NON-INVASIVE PRENATAL DETECTION OF DELF508 CFTR MUTATION STATUS

by

Patricia Ann Shaw

BA, Carnegie Mellon University, 2009

Submitted to the Graduate Faculty of

Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Patricia Ann Shaw

It was defended on

April 13, 2016

and approved by

Thesis Advisor: David Peters, Ph.D., Associate Professor, Department of Obstetrics, Gynecology, and Reproductive Sciences, School of Medicine, University of Pittsburgh

Candace Kammerer, Ph.D., Associate Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Ryan L. Minster, Ph.D., M.S.I.S., Assistant Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Copyright © by Patricia Ann Shaw

2016

David Peters, Ph.D.

NON-INVASIVE PRENATAL DETECTION OF DELF508 CFTR MUTATION STATUS

Patricia Ann Shaw, MS

University of Pittsburgh, 2016

ABSTRACT: Cystic fibrosis is the most common fatal genetic disorder in the Northern European population; however early diagnosis and treatment can lead to increased quality and length of life for affected individuals and therefore its diagnosis is of great public health significance. Cystic fibrosis is a monogenic recessive disorder caused by mutation in the CFTR gene. Current screening methods include newborn screening tests and invasive prenatal diagnostic tests. Newborn enzymatic screening tests are prone to false positive results, leading to unwarranted stress and anxiety for new parents. Invasive prenatal testing is available, but these invasive procedures carry some risks including miscarriage which also causes anxiety and stress for parents. Non-invasive prenatal testing for fetal aneuploidy using cell-free fetal DNA derived from maternal plasma is now part of the regular clinical offering for women in their first trimester of pregnancy. No commercial non-invasive prenatal test for monogenic disease is currently available. The goal of this study is to develop a non-invasive prenatal test for the detection of the mutation status of the fetus. Proposed here are methods to haplotype the parents in the CFTR region and to genotype the plasma to establish the mutation status of the fetus. In comparison to newborn screening or invasive diagnostic tests, a non-invasive prenatal test for cystic fibrosis would have little to no risk and also provide parents more time to identify clinical and psychosocial resources to facilitate their family's needs and decision-making processes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 BACKGROUND

1.1 PRENATAL SCREENING TESTS

Prenatal screening tests are designed to indicate whether the possibility of a fetal genetic condition is higher or lower than would be expected in the general population. These screenings do not serve as a definitive diagnosis of a fetal condition. Currently three types of screening are available to pregnant women: Ultrasound, First Trimester Screening, and Multiple Marker Screening[1]. These screening methods are not exclusive and many times will be combined.

Ultrasound

Ultrasound is used to obtain accurate dating of a pregnancy and to assess fetal growth and development. Birth defects resulting from chromosomal abnormalities, such as Down syndrome, can be observed as well as other defects not caused by chromosome abnormalities, such as spina bifida. Ultrasounds done at 18-20 weeks can detect Down syndrome (50% detection rate), Trisomy 18 or 13 (90% detection rate), spina bifida (90% detection rate), and anencephaly (100% detection rate)[1].

First Trimester Screening (FTS)

The first trimester screening consists of two parts: an ultrasound and a maternal serum blood test. FTS is performed between 11-14 weeks. The ultrasound measures nuchal translucency and the blood test measures the levels of pregnancy associated plasma protein (PAPP-A) and total human chorionic gonadotrophin (total hCG). Abnormal nuchal translucency and/or abnormal

levels of PAPP-A or total hCG can be an indication of Trisomy 18 or Down syndrome. FTS detects 85% of pregnancies with Down syndrome or Trisomy 18[1].

Multiple Marker Screening (MMS)

Like FTS, multiple marker screening detects pregnancies at increased risk for Down syndrome and Trisomy 18. Additionally it can also identify spina bifida and anencephaly. MMS is done at 15-20 weeks and is a maternal serum blood test which measures alpha-fetoprotein (AFP), hCG, unconjugated estriol (uE3) and inhibin-A. MMS can identity Down syndrome (70% detection rate), Trisomy 18 (60% detection rate), spina bifida (80% detection rate), and anencephaly (90% detection rate)[1].

While these prenatal screening tests are non-invasive, they are not diagnostic. These screenings also result in false positives. The false positive rate for Down syndrome detection is 5% using FTS and 5.2% using MMS[2]. Combining these screening methods and performing them at multiple gestational ages, known as integrated screening, can lower the false positive rate to 1.9%[2]. These false positive results are a great source of anxiety for expectant parents. When a patient receives abnormal screening results, they are advised that a definitive diagnosis can only be given by chorionic villus sampling or amniocentesis.

1.2 PRENATAL DIAGNOSTIC TESTS

When prenatal genetic testing is warranted to confirm the presence or absence of a condition, the current standard of care for the patient is to undergo an invasive procedure, either chorionic villus sampling (CVS) or amniocentesis, to extract fetal DNA for further testing. Further testing

on the fetal sample may include karyotype analysis, fluorescence *in situ* hybridization (FISH) analysis, microarray analysis, and specific testing for certain monogenic disorders. CVS is generally performed between 10-13 weeks of gestation and amniocentesis is generally performed after 15 weeks gestation[3]. Because of the invasive nature of these tests there are risks associated with them for the patient and also to the fetus including a risk of miscarriage[4-8]. While the risk of miscarriage may be low, 0.1% at experienced centers, this can lead to additional parental anxiety. Due to these risks, a less invasive, yet still definitive test is desired for fetal genomic testing.

1.3 NON-INVASIVE PRENATAL DIAGNOSTIC TESTING FOR CHROMOSOMAL ABNORMALITIES

It is well established that fetal cells are present in the maternal blood stream [9-12]. Fetal cells isolated from maternal blood have been used to test for aneuploidy and monogenic disorders [11, 12]. However these cells are extremely rare and estimated to only be present at one fetal cell per milliliter of maternal blood[11]. Isolation of these cells is very difficult and clinical trials for aneuploidy testing via the isolation and analysis of fetal cells showed that the sensitivity was very similar to existing serum screening tests for Down syndrome due to the difficult nature of actually isolating fetal cells [13, 14].

In 1997, Dennis Lo showed that fetal DNA is present as part of the circulating cell-free DNA in plasma and serum by detection of the Y chromosome in mothers of male fetuses[15]. Using a digital PCR method, it was shown that the median fetal-DNA concentration in the maternal plasma for the first, second, and third trimesters was 9.7%, 9.0%, and 20.4%

respectively[16]. In 2008, two groups demonstrated that it was possible to diagnose fetal aneuploidy from the maternal plasma using newly developed shotgun sequencing technology [17, 18]. Non-invasive prenatal detection of fetal aneuploidy has now become part of the regular clinical offering for women in their first trimester with commercial tests available such as MaterniT21 (Sequenom), Verifi (Progenity), Harmony (Ariosa), and Panorama (Natera).

With the clinical success of aneuploidy testing using cell-free fetal DNA, the field is now looking to the detection of other fetal genetic anomalies such as deletions and duplications as well as resolving the entire fetal genome using cell-free DNA from the mother to enable diagnosis of Mendelian disorders.

1.4 NON-INVASIVE PRENATAL DIAGNOSTIC TESTING FOR MONOGENIC DISORDERS

The original discovery of the presence of cell-free DNA in maternal blood was based on the detection of sequence, the Y chromosome, not present in the maternal genome. These paternally inherited sequences, where the mother lacks the paternal allele, are readily detectable using standard PCR methods because no sequence specific maternal background exists. In this fashion, clinical tests are currently in use to detect fetal RHD status with cell-free DNA[19]. Detection of the paternal allele can be used to diagnose Mendelian disorders when the allele is autosomal dominant and can also be used to exclude when shown that the fetus did not receive the paternal allele, if different than the maternal, in an autosomal recessive disorder. This has been used for diseases such as Huntington's, myotonic dystrophy, early onset primary dystonia, thalassemia, and achondroplasia [20-24].

In the case of an autosomal recessive disorder where the mother is a carrier and the father is also a carrier of the same allele, definitive diagnosis is much more difficult to make since the fetal DNA must be detected among the large background of maternal DNA with the same allele.

Table 1: Expected allelic ratio change in plasma by fetal fraction.

Fetal Fraction	Maternal Genotype	Fetal Genotype	Allele Ratio (A/B)
20%	AB	AB	50/50
		AA	60/40
		BB	40/60
10%	AB	AB	50/50
		AA	55/45
		BB	45/55
5%	AB	AB	50/50
		AA	52.5/47.5
		BB	47.5/52.5

As shown in to Table 1, when the mother and father are heterozygotes, the fetal genotype will be either: AB, AA, or BB. Since the cell free plasma DNA contains both mother and fetal DNA the allelic ratio in the plasma DNA will be skewed from 50/50 when the fetus is homozygous for one of the alleles. The degree to which that ratio will be different from 50/50 depends on the fetal fraction. If the B allele represents the deleterious allele then anytime B counts are greater than A counts, the fetus inherited the deleterious allele. Plasma DNA is highly fragmented and is found at very low levels. In general, conventional methods to measure allele frequency such as Taqman SNP assays are not sensitive enough to be able to detect these minor allele ratio shifts. In 2008, Lun et al published two digital PCR based methods, Digital Relative Mutation Dosage (RMD) and Digital Nucleic Acid Size Selection (NASS) to detect these allelic ratio shifts for two mutations in the HBB gene for which the mother was heterozygous[25]. When used in tandem, they were only able to accurately classify 6 of 10 second trimester

maternal plasma samples[25]. Improved methods and droplet digital PCR machinery have made digital PCR a more viable option for diagnosis. In 2014, a study was published where this approach was used to diagnose whether a fetus had an autosomal recessive disorder, methylmalonic acidemia[26]. Although digital PCR has become increasingly more accessible, the technology still relies on single amplicon amplification such that multiple PCR reactions would be required to test for more than one disorder. Also, since cell free plasma DNA is found in such low quantities, the number of distinct tests that could be run on a single sample is limited. Ideally one test would be able to screen for multiple disorders with the same sample and all testing would be completed at one time.

In 2012, two groups published papers showing that the entire fetal genome could be derived by molecular counting of the parental haplotypes within plasma DNA by shotgun sequencing [27, 28]. Instead of depending on counting the allelic ratio shift at one locus, they used the allelic ratio shifts at multiple nearby loci to detect which haplotype block the fetus inherited and thus are able to deduce whether the fetus inherited a specific allele. Lam et al also used parental haplotypes and shotgun sequencing to derive the fetal genome, but they chose to do it in a targeted manner[29]. By targeting parts of the genome instead of the entire genome, higher read depth is possible with fewer sequencing reads, decreasing costs and allows focus on clinically relevant loci. They targeted a 288Kb region of the HBB gene and were able to determine if the fetus inherited either mutation from the mother or the father[29]. Each of these approaches relied on first classifying the maternal haplotypes. The paternal haplotype is deduced from the presence of paternally inherited alleles, where the mother does not share the paternal allele, and are detectable after shotgun sequencing. The methods presented in these three papers

for determining the haplotypes for the mother are time and cost intensive. Therefore a more cost effective and less time intensive method is needed to haplotype the mother.

Currently there are no non-invasive prenatal commercial tests available for the detection of monogenic disorders using either cell-free DNA or fetal cells from a maternal blood sample. There are tests available for many monogenic disorders which are used regularly in the clinical setting such as cystic fibrosis, sickle cell, thalassemia's, and panels specific to disorders common in certain populations such as Ashkenazi Jews. These tests require pure genomic DNA so they can only be used to test parental samples and a pure fetal sample from a CVS or amniocenteses. The prevalence of monogenic disorders at birth is 10/1000 births[30]. The prevalence of monogenic disorders at birth is 10 times higher than the prevalence of Down syndrome, a now commonly screened for chromosome defect, which occurs in about 1:1000 live births[30]. Unlike Down syndrome there is no non-invasive prenatal test for these autosomal recessive monogenic disorders. A non-invasive prenatal test needs to be developed to detect autosomal recessive monogenic diseases. Cystic Fibrosis is the most common monogenic disorder worldwide and developing a non-invasive test to detect Cystic Fibrosis prenatally would yield a significant public health impact.

1.5 CYSTIC FIBROSIS

Cystic fibrosis of the pancreas, more commonly known as cystic fibrosis, is a life threatening progressive multi-system disorder. "Cystic fibrosis affects the epithelia of the respiratory tract, exocrine pancreas, intestine, male genital tract, hepatobiliary system, and exocrine sweat glands…"[31]. Mucus protects the linings of airways, digestive system, reproductive system, and

other organs and tissue, but in those affected with cystic fibrosis, the mucus produced is abnormally thick and sticky[32]. The disease shows great phenotypic heterogeneity. In the respiratory tract, the abnormal mucus can clog the airways leading to inflammation and chronic endobronchial infection, which leads to immune system failure in the lungs causing bacterial endobronchitis, which leads to further obstruction and inflammation[31]. Digestive problems are also seen with the majority of those affected having pancreatic insufficiency and less than 10% affected having pancreatic sufficiency[31]. “In people with cystic fibrosis, mucus blocks the ducts of the pancreas, reducing the production of insulin and preventing digestive enzymes from reaching the intestines to aid in digestion… a shortage of insulin can cause a form of diabetes known as cystic fibrosis-related diabetis mellitus.”[32]. Infertility due to azoospermia is seen in over 95% of males[31]. The leading cause of death is pulmonary disease and the second leading cause of death is liver disease[31]. Median survival is 36.5 years with greater survival in males and in those with pancreatic sufficiency[31].

