	T21 (4.4)	Not-T18 (2)	Female	Male (101.9)
FD.15.00142	Normal	Not-T21 (0.2)	Not-T18 (-0.11)	Female	Male (98.8)
FD.15.00141	Normal	Not-T21 (0.17)	Not-T18 (1.81)	Female	Female (-1.05)
FD1500107	T18 (11Mb)	Not-T21 (-0.55)	Not-T18 (1.91)	Male	Male (27.4)

Leia-se:

Sample	Trisomy*	Test Result T21 (Z-score)	Test Result T18 (Z-score)	Fetus Gender	Test Result Fetal Sex (Z-score)
FD1500110	T21	T21 (12.9)	Not-T18 (1.39)	Male	Male (55.7)
FD1500068	T21	T21 (8.5)	Not-T18 (1.2)	Male	Male (27.41)
FD1500092	T21	T21 (7.8)	Not-T18 (0.68)	Male	Male (101.9)
FD1500073	T21	T21 (6.8)	Not-T18 (2.12)	Male	Male (51.44)
FD1500098	T21	T21 (4.4)	Not-T18 (2)	Female	Female (-1.05)
FD.15.00142	Normal	Not-T21 (0.2)	Not-T18 (-0.11)	Female	Female (-1.05)
FD.15.00141	Normal	Not-T21 (0.17)	Not-T18 (1.81)	Female	Female (-1.05)
FD1500107	T18 (11Mb)	Not-T21 (-0.55)	Not-T18 (1.91)	Male	Male (98.8)