Cystic fibrosis is caused by mutation in the cystic fibrosis transmembrane conductance regulator gene CFTR on chromosome 7. The CFTR gene codes for a protein channel responsible for transporting chloride ions which helps to control the movement of water in the tissues, necessary for the production of normal mucus[32]. Without enough water, this mucus becomes thick and sticky. Inheritance of cystic fibrosis is autosomal recessive. Two deleterious alleles are required and carriers of one allele do not show symptoms. Over 1900 mutations in the CFTR gene have been identified that might cause dysfunction of the CFTR protein [33]. Only a few dozen of those have been well classified for their impact on the CFTR protein [34]. These mutations can be classified into six main categories (I –VI) causing either complete absence of

protein to decreased stability of the membrane [33]. Figure 1 below summarizes the six classifications of CFTR mutations and their effect on the resulting protein product.

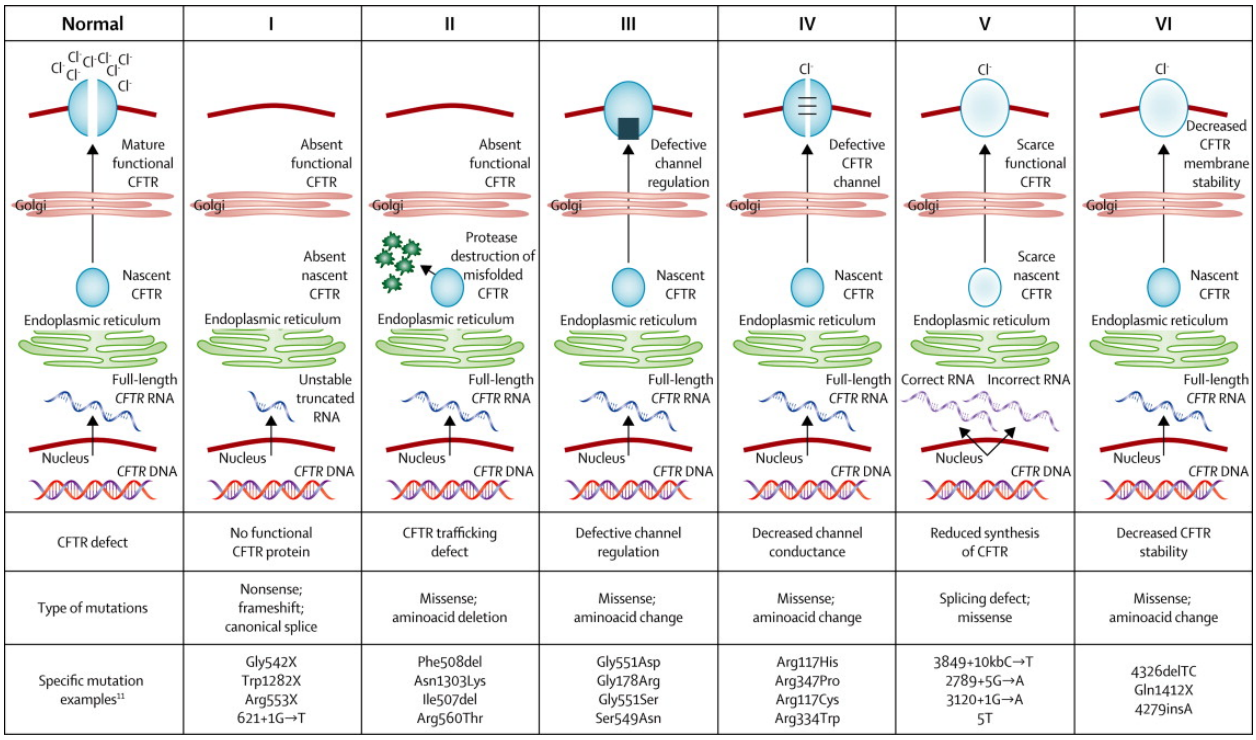

Figure 1: Classes of CFTR mutations.

CFTR mutations are classified into 6 categories. The delF508 mutation is class II. Reprinted from The Lancet Respiratory Medicine, Volume 1, Issue 2, Boyle, Michael P., De Boeck, Kris, A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect, Pages 158-163, Copyright 2013, with Permission from Elsevier.

Cystic fibrosis is the most common fatal genetic disease in persons of northern European heritage[35]. It occurs in 1:3200 live births of northern European ancestry, 1:15000 in African-Americans, and 1:31000 in Asians[35]. The risk of being a cystic fibrosis carrier is 1:29 (3.44%), 1:28 (3.57%), and 1:61 (1.64%) in Ashkenazi Jews, North American of northern European heritage, and African-American respectively[36, 37]. Those in the at-risk population for cystic fibrosis are advised to have cystic fibrosis carrier screening by genetic counselors during pregnancy. If both parents are carriers, an invasive procedure may be performed to determine if

the fetus is affected. The current cystic fibrosis mutation panel endorsed by the American College of Medical Genetics (ACMG) covers 23 mutations with frequencies over 0.1%, which account for roughly 84% of all the mutations in a pan-ethnic U.S. population[38]. Sosnay et al analyzed the phenotypes and penetrance of all variants at or above 0.01% in cystic fibrosis patients and found an additional 104 variants to be pathogenic [34]. They estimated that testing for all 127 variants accounted for 95.4% in their sample versus 87.5% by testing only the 23 ACMG genes [34]. The carrier screening cannot actually provide a true negative result because of the limited markers that are tested. However, having a negative result on a carrier screening significantly decreases the patient's chance of being a cystic fibrosis mutation carrier. Next generation sequencing technologies have an advantage because more mutations can be detected and one study indicated that using sequencing technologies for CFTR testing would increase mutation detection to 98.7%[39].

The delF508 mutation is the most common CFTR mutation. Worldwide, roughly 90% of affected individuals have at least 1 copy of this mutation in conjunction with another mutation and 50% are homozygous for the mutation [33]. According to the ACMG, the delF508 mutation accounts for the highest percentage of CFTR mutations found in all but one of the US populations studied. Among US populations the frequency of the delF508 mutation is 72.42% in Non-Hispanic Caucasians, 54.38% in Hispanic-Caucasians, 44.07% in African-Americans, and 31.41% in Ashkenazi Jews[38]. The delF508 mutation is a 3 base pair nucleotide deletion resulting in the loss of an amino acid, phenalynine, causing the resulting protein to misfold and not be trafficked to the cell-surface [33]. Studies on European populations revealed "… a clear-cut north-west to south-east gradient of delF508 prevalence in CF patients showing a maximum prevalence of 88% in Denmark to a minimum of 24.5% in the Turkish population"[40]. It is

estimated that the delF508 mutation is at least 52,000 years old and may have arisen in the Baluchi population in the Middle East [40, 41].

1.6 DESIGN CONSIDERATIONS

1.6.1 Haplotyping Assay Design

In order to reduce the cost of sequencing on a per sample basis, the haplotyping and plasma genotyping assay needs to be targeted to the disease allele region. Whole genome haplotyping is not necessary as the clinical question revolves around the fetus inheriting the delF508 mutation and nothing else. The size of the region needs to be long enough so that it contains enough loci to differentiate parental haplotypes, but not cover such a long span that would contain a site where a recombination event may occur. A meiotic recombination event is estimated to occur about once per chromosome arm per generation[42]. In order to minimize the chance of spanning a recombination site, the degree of linkage disequilibrium (LD) should be maximized across those loci. A region of high LD means that rather than alleles segregating completely independently, the alleles segregate as small blocks. The degree of LD between SNPs will also vary by population. In the CEU population (Utah residents with ancestry from northern and western Europe), the CFTR gene consists of 3 haplotype blocks, each outlined in black, spanning 108Kb, 70Kb and 2Kb respectively[43]. The delF508 mutation is found within the first block. Since there is less of a chance that recombination will occur within the haplotype block, only SNPs within block one were chosen.

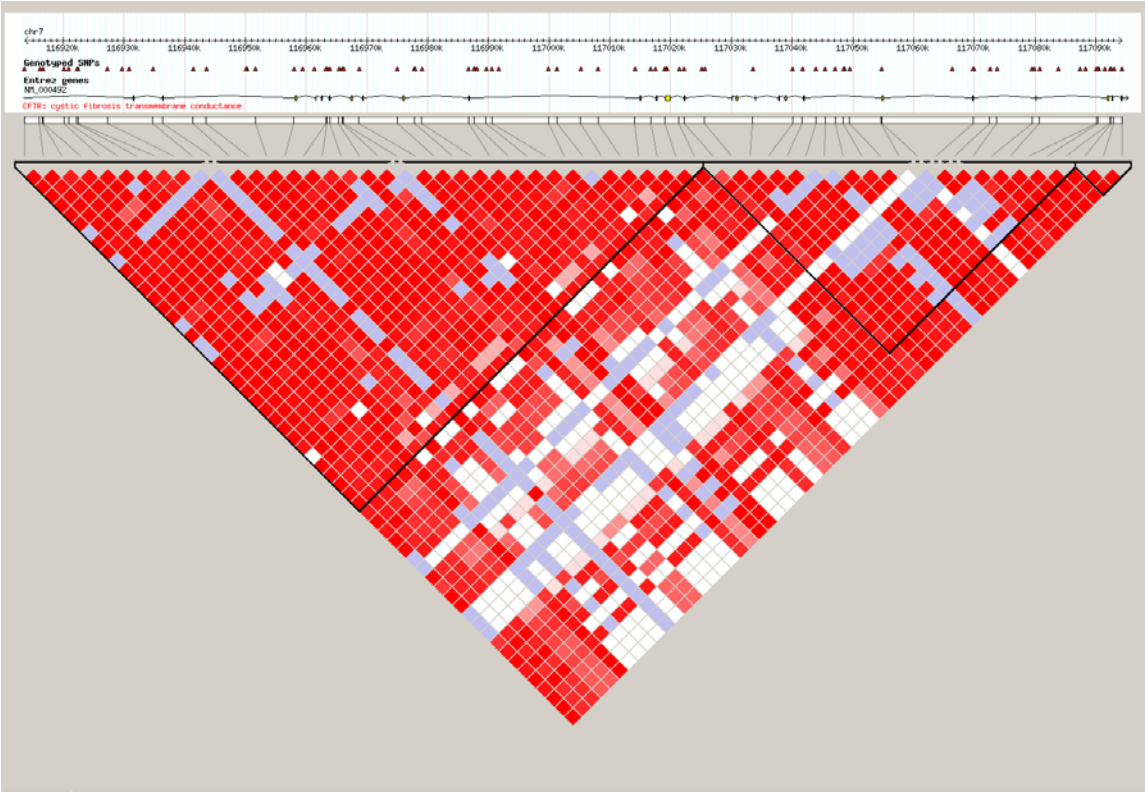

Figure 2: LD Plot of the CFTR gene (Hapmap CEU Population)

The CFTR gene consists of 3 haplotype blocks spanning 108Kb, 70Kb and 2Kb respectively[43].

Figure 3 outlines the method by which loci were targeted and connected for haplotyping. Within haplotype block 1, from Figure 2, 12 SNPs plus the delF508 mutation were targeted via long range PCR. SNPs were chosen based on their inclusion within the Hapmap database and those that possess a high heterozygosity rate (>40%) to maximize the genotype differences within the samples. Additionally, these loci had to be close enough together for long range PCR. These 13 sites span 35.5 Kb. In order to link heterozygous loci, it is necessary for each locus to link to multiple loci. In total 42 long range amplicons were designed to connect 13 loci. These long range amplicons were 269-14,296bp in length. While the Illumina sequencers are capable of up to 300bp paired-end sequencing, the total fragment length still poses a limitation for the on-

board sequencing chemistry. Typical amplicon lengths for sequencing are 200-500bp, but amplicons up to 1500bp in size can be used. Since the majority of the long range amplicons exceed the optimum length, it is necessary to make the final amplicons smaller while still containing the two far away loci. Therefore, after long range amplification, the amplicons are circularized and an inverse PCR is performed on the circularized fragment. These new inverse PCR amplicons are 161-312bp.

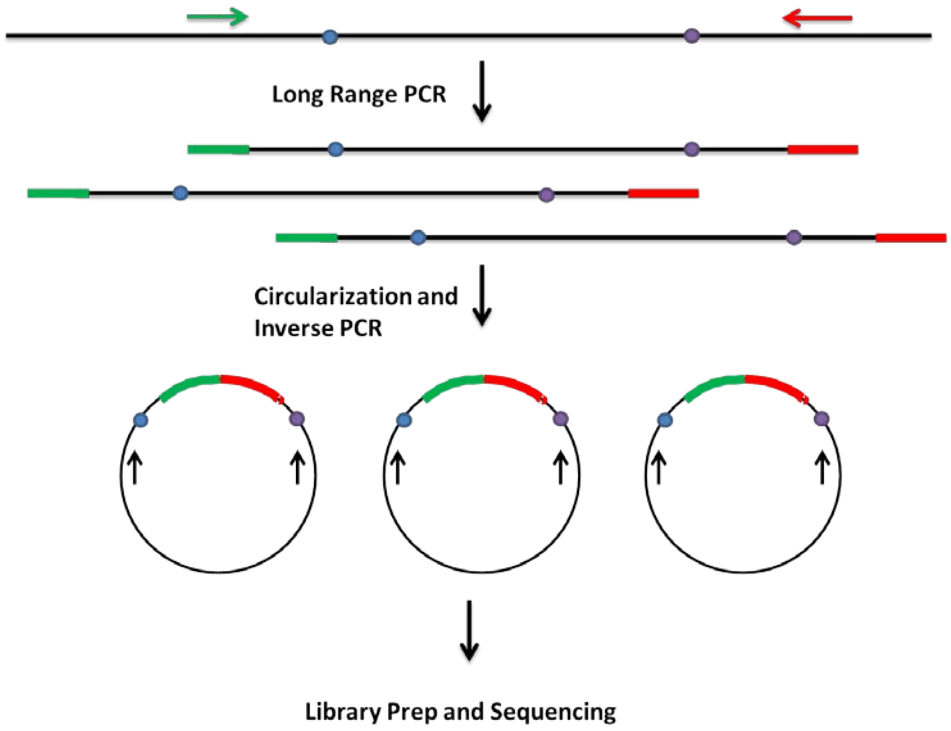

Figure 3: Strategy for the haplotyping of nuclear DNA.

Long range PCR was performed on each amplicon individually and per sample. Long range amplicons are circularized and then an inverse PCR is performed on the circularized products.

1.6.2 Plasma Genotyping

To non-invasively haplotype the fetus using the parental haplotypes, allele counts need to be attained across the haplotyped loci. In order to obtain sufficient sequencing depth to calculate allele frequencies economically a targeted approach must be used to sequence the plasma DNA. Agilent SureSelect, which uses RNA baits, was used previously for this purpose[29]. However, custom panel designs have a high up front cost and are not economical for the small amount of loci that this parental haplotyping method was initially designed to cover. Multiplex PCR of plasma DNA is more economical for this purpose. However, since PCR results in exponential amplification rather than linear amplification, the allele frequency at heterozygous sites, in practice, is often skewed from 50/50 in a pure sample due to uneven amplification of alleles. This was discussed previously in section 1.4. The expected change in ratio in the plasma sample is very slight (Table 1) and the unequal amplification resulting from PCR will create noise, making the allele frequency change from the fetus potentially undetectable. Since amplification of the plasma DNA is necessary, a unique molecular identifier (UMI) is attached to the DNA prior to amplification. The UMI is a 12bp nucleotide sequence which tags each starting DNA molecule so that after amplification, PCR duplicates can be removed and only unique molecules are used for the allele counts. A conventional PCR design could be used to target the loci, but would not allow individual tagging of each molecule prior to amplification. This method allows for tagging each molecule and targeting the assay to specific loci to decrease sequencing costs.

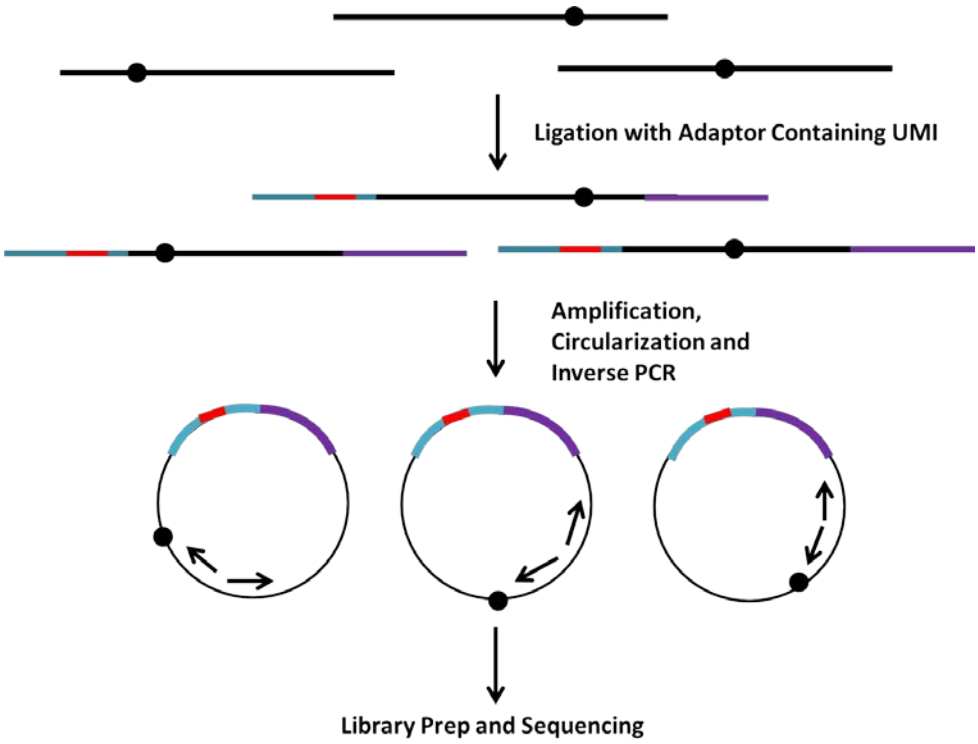

Figure 4: Strategy for genotyping plasma DNA.

Custom adaptors containing a 12bp unique molecular identifier (UMI) were ligated onto plasma and used for library amplification. Amplification products were circularized and inverse PCR was performed to linearize the products for subsequent sequencing library preparation.

1.7 PUBLIC HEALTH SIGNIFICANCE

All states now test for cystic fibrosis as part of their newborn screening panel[44]. The test for cystic fibrosis as part of newborn screening is two-tiered (IRT+DNA) in most states including Pennsylvania[44]. The first part of the test measures the levels of immunoreactive trypsinogen (IRT). If the level of IRT is high and a two-tiered test is performed, DNA is extracted from the blood spot for CFTR mutation testing. If the state uses an IRT-Only method then a subsequent newborn screening test is performed a few weeks later. If the child still has an elevated IRT

level, then a sweat test will be performed for definitive diagnosis. Newborn screening accounted for 60% of newly detected cystic fibrosis cases[44]. The IRT test used in newborn screening has significant drawbacks. Studies have shown that IRT/DNA screening has a very high sensitivity and specificity, but the initial IRT testing is prone to false positives [45]. In a study from the Netherlands, 1,493 out of 145,499 (1%) newborns had a high IRT level. Further genetic testing was performed and only 37 of the newborns were affected with cystic fibrosis (67 were carriers) [45]. 93%.of the newborns screened with high IRT levels had no CFTR mutations detected [45]. Based on this study, out of 10,000 births, 100 newborns will have a high IRT level and be directed for further testing, but 93 of those 100 will be false positives. This high false positive rate is concerning due to the likely parental anxiety that would occur should parents be notified that their newborn might have a serious, life-limiting genetic condition while they wait for a definitive answer. 93 out of those 100 were caused additional unwarranted stress during an already stressful time.

A definitive, one-time genetic based test for cystic fibrosis would likely be preferable to most parents rather than the current newborn based screening test with high initial false positive rates especially in states where DNA testing is not done by default. Additionally, performing this test prenatally would give parents additional time to consider their options. Currently, this testing is offered through invasive procedures (CVS and amniocentesis), but since these invasive procedures carry some risk to the fetus, this causes additional parental anxiety. A non-invasive prenatal test for cystic fibrosis would have little to no risk and also provide parents more time to make reproductive decisions.

Increased prenatal screening for cystic fibrosis will also likely have a positive economic impact. Life expectancy has increased dramatically due to medical advancements in treatment

over the last few decades for those with cystic fibrosis, but these new treatments come at a high cost. A 2013 study found that the average cost of care for treating cystic fibrosis was $15,571 annually [46]. The mean annual costs for mild, moderate, and severe disease are $10,151, $25,647, $33,691 respectively [46]. Certain mutation combinations in the CFTR gene are associated with increased severity of the disease. Earlier diagnosis for cystic fibrosis would give parents more time to identify clinical and psychosocial resources to facilitate their needs and decision-making processes. For example, parents may decide to undergo pre-implantation genetic diagnosis procedures in the future, and this may lead to at least some decrease in the births of affected fetuses.

Proposed here is a method to haplotype genomic maternal and paternal samples to aid in the assignment of the fetal haplotype from sequencing maternal plasma cell-free DNA. Additionally, the incorporation of unique molecular identifiers (UMI) allows plasma reads resulting from PCR amplification to be removed so that only the original molecules are used in the allele frequency count. This assay is designed to haplotype the region in the CFTR gene containing the delF508 mutation. Cystic fibrosis is the most common monogenic disorder worldwide and the delF508 mutation is the most common mutation causing cystic fibrosis. Currently there are no commercially available non-invasive prenatal tests for cystic fibrosis or other autosomal recessive monogenic disorders. This design could be used to target other CFTR mutations as well as other loci linked to monogenic disorders.

2.0 SPECIFIC AIMS

Specific Aim 1: To evaluate a new method to haplotype maternal and paternal samples of trios in the CFTR region.

Specific Aim 2: To evaluate whether the delF508 mutation is likely associated with one or multiple haplotypes within this sample population.

Specific Aim 3: To evaluate a new method to count allele frequency in plasma using a unique molecular identifier and assign the fetal haplotype.

3.0 METHODS

Recruitment

Samples from six families, containing paternal blood, maternal blood, and pure fetal sample were obtained and used for evaluation of the haplotyping method. Two of these families consist of paternal and maternal delF508 carriers with the fetus not inheriting the delF08 mutation in either family. Five of the six maternal samples came from patients who reported their ethnicity as white and one of the maternal samples reported themselves as multiple. Only one of the six paternal samples was reported as white and there was no data for the other five. Ethnicity is not reported for the fetus. Due to a limited number of available samples, use of only trios with reported ethnicity information was not practical. Since the delF508 mutation is most prevalent in the Caucasian population and a vast majority of patients seeking prenatal genetic testing at Magee-Womens Hospital identify themselves as Caucasian, the CEU population (Utah residents with ancestry from northern and western Europe) in the Hapmap database will be used for reference LD and haplotype information. The University of Pittsburgh Institutional Review Board approved the patient consenting process and collection of all samples used in this study. Written informed consent was obtained for every sample.

DNA extraction

Parental genomic DNA was extracted from frozen blood minus the plasma portion using the QIAmp DNA Blood Midi Kit (Qiagen) as per protocol. Fetal genomic DNA was extracted from

CVS or amniocyte cultures. Cells were dissociated from culture dishes by treating with 0.05% Trypsin with EDTA (Invitrogen) for 10 minutes at 37°C followed by dilution with 1ml of DPBS (Gibco). The cellular solution was centrifuged at 500 x g for 5 minutes at 4°C and supernatant was removed. The cell pellet was resuspended in 200µl DPBS (Gibco). DNA from cultured CVS or amniocytes was extracted using the QIAmp DNA Mini Kit (Qiagen) as per protocol. Extracted DNA was quantified by nanodrop.

Haplotyping of Nuclear DNA

Long range PCR was performed on each amplicon individually and per sample. Input DNA for parental samples was 5ng and 6ng for fetal samples per PCR. The PCR cycling conditions were 1 minute at 98°C, 35 cycles of 98°C for 10 seconds, annealing for 30 seconds, extension at 72°C, and a final extension at 72°C for 5 minutes using Phusion HF Polymerase (Fisher Scientific). Annealing temperature and extension time varied per amplicon according to appendix Table 11. Long range amplicons were purified then treated with 2.5U T4 Polynucleotide Kinase (NEB) in 1x T4 DNA Ligase Buffer at 37°C for 30 minutes. 100U T4 DNA Ligase (NEB) was added directly to the reaction which was incubated for 2 hours at 25°C for circularization of fragments. Ligation products were purified then amplified. PCR cycling conditions were 2 minutes at 98°C, 35 cycles of 98°C for 10 seconds, annealing for 30 seconds, extension at 72°C for 30 seconds and a final extension at 72°C for 5 minutes. Annealing temperatures varied per amplicon according to Appendix Table 12. PCR products were purified. All purification steps were done using 1x volume of AmpureXP Beads (Beckman Coulter) with 80% ethanol washes. A portion of each final PCR product was run on a gel to confirm amplification of the correct product and to estimate relative concentrations. Per sample, each PCR product was combined based on relative

concentrations. Pooled PCR products were gel purified, cutting fragments 100-400bp. Min-Elute Gel Extraction Kit (Qiagen) was used for cleanup.

Sequencing Library Preparation

Gel purified PCR products were quantified using the 2100 Bioanalyzer (Agilent) and 81-300ng of pooled products were used as input for library prep. Amplicon libraries were prepared using NEBNext Ultra Library Prep Kit as per protocol except Illumina TruSeq Indexed Adaptors were used during adaptor ligation and final libraries were amplified 12 cycles using Phusion HF Polymerase (Fisher Scientific) and custom primers: forward 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA-3' and reverse 5'-CAAGCAGAAGACGGCATACGAGAT-3'. Libraries were quantified via real time PCR and via bioanalyzer (Agilent).

Genotyping of Plasma DNA

Plasma was separated from whole blood via centrifugation at 1,600 x g for 10 minutes, followed by a second centrifugation to remove contaminating nucleated cells at 16,000 x g for 10 minutes. DNA was extracted from 3.6ml of plasma using the QIAamp DNA Blood Mini kit (Qiagen) as per protocol. Amplicon libraries were prepared using NEBNext Ultra Library Prep Kit as per protocol except custom adaptors were used during adaptor ligation. An equimolar mixture of AGCCGCCTTTCCATAGTGAGTCCTTCGTCTGTGACTAACTGTGCCAAATCGTCTAGC AA*T (* = phosphorothiate bond) and 5' phos-TTGCTAGACGATCAATTTGACCTANNNNNNNNNNNNAGTCGACGGTTAGAACCTAG GGTC oligos were combined and heated for 5 minutes at 95°C then were slowly cooled to room temperature. Final adaptor concentration in ligation mixture was 5nM. Libraries were amplified 12 cycles using Phusion HF Polymerase (Fisher Scientific) and custom primers: forward 5'-

AGCCGCCTTTCCATAGTGAGTCCTTCGT-3' and reverse 5'-GACCCTAGGTTCTAACCGTCGACT-3'. Amplified libraries were purified then treated with 2.5U T4 Polynucleotide Kinase (NEB) in 1x T4 DNA Ligase Buffer at 37°C for 30 minutes. 400U T4 DNA Ligase (NEB) was added directly to the reaction which was incubated for 2 hours at 25°C for circularization of fragments. Ligation products were purified then amplified. Inverse PCR primers for each of 24 loci were combined in equimolar ratios and final total primer concentration was 12μM in the PCR. Sequences for plasma inverse PCR primers are located in Appendix Table 13. PCR cycling conditions were 2 minutes at 98°C, 35 cycles of 98°C for 10 seconds, 60°C for 30 seconds, extension at 72°C for 30 seconds and a final extension at 72°C for 5 minutes. PCR products were purified. Libraries were prepared for sequencing according to "Sequencing Library Preparation" protocol stated previously. All purification steps were done using 1x volume of AmpureXP Beads (Beckman Coulter) with 80% ethanol washes.

Sequencing and Bioinformatics

All libraries were sequenced using an Illumina Miseq with 150bp paired-end reads. Reads for haplotyping were quality and adaptor trimmed using TrimGalore[47]. Reads were trimmed to 75bp in order to remove the junction point of the adjoining ends of the long-range PCR amplicon using the Fastx Toolkit [48]. Reads were aligned in single-end mode to a custom fasta reference file containing only the amplicons of interest using Bowtie and allowing no mismatches and keeping only uniquely mapped reads[49]. Aligning with no mismatches allowed for exact assignment of the allele and only uniquely mapped reads would contain the allele because both sequences, which only varied at the mutation, were present in the fasta reference file. Single end reads were merged keeping only pairs that mapped uniquely and with no mismatches from both ends.

Sanger Sequencing

25ng of fetal DNA were used for each PCR reaction. The PCR cycling conditions were 30 seconds at 98°C, 35 cycles of 98°C for 10 seconds, annealing for 30 seconds, extension at 72°C, and a final extension at 72°C for 5 minutes using Phusion HF Polymerase (Fisher Scientific). Annealing temperature varied per amplicon according to Appendix Table 15. Products were gel purified using MinElute Gel Extraction Kit (Qiagen).

4.0 RESULTS

4.1 SPECIFIC AIM 1

13 SNPs were chosen for haplotyping and a total of 42 amplicons were designed to inter-connect the SNPs. Each long range PCR and inverse PCR reaction was optimized for annealing temperature and extension time (Figure 5).

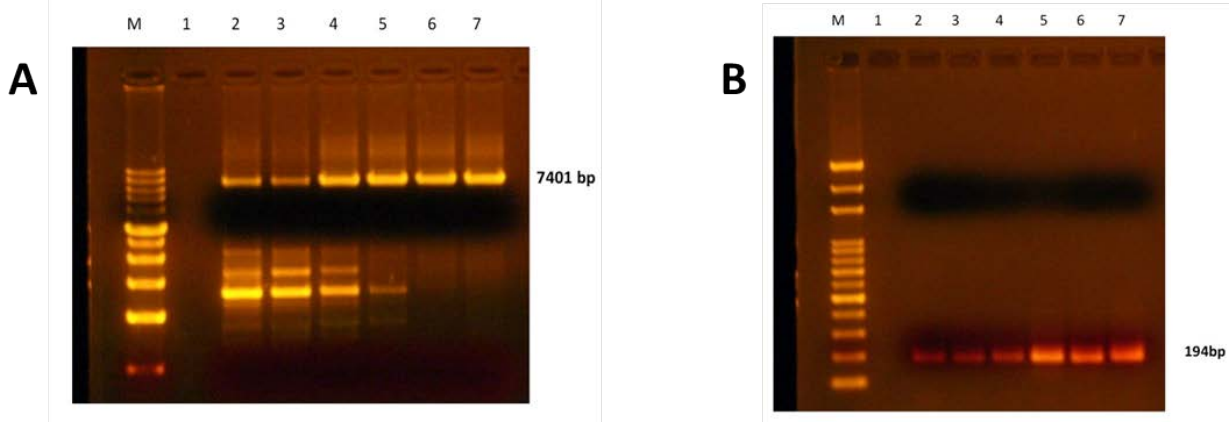

Figure 5: Example long range and inverse PCR annealing profile.

Example long range and inverse PCR annealing profile. PCR annealing profile of A) long-range PCR amplicon (7401bp) and B) corresponding inverse PCR amplicons 194bp). M = 100bp DNA ladder. Lanes 2-7 contain PCR reactions run at increasingly higher annealing temperatures.

Six trios consisting of maternal, paternal and fetal genomic DNA were sequenced on an Illumina Miseq with 150bp paired-end reads. After trimming, alignment and read merging an average of 342,651 pairs (41.13%) per sample were kept for further analysis. Detailed sequencing statistics can be found in Appendix Table 16. The average depth at each SNP site varied (Table 3) drastically, from 196X-122,159X, due to PCR amplification efficiency and the number of amplicons that were designed to cover a particular SNP. Exact concentrations were not calculated, because of cost and time, for each amplicon prior to pooling so it was expected that the reads for each amplicon would vary. Additionally, some SNPs had more amplicons covering them because of limitations of the surrounding genomic sequence for PCR design. SNPs rs2082056, rs213965, and rs213971 had the lowest average depths at 196X, 214X, and 1040X respectively. The low depths at these 3 SNPs indicate a very poor PCR efficiency of the amplicons covering those SNPs versus the PCR efficiencies of the 10 others. Further optimization would be needed to either redesign these amplicons or to remove those SNPs from the design and choose others. All SNPs still contained enough depth to determine genotype and to assess the haplotyping method.

Table 2: Sample information.

Trio Number	Sample Name	Sample Type	delF508 status	Reported Ethnicity
1	G1342	Maternal blood	Non-carrier	White
	G1343	Paternal blood	Non-carrier	
	12-032 CVS	Cultured CVS	Non-carrier	
2	G1687	Paternal blood	Carrier	
	G1688	Maternal blood	Carrier	White
	13-014	Cultured amnio	carrier	
3	G1700	Maternal blood	Carrier	Multiple
	G1701	Paternal blood	Carrier	
	13-019	Cultured amnio	Non-carrier	
4	G1747	Maternal blood	Non-carrier	White
	G1748	Paternal blood	Non-carrier	White
	13-049 CVS	Cultured CVS	Non-carrier	
5	G1831	Maternal blood	Non-carrier	White
	G1832	Paternal blood	Non-carrier	
	13-089 CVS	Cultured CVS	Non-carrier	
6	G1841	Maternal blood	Non-carrier	White
	G1842	Paternal blood	Non-carrier	
	13-093 CVS	Cultured CVS	Non-carrier	

Table 3: Read depth at targeted sites.

SNP	Average Read Depth Across Samples	Number of Amplicons Designed to Cover SNP
rs213950 G/A	122,159	5
rs113993960 CTT/-	60,604	5
rs1896887 G/C	55,262	5
rs718829 C/A	107,084	7
rs2082056 T/C	196	8
rs213934 T/C	115,637	9
rs2518873 C/T	19,302	7
rs213955 A/G	31,458	8
rs6949974 C/T	36,024	8
rs17547853 G/A	65,762	7
rs213965 T/A	214	6
rs213971 G/A	1,040	5
rs1042077 T/G	66,677	4

Genotypes were assigned based on allele frequencies of the properly aligned reads. Assigned genotypes for each sample can be found in Tables 4-6. Detailed allele counts for each sample can be found in Appendix Table 17. The genotypes of each of the fetal samples were consistent with the genotypes of the parents in every case. The genotype at the delF508 mutation (rs113993960) was also consistent with the clinical data received from each patient. Sites that were homozygous in each sample could be linked as part of a haplotype. For example, G1343 is homozygous at all sites genotyped so the haplotype can be assigned (see Table 8) without any further analysis.

Table 4: Assigned genotypes in trios 1 and 2.

	Trio 1			**Trio 2**		
	G1342	**G1343**	**12-032**	**G1688**	**G1687**	**13-014**
rs213950 G/A	a/a	g/g	g/a	g/a	g/a	g/a
rs113993960 CTT/-	ctt/ctt	ctt/ctt	ctt/ctt	ctt/-	ctt/-	ctt/-
rs1896887 G/C	g/c	g/g	g/g	g/g	g/g	g/g
rs718829 C/A	a/a	c/c	c/a	c/a	c/a	c/a
rs2082056 T/C	c/c	t/t	t/c	t/c	t/c	t/c
rs213934 T/C	c/c	t/t	t/c	t/c	t/c	t/c
rs2518873 C/T	t/t	c/c	c/t	c/c	c/c	c/c
rs213955 A/G	g/g	a/a	a/g	a/g	a/g	a/g
rs6949974 C/T	c/t	c/c	c/t	c/c	c/c	c/c
rs17547853 G/A	g/g	g/g	g/g	g/g	g/g	g/g
rs213965 T/A	a/a	t/t	t/a	t/a	t/a	t/a
rs213971 G/A	a/a	g/g	g/a	g/a	g/a	g/a
rs1042077 T/G	t/g	t/t	t/g	t/t	t/t	t/t

Table 5: Assigned genotypes in trios 3 and 4.

	Trio 3			Trio 4		
	G1700	**G1701**	**13-019**	**G1747**	**G1748**	**13-049**
rs213950 G/A	g/a	g/a	g/g	g/a	g/a	g/a
rs113993960 CTT/-	ctt/-	ctt/-	ctt/ctt	ctt/ctt	ctt/ctt	ctt/ctt
rs1896887 G/C	g/g	g/g	g/g	g/g	g/g	g/g
rs718829 C/A	c/a	c/a	c/c	c/a	c/a	c/a
rs2082056 T/C	t/c	t/c	t/t	t/c	t/c	t/c
rs213934 T/C	t/c	t/c	t/t	t/c	t/c	t/c
rs2518873 C/T	c/c	c/c	c/c	c/t	c/t	c/t
rs213955 A/G	a/g	a/g	a/a	a/a	a/a	a/a
rs6949974 C/T	c/c	c/c	c/c	c/t	c/t	c/t
rs17547853 G/A	g/g	g/g	g/g	g/g	g/a	g/a
rs213965 T/A	t/a	t/a	t/t	t/a	t/a	t/a
rs213971 G/A	g/a	g/a	g/g	g/a	g/a	g/a
rs1042077 T/G	t/t	t/t	t/t	t/g	t/g	t/g

Table 6: Assigned genotypes in trios 5 and 6.

	Trio 5			Trio 6		
	G1831	**G1832**	**13-089**	**G1841**	**G1842**	**13-093**
rs213950 G/A	g/a	g/a	g/a	a/a	a/a	a/a
rs113993960 CTT/-	ctt/ctt	ctt/ctt	ctt/ctt	ctt/ctt	ctt/ctt	ctt/ctt
rs1896887 G/C	g/g	g/g	g/g	g/g	g/g	g/g
rs718829 C/A	c/a	c/a	c/a	a/a	a/a	a/a
rs2082056 T/C	t/c	t/c	t/c	c/c	c/c	c/c
rs213934 T/C	t/c	t/c	t/c	c/c	c/c	c/c
rs2518873 C/T	c/t	c/t	c/t	t/t	t/t	t/t
rs213955 A/G	a/a	a/g	a/a	a/a	a/g	a/a
rs6949974 C/T	c/t	c/t	c/t	t/t	t/t	t/t
rs17547853 G/A	g/g	g/g	g/g	a/a	g/a	a/a
rs213965 T/A	t/a	t/a	t/a	a/a	a/a	a/a
rs213971 G/A	g/a	g/a	g/a	a/a	a/a	a/a
rs1042077 T/G	t/g	t/g	t/g	g/g	g/g	g/g

Within each trio, counts were assigned to each possible combination of heterozygous loci. Table 7 shows the counts for trio 4 at the sites that were heterozygous. There are 4 possible categories for 2 heterozygous loci: ref-ref, ref-alt, alt-ref, alt-alt. However, only 2 combinations are biologically possible: 1) ref-ref/alt-alt or 2) ref-alt/alt-ref. What the counts consistently showed across all samples was that almost all 4 categories were present, meaning that the reference allele at one site was associated with both the reference and the alternate allele at the adjoining site. For example in sample G1747 between SNPs rs213950 and rs718829 there was a total of 46,414 counts. The expectation would be that close to 100% of the counts would be split relatively evenly in the ref-ref and alt-alt category with very close to 0% being in the opposite categories or vice versa with the counts in the ref-alt and alt-ref categories. Instead the data is showing that the reference allele for SNP rs213950 is associating with the reference allele rs718829 33.44% of the time, but is also associating with the rs718829 alternate allele 22.92% of the time. Similar results are seen for the alternate allele of rs213950 and rs718829.With few exceptions, all amplicons showed this result and were inconclusive. There were very few instances where the counts appeared to favor a particular SNP pairing. For example in sample G1747 between SNPs rs213950 and rs213934 the counts favored a pairing of rs213950 reference with rs213934 reference and a pairing of the alternate alleles. 42.63% of the time rs213950 reference associated with rs213934 reference and the alternate alleles associated with each other 44.34% of the time. These values are much closer to the expected 50% distribution between categories. Given that almost every other amplicon showed inconclusive results, this “positive” result is most likely just random and should not be used to construct the haplotype. Counts from the remaining trios can be found in Appendix Tables 18-22.

Table 7: Haplotype category counts at heterozygous loci for trio 4.

SNP Pair	Category	G1747		G1748		13-014	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213950_rs718829	ref-ref	15520	33.44%	17317	36.15%	8834	27.80%
	ref-alt	10640	22.92%	10802	22.55%	7559	23.79%
	alt-ref	10335	22.27%	10509	21.94%	7199	22.65%
	alt-alt	9919	21.37%	9272	19.36%	8186	25.76%
rs213950_rs213934	ref-ref	17166	42.63%	3381	23.00%	2048	27.76%
	ref-alt	2807	6.97%	4650	31.63%	2192	29.71%
	alt-ref	2439	6.06%	2532	17.22%	1336	18.11%
	alt-alt	17856	44.34%	4138	28.15%	1801	24.41%
rs718829_rs213934	ref-ref	6108	25.57%	4369	24.84%	11213	28.05%
	ref-alt	6440	26.96%	5180	29.45%	11613	29.05%
	alt-ref	5171	21.65%	3645	20.72%	8321	20.82%
	alt-alt	6164	25.81%	4396	24.99%	8827	22.08%
rs718829_rs2518873	ref-ref	1228	30.76%	363	29.42%	1971	31.36%
	ref-alt	625	15.66%	209	16.94%	1071	17.04%
	alt-ref	1454	36.42%	392	31.77%	2058	32.74%
	alt-alt	685	17.16%	270	21.88%	1185	18.85%
rs2082056_rs213934	ref-ref	sample homozygous		15	13.76%	28	21.21%
	ref-alt			11	10.09%	17	12.88%
	alt-ref			38	34.86%	43	32.58%
	alt-alt			45	41.28%	44	33.33%
rs213934_rs2518873	ref-ref	2260	36.45%	2250	31.84%	2690	37.42%
	ref-alt	911	14.69%	1043	14.76%	1159	16.12%
	alt-ref	2088	33.67%	2512	35.55%	2188	30.44%
	alt-alt	942	15.19%	1262	17.86%	1152	16.02%
rs213934_rs17547853	ref-ref	sample homozygous		1785	23.61%	3261	66.73%
	ref-alt			1063	14.06%	165	3.38%
	alt-ref			1565	20.70%	252	5.16%
	alt-alt			3146	41.62%	1209	24.74%
rs2518873_rs17547853	ref-ref	sample homozygous		643	46.09%	463	47.20%
	ref-alt			358	25.66%	224	22.83%
	alt-ref			188	13.48%	156	15.90%
	alt-alt			206	14.77%	138	14.07%
rs6949974_rs17547853	ref-ref	sample homozygous		2414	29.21%	2039	33.28%
	ref-alt			1919	23.22%	1412	23.05%
	alt-ref			2230	26.98%	1549	25.28%
	alt-alt			1702	20.59%	1127	18.39%

Table 7 (continued)

SNP Pair	Category	G1747		G1748		13-014	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs6949974_rs1042077	ref-ref	sample homozygous		7631	32.48%	9023	43.60%
	ref-alt			4143	17.64%	4379	21.16%
	alt-ref			4401	18.73%	4594	22.20%
	alt-alt			7317	31.15%	2701	13.05%
rs17547853_rs213971	ref-ref	sample homozygous		66	41.51%	35	39.33%
	ref-alt			41	25.79%	28	31.46%
	alt-ref			27	16.98%	13	14.61%
	alt-alt			25	15.72%	13	14.61%
rs17547853_rs1042077	ref-ref	sample homozygous		1286	20.78%	2234	27.18%
	ref-alt			1410	22.78%	2720	33.09%
	alt-ref			1019	16.46%	1509	18.36%
	alt-alt			2474	39.97%	1756	21.37%
rs213971_rs1042077	ref-ref	sample homozygous		147	32.67%	70	31.25%
	ref-alt			126	28.00%	60	26.79%
	alt-ref			83	18.44%	43	19.20%
	alt-alt			94	20.89%	51	22.77%

Each row shows the 2 linked loci and the counts associated connecting the reference or alternate alleles at each locus. If the 2 linked loci are homozygous at either linked site then counts are not given and “sample homozygous” is indicated.

The haplotyping method proposed did not produce conclusive results to link heterozygous loci for a complete haplotype. However, using the homozygous alleles and the assumed inheritance within the trios, 3 of the 6 trios could be completely haplotyped across the 13 loci. Table 8 shows the haplotypes determined. In table 8, each column represents 1 of the 13 targeted loci. Each row corresponds to a haplotype for a given sample with each sample having two haplotypes. The reference allele for rs113993960 (delF508) is “ctt” while the alternate (mutation) allele is the deletion of “ctt” denoted as “-“. For example, paternal sample G1343 of trio 1 was homozygous at all the loci meaning he has 2 copies of the same haplotype. G1343 is not a carrier for the delF508 mutation. Homozygous sites for sample G1342 and 12-032 are also filled in with lower case letters. It is expected that the fetus would inherit one

haplotype from the mother and one from the father. Since the paternal sample haplotype is known, 1 of the fetal haplotypes can be assumed and those sites are filled in with **bold capitalized** letters. The sites remaining in sample 12-032 are heterozygous, so whatever allele was not part of the paternal haplotype is part of the inherited maternal haplotype. Both fetal haplotypes are known and the maternal haplotypes in G1342 can be deduced based on the haplotype that the fetus inherited. Similar logic was followed to haplotype trios 3 and 6 starting with the completely homozygous sample within each trio. Trios 2, 4, and 5 could not be completely haplotyped due to the failure of the method to conclusively link heterozygous sites and the lack of homozygous sites to assume inheritance.

Table 8: Assigned haplotypes.

	rs213950	**rs113993960 (delF508)**	**rs1896887**	**rs718829**	**rs2082056**	**rs213934**	**rs2518873**	**rs213955**	**rs6949974**	**rs1754785**	**rs213965**	**rs213971**	**rs1042077**
Trio 1													
G1342	a	ctt	**G**	a	c	c	t	g	**T**	g	a	a	**G**
	a	ctt	**C**	a	c	c	t	g	**C**	g	a	a	**T**
G1343	g	ctt	g	c	t	t	c	a	c	g	t	g	t
	g	ctt	g	c	t	t	c	a	c	g	t	g	t
12_032	**G**	ctt	g	**C**	**T**	**T**	**C**	**A**	**C**	g	**T**	**G**	**T**
	A	ctt	g	**A**	**C**	**C**	**T**	**G**	**T**	g	**A**	**A**	**G**
Trio 2													
G1688			g				c		c	g			t
			g				c		c	g			t
G1687			g				c		c	g			t
			g				c		c	g			t
13_014			g				c		c	g			t
			g				c		c	g			t

Table 8 (continued)

	rs213950	rs113993960 (delF508)	rs1896887	rs718829	rs2082056	rs213934	rs2518873	rs213955	rs6949974	rs1754785	rs213965	rs213971	rs1042077
Trio 3													
G1700	**G**	ctt	g	**C**	**T**	**T**	c	**A**	c	g	**T**	**G**	t
	A	-	g	**A**	**C**	**C**	c	**G**	c	g	**A**	**A**	t
G1701	**G**	ctt	g	**C**	**T**	**T**	c	**A**	c	g	**T**	**G**	t
	A	-	g	**A**	**C**	**C**	c	**G**	c	g	**A**	**A**	t
13-019	g	ctt	g	c	t	t	c	a	c	g	t	g	t
	g	ctt	g	c	t	t	c	a	c	g	t	g	t
Trio 4													
G1747		ctt	g					a		g			
		ctt	g					a		g			
G1748		ctt	g					a					
		ctt	g					a					
13-049		ctt	g					a					
		ctt	g					a					
Trio 5													
G1831		ctt	g					a		g			
		ctt	g					a		g			
G1832		ctt	g							g			
		ctt	g							g			
13-049		ctt	g					a		g			
		ctt	g					a		g			
Trio 6													
G1841	a	ctt	g	a	c	c	t	a	t	a	a	a	g
	a	ctt	g	a	c	c	t	a	t	a	a	a	g
G1842	a	ctt	g	a	c	c	t	**A**	t	**A**	a	a	g
	a	ctt	g	a	c	c	t	**G**	t	**G**	a	a	g
13-093	a	ctt	g	a	c	c	t	a	t	a	a	a	g
	a	ctt	g	a	c	c	t	a	t	a	a	a	g

Assigned haplotypes. The row names represent the 2 haplotypes for each sample and the column names are the loci. Homozygous sites are filled in by sample in lower case letters. **Bold Capitalized** letters represent heterozygous sites that could be phased due to inheritance.

4.2 SPECIFIC AIM 2

The targeted loci are all located within the first 108 Kb haplotype block in Figure 2 of the CFTR gene. Figure 6 shows a 38Kb sub-region of haplotype block 1. Within this sub-region there are 2 blocks, denoted 1A and 1B, of linkage disequilibrium (LD). The alleles in block 1A are more likely to be inherited together and the alleles in 1B are also more likely to be inherited together. If a recombination event were to happen within this sub-region, the most likely location would be between blocks 1A and 1B which splits SNPs 13 and 15 into 2 separate blocks of LD. This assay only included SNPs 3, 4, 8, 12, 13, 17, 18 from Figure 6 so the other SNPs are blacked out. Based on the order of haplotypes in Figure 6, a number (1-15) was arbitrarily assigned to each haplotype. In Figure 6, the lines between the blocks represent the most likely continuous haplotype with the weight of the line indicating the most frequent combination of the 2 blocks if more than 1 is observed. Where there is no line, such as for haplotypes 6-15, no information was provided from Hapmap. For example the most likely continuous haplotype for haplotype 1 across the blocks would be G/G/T/G/T/G/T. Any haplotype that could not be distinguished from the SNPs chosen in this assay are noted by connecting arrows. For instance, based on the 5 SNPs in block 1A, haplotypes 2 and 15 could not be distinguished from one another and it is marked 2/15 in Table 9.

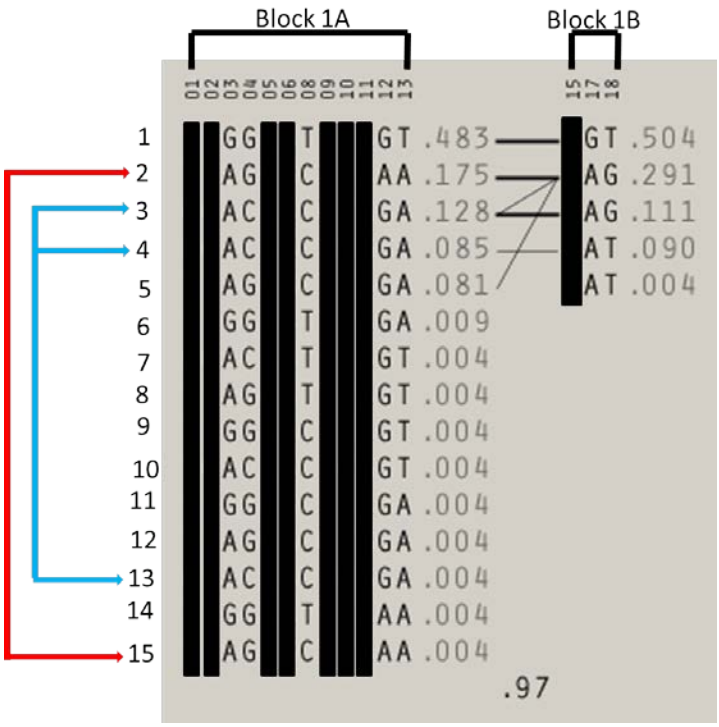

Figure 6: CEU haplotypes in a 38Kb sub-region of CFTR.

SNPs not targeted in this assay are blacked out. Each haplotype in block 1A is given a number 1-15. Haplotypes connected by arrows in left margin are indistinguishable from one another based on loci included[43].

Each column of Table 9 shows the SNP and its corresponding Hapmap number (see Figure 6). 5 of the 13 SNPs targeted by this assay were not in Hapmap while 8 of the 13 were. The SNPs that weren't in Hapmap are placed in order. For instance, the delF508 mutation (rs113993960) falls between SNPs 3 and 4. Each row of Table 9 shows the haplotype (1-15) from Figure 6. Because of the ambiguity of the continuous haplotypes between block 1A and block 1B of Figure 6 for all 15 haplotypes, SNPs 17 and 18, rs213971 and rs1042077 respectively, are not included. The resulting haplotypes were compared to the known haplotypes from the CEU population in Hapmap and the loci not in Hapmap were assigned to specific

haplotypes. In Table 9, the last column on the right shows the number of times this haplotype was seen in the 18 samples in this data set. For instance, haplotype 1 was observed 7 times.

Table 9: Summary of observed haplotypes and corresponding Hapmap haplotypes.

	(3) rs213950	rs113993960	(4) rs1896887	rs718829	rs2082056	(8) rs213934	rs2518873	(11) rs213955	rs6949974	(12) rs17547853	(13) rs213965	Observed
2/15	A		G			C		A		A	A	5x
5	A		G			C		A		G	A	
12	A	ctt	G	a	c	C	t	G	t	G	A	3x
12a	A	-	G	a	c	C	c	G	c	G	A	2x
8	A		G			T		A		G	T	
3/4/ 13	A		C			C		G		G	A	1x
10	A		C			C		G		G	T	
7	A		C			T		A		G	T	
1	G	ctt	G	c	t	T	c	A	c	G	T	7x
6	G		G			T		A		G	A	
14	G		G			T		A		A	A	
9	G		G			C		A		G	T	
11	G		G			C		A		G	A	

Table is derived from Figure 6 and includes the SNPs that were covered in the assay. Numbers in parenthesis across the top reflect the SNP number from Figure 6 that was given in Hapmap. The observed column shows how many times this haplotype was seen in our data set. Alleles between the Hapmap SNPs covered in our assay are filled in to complete the haplotype where possible.

As seen in Table 9, two distinct versions of haplotype 12 were seen in our data set. These varied at 2 SNP sites that were not part of the Hapmap data. Two unrelated individuals, G1700 and G1701, are delF08 carriers and the haplotype that the mutation corresponds to in the Hapmap data is haplotype 12. However, two other unrelated individuals, G1342 and G1842, also have haplotypes that correspond to haplotype 12. They are not delF508 carriers. There is a

subtype of haplotype 12 that carries the delF508 mutation. The mutation carrying version of haplotype 12 will be called 12a.

Two fetal samples that are homozygous for the delF508 mutation were sanger sequenced at 12 of the 13 sites. The delF508 site was not sequenced in these samples as clinical testing already confirmed the deletion. Both fetal samples were homozygous for the same haplotype. This haplotype corresponds with haplotypes 3/4/13 from Table 9. It is identical to one of the haplotypes found in sample G1342 except at the delF508 site where G1342 does not carry the delF508 mutation.

Table 10: Haplotypes from homozygous delF508 samples.

	rs213950	**rs113993960**	**rs1896887**	**rs718829**	**rs2082056**	**rs213934**	**rs2518873**	**rs213955**	**rs6949974**	**rs17547853**	**rs213965**	**rs213971**	**rs1042077**
11-067	A	-	C	A	C	C	T	G	C	G	A	A	T
12-243	A	-	C	A	C	C	T	G	C	G	A	A	T

4.3 SPECIFIC AIM 3

Two libraries were made from 3.6ml of plasma each. Plasma samples were not related to the trios. Products from plasma were checked throughout the preparation to confirm that the protocol was working as intended. Figure 7A shows the plasma DNA post-ligation with custom adaptors with a unique molecular identifier after amplification. After circularization, targeted amplification was performed via a multiplex inverse PCR (Figure7B). The resulting circularization products are various sizes due to the random fragmentation of plasma DNA. The products from the inverse PCR are also random because the SNP can fall anywhere within the circle. After inverse PCR, Illumina compatible adaptors were ligated to the products and the library was amplified. The addition of the final adaptors adds 120bp to the fragment. A fragment size shift of approximately 120bp can be seen in the final products (Figure 7C).

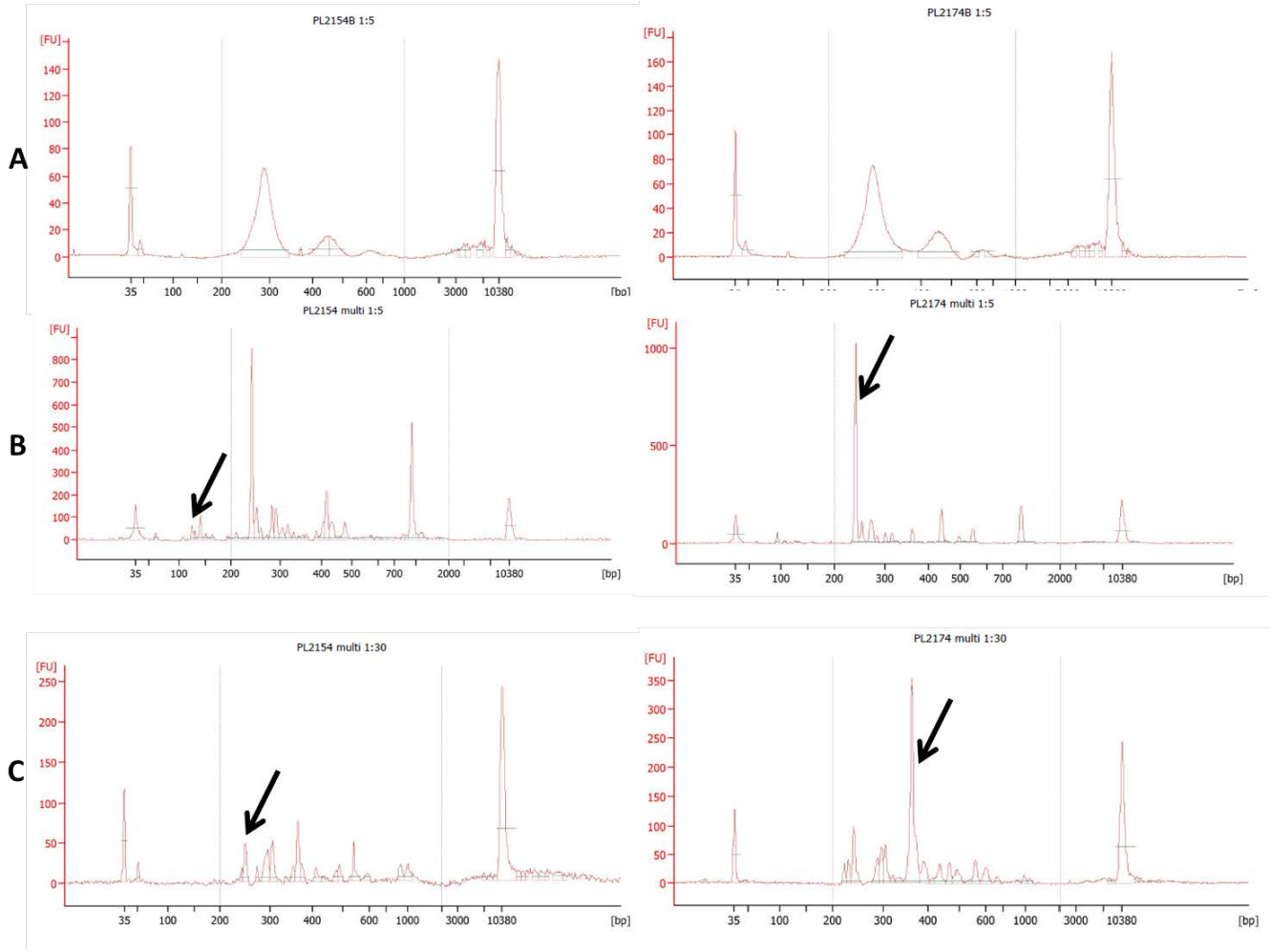

Figure 7: Bioanalyzer traces of plasma libraries throughout preparation.

Paired end reads generated for the 2 plasma libraries were 926,819 and 1,149,907. Further bioinformatics analysis is ongoing and nothing further can be presented to evaluate this method for counting allele frequency in plasma DNA.

5.0 DISCUSSION

The aim of this project was to develop a new method for haplotyping pure paternal and maternal samples to aid in the assignment of the fetal haplotype after sequencing the maternal plasma DNA.

Six trios comprising maternal, paternal, and fetal genomic DNA underwent a novel library preparation method and were sequenced in order to haplotype a 35.5Kb region of the CFTR gene which includes the delF508 mutation. All genotyping results within the trios were consistent with presumed inheritance among the trios. After analyzing the counts reflecting the linkage of the loci, it is clear that the method did not perform as intended. The counts showed that between 2 heterozygous sites, any given allele at one site was associated with both of the alleles at the linked site; a biological impossibility. One distinct possibility that would lead to these results is if prior to circularization, individual long range amplicons ligated to each other forming a concatemer. The resulting PCR product from the inverse PCR of a concatemer is the same size as the PCR product from the proper intra-molecular ligation. Figure 8 outlines the process by which these PCR products connecting 2 loci that are not originally connected might occur.

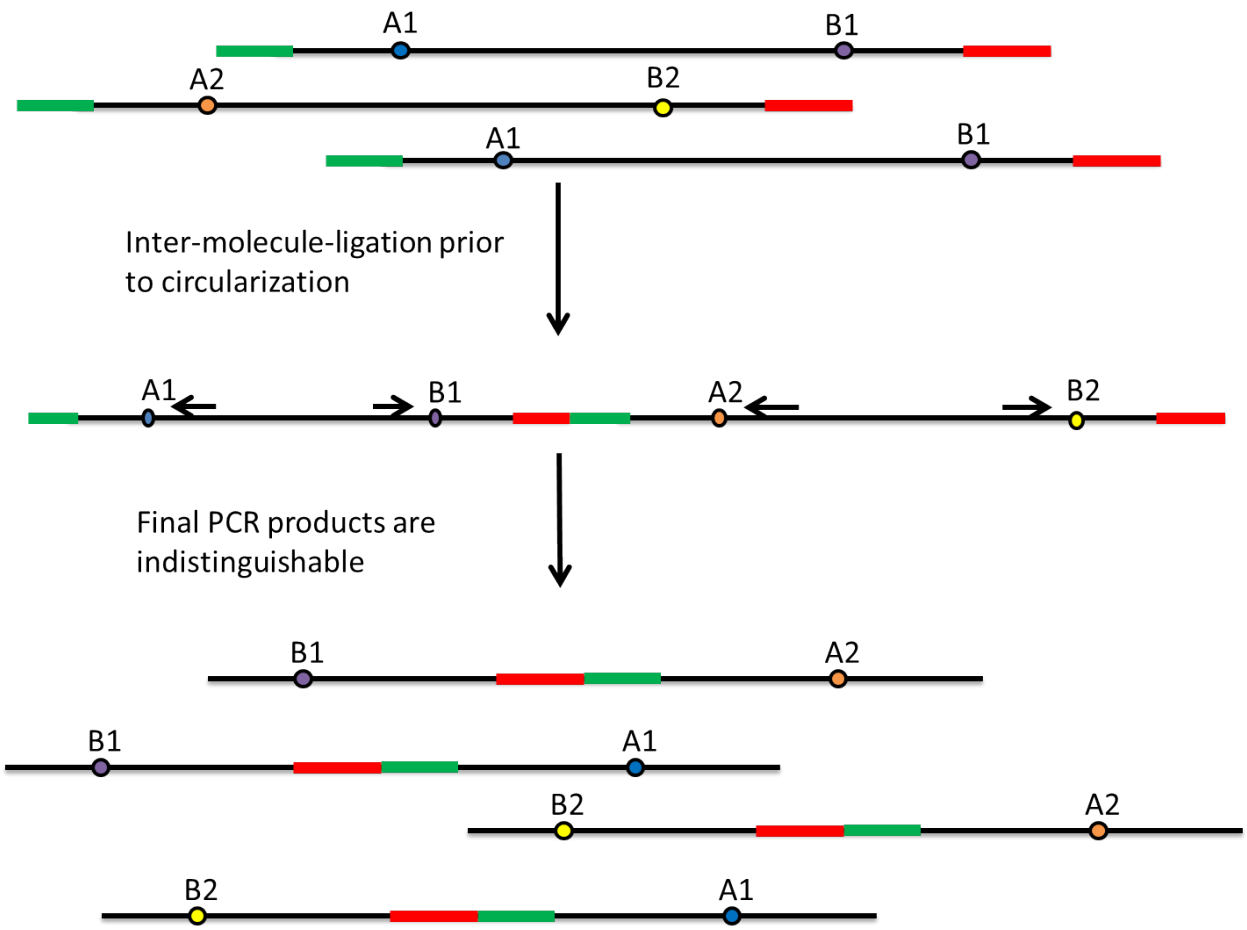

Figure 8: Formation of inverse PCR products not reflecting original linked loci.

In order for this approach to work for haplotyping, it is necessary to minimize or completely remove the possibility of intermolecular ligation prior to circularization. Ligation by T4 DNA ligase can result in both inter and intra molecule ligation. Dugaiczyk et al showed that this was dependent on the effective concentration of one end of the DNA molecule to the other (j) and the effective DNA concentration(i)[50]. At a j:i ratio of 2 or higher, circular products were favored[50]. They proposed that the most effective way to promote circularization was to minimize i and lower the DNA concentration[50]. Since this method also relies on a range of long range amplicon sizes, this j:i ratio would be different for each as the length of DNA is

factored into variable j so potentially each reaction would have to be optimized separately[50]. Shoaib et al also found that low DNA concentrations led to self-circularization[51]. In their experiments, with concentrations of less than 1ng/ul, self-circularization was favored[51]. I also observed that intermolecular ligation was enhanced by increased concentrations of divalent cations[52]. T4 DNA ligase buffer contains the divalent cation $MgCl_2$. Possibly decreasing the concentration of $MgCl_2$ could limit the amount of intermolecular ligation.

A better approach would be to remove the chance of formation of concatemers from the reaction. This could be achieved by physically separating the long range amplicons prior to ligation so that only 1 molecule is present. Similar to digital PCR workflows, the long range amplicons would be diluted to a point where only one molecule would be placed in a well on a plate. However, this approach would add significant time and cost to the protocol. Another option may be to use emulsions. An individual molecule, in this case the long range PCR amplicon, would be put into an individual droplet. This is very similar to current technology in droplet digital PCR machines except that instead of performing a PCR within the droplet, the ligation reaction would occur. At the present time, this is not technically feasible and would likely require very specific machinery to perform.

From the homozygous loci and assumed inheritance within the trios, 3 of the 6 trios could be haplotyped across all 13 loci. All of the assigned haplotypes were consistent with at least one of the haplotypes seen in the Hapmap CEU population (Figure 6). Trio 3 was the only completely haplotyped trio that contained samples who were delF508 mutation carriers. The maternal (G1700) and paternal (G1701) samples were carriers for the delF508 mutation and the CEU haplotype that was consistent with the observed haplotypes was 12. However, a haplotype that was consistent with haplotype 12 was also seen in trio 1, samples G1342 and 12-032. When

the number 12 haplotypes are compared between trios 1 and 3 they vary at two sites: rs2518873 and rs6949974. The version of haplotype 12 containing the delF508 mutation will be denoted as 12a and the non-deletion version will be denoted as 12. Since two unrelated individuals carried the delF508 mutation on 12a, I wanted to explore if this was also the haplotype carried by two delF508 homozygotes. Sanger sequencing was performed on 2 cultured amniocyte DNA samples, 11-067 and 12-243 to genotype the 12 SNPs surrounding the delF508 mutation. Both of these samples were homozygous for the same haplotype. Instead of matching the previous haplotype 12a, they both carried a haplotype that was identical to one carried by G1342 which was consistent with CEU haplotype 3, 4, or 13. Due to the SNPs chosen, the haplotypes 3, 4 and 13 could not be distinguished from one another. This means that the delF508 mutation could occur on 3, 4 or 13. Further genotyping at sites not originally covered in this assay, but within Hapmap, on samples in the 3/4/13 haplotype category could be done in order to elucidate if any one of these three might harbor the delF508 mutation. The results here indicate that the delF508 mutation can be found on at least 2 haplotypes and is not associated with just one haplotype.

In 2015, Zeevi et al published a study which they used consensus haplotype-mediated NIPD to non-invasively detect the N370S mutation causing Gaucher disease in the Ashkenanazi Jewish population in plasma [53]. They used next generation sequencing to perform homozygosity mapping around the GBA gene for 7 unrelated individuals who were homozygous for the N370S mutation [53]. Using the consensus sequence of homozygous unrelated individuals would remove the requirement for haplotyping the parents because a universal haplotype containing the mutation would be used. This would save significant time and cost. For delF508 detection, this approach would require finding the haplotype(s) associated with the deletion. In the N370S study, they were able to identify one founder haplotype carrying the

mutation among the homozygous individuals [53]. From this small delF508 experiment it is clear that at least 2 haplotypes are associated with the mutation and not just one. The N370S mutation is a young mutation compared to the delF508 mutation. The N370S mutation in the Ashkenazi Jewish population is estimated to be about 1000 years old or less whereas the delF508mutation is at least 52,000 years old[54]. With so many generations, it is likely that the delF508 mutation is associated with many more haplotypes than the N370S mutation. It would be feasible to perform homozygosity mapping on homozygous delF508 individuals, but a larger sample size would be needed to detect the likely multiple haplotypes containing the delF508 mutation. This same homozygosity mapping would also be required across multiple worldwide populations for validation that the test could be applied universally.

While the method presented here would detect and utilize the paternal haplotype, it is not actually necessary to haplotype pure paternal DNA. It has been shown that the paternal haplotype can be deduced computationally from the sequencing of the plasma which contains the paternally inherited haplotype[27]. Not having to haplotype pure paternal DNA would save on time and cost, but is also important because the incidence of non-paternity is estimated to be between 3% and 10%[27].

The use of unique molecular identifiers in plasma DNA sequencing has the potential to negate the requirement for parental haplotypes. By individually tagging molecules prior to amplification, the allele frequency of only the “original” molecules can be counted rather than counting all of the PCR copies. Not taking into account the PCR copies may significantly reduce the noise enough that a very slight allelic ratio change could be detected. Analysis is ongoing for that part of this study.

For now, it appears that maternal haplotyping will be necessary to diagnose recessive monogenic disorders in the fetus. The proposed genomic haplotyping method described here is too time and labor intensive for everyday clinical use. Modification of the method to minimize or reduce the potential formation of concatemers may result in a workflow that would be more expensive and also take more time with the technologies available today. Additionally, this design only covered 35.5Kb and increasing the bases targeted to other monogenic disorder genes would need to involve multiplex PCRs that require loads of optimization. The consensus haplotype mediated method is probably not ideal due to the amount of testing across populations that would be required to validate it for just 1 mutation site.

The most promising method moving forward for haplotyping parental genomic DNA would be to utilize longer sequencing reads. The longer the sequence, the more SNPs it will contain to form a haplotype. If the reads are long enough and contain multiple SNPs, complete de novo assembly of the haplotype is possible. Yang et al designed an algorithm called HARSH (Haplotyping with Reference and Sequencing technology) for haplotype phasing which uses a probabilistic model to incorporate the SNP information from the sequencing reads as well as a provided reference panel of haplotypes[55]. To test their algorithm, they simulated sequencing data of 100bp paired-end reads on 1000-4000bp fragments at 1X-10X sequencing coverage[55]. Using their algorithm, they were able to get as accurate results from 100bp paired end reads on 1000bp fragments as they could from 100bp paired end reads on 4000bp fragments de novo assembled[55]. A 4000bp fragment is currently not sequenceable on the Illumina platform, but a 1000bp fragment is. Moving forward with this experiment, a targeted sequencing method using 100bp paired end reads on 1000bp fragments is the next logical step. Briefly, a targeting method, such as Agilent SureSelect, would be used to capture long fragments in the CFTR region. A

custom targeting panel does have high up-front costs, but the sample reagents could potentially be used on the plasma cell-free DNA for targeted sequencing of the same area. Unlike the method proposed here which targets 1 mutation in 1 gene, a targeting panel could be designed to easily cover many genes linked to monogenic disorders and any mutation within that gene.

Another promising technology for long range haplotyping of genomic DNA is the SMRT (Single Molecule, Real-Time) sequencing utilized by PACBIO sequencing systems. Using this system, fragments that are >10,000bp can be sequenced and those long reads could be used for de novo haplotype assembly[56]. Currently this sequencing technology is not readily available for testing.

APPENDIX: TABLES

Table 11: Haplotyping long range and inverse PCR primer sequences.

Primer Name	Sequence
amp_1 LR forward	TTTGATAATGACCTAATAATGATGGGTTTTATTT
amp_1 LR reverse	GATTAAATGCTACTTCTCTAGACTTACATAAGAA
amp_1 circle forward	AAGGCTCCAGTTCTCCCATA
amp_1 circle reverse	CACAATCCCTGTCCTCAAGGA
amp_2 LR reverse	TCATATGCATAATCAAAAAGTTTTCACATAGTT
amp_2 circle reverse	CTGGATTATGCCTGGCACCA
amp_3 LR reverse	GATTATATAAAGGAATTGCAAATGCCAACTATC
amp_3 circle reverse	GTTTATTGACAGTATACTCC
amp_4 LR reverse	AATATAATGGATGTTACAGATTTTGTCACTAAT
amp_4 circle reverse	AGAGACATAAAGATATTAAACAGAGTTACA
amp_5 LR forward	AATTTCATTCTGTTCTCAGTTTTCCTGGATTAT
amp_5 circle forward	GCTTCTGTATCTATATTCATC
amp_8 LR forward	TAAATGTTAGGGCAGGGAATATGTTACTATGAA
amp_8B LR forward	TGTGTGATGGTGGGTTCAGT
amp_8 circle forward	GCCAACTATCAAAGATATTGC
amp_10 LR forward	CAAACTCTAGGAACATATTTGATGACAGAGAC
amp_10B_circle forward	CTTACCAAACTGTACTGGTTTTCA
amp_11 LR reverse	CACAAGGAAAATGGACAAGGGG
amp_11 circle reverse	ACTCTACTCTTCCCTTCCTCTGT
amp_14 LR forward	CTTTATAATACTCTTGGCTCTCTTACGTTCTCT
amp_14 LR reverse	ATTTTCAATCCTGTTACATAATAACAATGG
amp_14B_LR reverse	TTGCCCCACTTTCCAAAGATT
amp_14 circle forward	ATAAGTGTGGTAGGAGCAGAGG
amp_14 circle reverse	TTGCAAATACAGTACATCTAACAAGA
amp_17 LR forward	TTATCTAATGGCCTATAAAATGTGACTTTC
amp_17B_LR forward	CCTTTAGCAATGCTTTCCTCAGT
amp_17 LR reverse	TAAATCCCCACACTTGTTGCTAATTG
amp_17B_circle forward	CTATTTTGCATAGCAGCTGTGGT
amp_17 circle reverse	TTACAGTATCCTGATATTGACTTATCGAG

Table 11 (continued)

Primer Name	Sequence
amp_19 LR forward	GAAGTGCTTGAAGGCAGGTTTCATA
amp_19B_LR forward	CACAGAAGTGCTTGAAGGCAG
amp_19 LR reverse	AAGCAAAAATTGTCAGAAGTAGAAGTAGAT
amp_19B_LR reverse	TGTGTGCAGAGGAGAAAAAGAAAG
amp_19B circle forward	AATATTTTAATAATCAACCCCACAGTAAC
amp_19 circle reverse	TGTCCCCTTCAGTATTCTTGTCC
amp_22 LR forward	GTTCCTTATCTTCTATCTGTTTCTTACTGT
amp_22 LR reverse	AGAACAGTAATAAAGATGAAGACACAGTTC
amp_22 circle forward	TGGAGAAGCAAAAATTGTCAGAAGT
amp_22 circle reverse	GAGAATCCTATGTACTTGAGA
amp_27 LR forward	AACAAATACAAATCTGCACACATACAGAAATAA
amp_27 circle forward	ACAGATTTTGTCACTAATATCAAACACTTA
amp_31 LR reverse	AACAGAAGTATACAATAGATAGAATACCCTAGA
amp_31 circle reverse	GCATTCTTGAAGTCTCAGACCAAA
amp_34 LR reverse	AGTCAGACAGATCTGGATTTGAACC
amp_34 circle reverse	TTTTCACATTTGCAGATAAGGAAACTAAAG
amp_36 LR forward	TTTACTACCTATCTAACTCTTCGCATTCTT
amp_36 circle forward	ATAGAATACCCTAGAGGCTTTCAAAC
amp_37 LR forward	CCAACAATTACATGTATAAACAGAGAATCC
amp_37B_circle forward	GAGATAATTTCTTCATAACATTTAAAT
amp_40 LR reverse	CTTTACAATAGAACATTCTTACCTCTGCCA
amp_40 circle reverse	CCAGCAGTGACTACATGGAACA
amp_44 LR forward	ATATATTTAGTCCCCACAACTACCCTGG
amp_44 circle forward	CAACCTCCACCATTAACATCCTA

Table 12: Long range PCR amplicon properties.

	Forward Long Range Primer	Reverse Long Range Primer	Ext. Time (min)	Annealing Temp.	LR Amplicon Length
amplicon 1	amp_1 LR forward	amp_1 LR reverse	6	68	8480
amplicon 2	amp_1 LR forward	amp_2 LR reverse	0.5	68	269
amplicon 3	amp_1 LR forward	amp_3 LR reverse	0.5	68	1056
amplicon 4	amp_1 LR forward	amp_4 LR reverse	9	64	12462
amplicon 5	amp_5 LR forward	amp_3 LR reverse	0.5	68	950
amplicon 6	amp_5 LR forward	amp_1 LR reverse	6	68	8374
amplicon 7	amp_5 LR forward	amp_4 LR reverse	9	66	12356
amplicon 8	amp_8 LR forward	amp_1 LR reverse	5	64	7543

Table 12 (continued)

	Forward Long Range Primer	Reverse Long Range Primer	Ext. Time (min)	Annealing Temp.	LR Amplicon Length
amplicon 9	amp_8 LR forward	amp_4 LR reverse	9	68	11525
amplicon 10	amp_10 LR forward	amp_4 LR reverse	3	66	4139
amplicon 11	amp_1 LR forward	amp_11 LR reverse	9	66	13296
amplicon 12*	amp_8 LR forward	amp_11 LR reverse	9	66	12359
amplicon 13	amp_10 LR forward	amp_11 LR reverse	3.5	68	4973
amplicon 14	amp_14 LR forward	amp_14 LR reverse	3.5	62	5209
amplicon 15B	amp_10 LR forward	amp_14B LR reverse	7	68	10084
amplicon 17B*	amp_17B LR forward	amp_17 LR reverse	3	64	3317
amplicon 18	amp_14 LR forward	amp_17 LR reverse	6	68	8302
amplicon 19	amp_19 LR forward	amp_19 LR reverse	3	68	4429
amplicon 20*	amp_17 LR forward	amp_19 LR reverse	5	66	7561
amplicon 21	amp_14 LR forward	amp_19 LR reverse	9	60	12570
amplicon 22	amp_22 LR forward	amp_22 LR reverse	3	64	4526
amplicon 23B	amp_17B LR forward	amp_22 LR reverse	9	68	11989
amplicon 24	amp_19 LR forward	amp_22 LR reverse	6	68	8833
amplicon 25	amp_5 LR forward	amp_11 LR reverse	9	68	13190
amplicon 26	amp_10 LR forward	amp_17 LR reverse	9	62	13152
amplicon 27	amp_27 LR forward	amp_11 LR reverse	0.5	68	995
amplicon 28	amp_27 LR forward	amp_14 LR reverse	4	64	6081
amplicon 29	amp_27 LR forward	amp_17 LR reverse	7	68	9174
amplicon 30	amp_27 LR forward	amp_19 LR reverse	10	68	13442
amplicon 31B	amp_17B LR forward	amp_31 LR reverse	7	68	9331
amplicon 32	amp_19 LR forward	amp_31 LR reverse	4	68	6155
amplicon 33	amp_22 LR forward	amp_31 LR reverse	2	66	1848
amplicon 34	amp_19 LR forward	amp_34 LR reverse	10	68	13458
amplicon 35	amp_22 LR forward	amp_34 LR reverse	7	68	9151
amplicon 36	amp_36 LR forward	amp_34 LR reverse	5	68	7401
amplicon 37	amp_37 LR forward	amp_34 LR reverse	3.5	68	4775
amplicon 38	amp_14 LR forward	amp_31 LR reverse	10	68	14296
amplicon 39	amp_36 LR forward	amp_22 LR reverse	2	66	2776
amplicon 41	amp_22 LR forward	amp_40 LR reverse	7	68	10034
amplicon 42	amp_36 LR forward	amp_40 LR reverse	6	68	8284
amplicon 43	amp_37 LR forward	amp_40 LR reverse	4	66	5658
amplicon 44	amp_44 LR forward	amp_40 LR reverse	0.5	68	1041

* indicates use of GC phusion buffer rather than HF phusion buffer

Table 13: Haplotyping inverse PCR amplicon properties.

	Circle Forward Primer	Circle Reverse Primer	Annealing Temp.	Amplicon Length
amplicon 1	amp_1 circle forward	amp_1 circle reverse	66	202
amplicon 2	amp_1 circle forward	amp_2 circle reverse	60	215
amplicon 3	amp_1 circle forward	amp_3 circle reverse	60	162
amplicon 4	amp_1 circle forward	amp_4 circle reverse	64	197
amplicon 5	amp_5 circle forward	amp_3 circle reverse	56	181
amplicon 6	amp_5 circle forward	amp_1 circle reverse	62	221
amplicon 7	amp_5 circle forward	amp_4 circle reverse	60	216
amplicon 8	amp_8 circle forward	amp_1 circle reverse	66	220
amplicon 9	amp_8 circle forward	amp_4 circle reverse	66	215
amplicon 10	amp_10B circle forward	amp_4 circle reverse	66	217
amplicon 11	amp_1 circle forward	amp_11 circle reverse	66	161
amplicon 12	amp_8 circle forward	amp_11 circle reverse	62	179
amplicon 13	amp_10 circle forward	amp_11 circle reverse	66	181
amplicon 14	amp_14 circle forward	amp_14 circle reverse	66	264
amplicon 15B	amp_10B circle forward	amp_14 circle reverse	66	289
amplicon 17B	amp_17B circle forward	amp_17 circle reverse	66	231
amplicon 18	amp_14 circle forward	amp_17 circle reverse	64	221
amplicon 19	amp_19B circle forward	amp_19 circle reverse	66	190
amplicon 20	amp_17B circle forward	amp_19 circle reverse	66	177
amplicon 21	amp_14 circle forward	amp_19 circle reverse	66	191
amplicon 22	amp_22 circle forward	amp_22 circle reverse	66	253
amplicon 23B	amp_17B circle forward	amp_22 circle reverse	62	234
amplicon 24	amp_19B circle forward	amp_22 circle reverse	62	223
amplicon 25	amp_5 circle forward	amp_11 circle reverse	62	180
amplicon 26	amp_10B circle forward	amp_17 circle reverse	66	221
amplicon 27	amp_27 circle forward	amp_11 circle reverse	62	229
amplicon 28	amp_27 circle forward	amp_14 circle reverse	60	312
amplicon 29	amp_27 circle forward	amp_17 circle reverse	62	269
amplicon 30	amp_27 circle forward	amp_19 circle reverse	66	239
amplicon 31B	amp_17B circle forward	amp_31 circle reverse	66	182
amplicon 32	amp_19B circle forward	amp_31 circle reverse	66	171
amplicon 33	amp_22 circle forward	amp_31 circle reverse	64	201
amplicon 34	amp_19B circle forward	amp_34 circle reverse	66	211
amplicon 35	amp_22 circle forward	amp_34 circle reverse	66	241
amplicon 36	amp_36 circle forward	amp_34 circle reverse	66	194
amplicon 37	amp_37B circle forward	amp_34 circle reverse	62	218
amplicon 38	amp_14 circle forward	amp_31 circle reverse	66	172
amplicon 39	amp_36 circle forward	amp_22 circle reverse	62	206

Table 13 (continued)

	Circle Forward Primer	Circle Reverse Primer	Annealing Temp.	Amplicon Length
amplicon 41	amp_22 circle forward	amp_40 circle reverse	66	246
amplicon 42	amp_36 circle forward	amp_40 circle reverse	66	199
amplicon 43	amp_37B circle forward	amp_40 circle reverse	62	223
amplicon 44	amp_44 circle forward	amp_40 circle reverse	64	252

Table 14: Plasma inverse PCR primer sequences.

Primer Name	Sequence
deltaF508 primer	5' ATGCCTGGCACCATTAAAGAAAATA
deltaF508 reverse circle 2	5' CCACTGTGCTTAATTTTACCC
rs1896887 forward	5' GACAGTATACTCCAAATAGTG
rs1896887 reverse circle	5' CCCTGCCCTAACATTTACAGC
rs718829 forward	5' CAATCCCTGTCCTCAAGGAGC
rs718829 reverse circle	5' TTCCTAGAGTTTGGCCCAGTCCC
rs2082056 forward	5' GAGACATAAAGATATTAAACAGAG
rs2082056 reverse circle	5' CTGTATGTGTGCAGATTTG
rs213934 forward	5' CTCTACTCTTCCCTTCCTCTGTTC
rs213934 reverse circle	5' GAGAACGTAAGAGAGCCAAGAG
rs2518873 forward	5' TGCAAATACAGTACATCTAAC
rs2518873 reverse circle	5' CCATTAGATAATACTGAGGAAAGC
rs213955 forward	5' CAGTATCCTGATATTGACTTATCG
rs213955 reverse circle	5' ACCTGCCTTCAAGCACTTCT
rs6949974 forward	5' CCCCTTCAGTATTCTTGTCCTT
rs6949974 reverse circle	5' GATAAGGAACTACAGTCTATAC
rs17547853 forward	5' GAAGTCTCAGACCAAATCCC
rs17547853 reverse circle	5' GCGAAGAGTTAGATAGGTAGT
rs213965 forward	5' CCTATGTACTTGAGATATAAGTAAGG
rs213965 reverse circle	5' CTCTGTTTATACATGTAATTGTTGG
rs213971 forward	5' GCAGATAAGGAAACTAAAGTTCAG
rs213971 reverse circle	5' ATACCTTCCAGGGTAGTTGTG
rs1042077 forward	5' ACATGGAACACATACCTTCGA
rs1042077 reverse circle	5' CACTGCTGGTATGCTCTCCA

Table 15: Sanger sequencing amplicon properties.

Targeted Loci	Forward Primer	Reverse Primer	Annealing Temp.
rs213950	CATAGCAGAGTACCTGAAACAGGA	AATGGTGCCAGGCATAATCCA	64
rs1896887	TGTGTGATGGTGGGTTCAGT	ATCTAATCCACGGTTTGCCC	66
rs718829	TGTTTGGCACCAGGGATGT	TGTTCTCCCTGTTGTTCACTC	66
rs2082056	TCCGAACCTACCCCAAACAC	CCTTCACTGATAGGAACTGACCA	66
rs213934	AACACTCTCAGATACACGTTCACA	ATTGGGTGGTTAGTGTCATTGTT	66
rs2518873	CTTTAGCAATGCTTTCCTCAGT	TTGCCCCACTTTCCAAAGATT	66
rs213955	TGTGAGCAAGTGAATTCAAGCA	AAATCCCCACACTTGTTGCTAAT	66
rs6949974	TCTGTTTCTTACTGTCCCCTTCAG	GGAAAAGAGACAGGGAAGGGA	66
rs17547853	AAGGCGAGAGTTTACTACCTATC	GTCCTTGACTCCTAGGCATTGT	66
rs213965	GGGATCCATTATGTAGCTCTTGC	AAAGATGAAGACACAGTTCCCA	66
rs213971	TTAGTCCCCACAACTACCCT	ACCCAACCTCCACCATTAACA	64
rs1042077	ATGGAGAGCATACCAGCAGTG	ACAATAGAACATTCTTACCTCTGCC	66

Table 16: Haplotyping method sequencing statistics.

	Number Reads	Reported Reads R1	Reported Reads %	Reported Reads R2	Reported Reads %	Proper Pairs	% Reads Used
G1342	995,469	690,332	69.35%	630,565	63.34%	480,910	48.31%
G1343	991,811	619,498	62.46%	528,968	53.33%	358,383	36.13%
12-032	805,540	483,589	60.03%	436,288	54.16%	268,703	33.36%
G1687	685,653	420,799	61.37%	376,351	54.89%	248,537	36.25%
G1688	718,333	466,721	64.97%	412,056	57.36%	295,712	41.17%
13-014	894,701	591,103	66.07%	524,764	58.65%	380,109	42.48%
G1700	765,353	478,595	62.53%	408,538	53.38%	284,100	37.12%
G1701	915,672	619,375	67.64%	539,220	58.89%	407,935	44.55%
13-019	835,064	500,247	59.91%	455,387	54.53%	297,311	35.60%
G1747	833,965	550,618	66.02%	454,417	54.49%	315,173	37.79%
G1748	673,957	427,823	63.48%	391,441	58.08%	273,341	40.56%
13-049	618,913	416,681	67.32%	375,115	60.61%	267,481	43.22%
G1831	1,122,707	752,920	67.06%	699,677	62.32%	495,604	44.14%
G1832	995,539	624,969	62.78%	530,235	53.26%	345,194	34.67%
13-089	627,893	410,099	65.31%	369,650	58.87%	251,294	40.02%
G1841	874,031	579,404	66.29%	529,983	60.64%	361,329	41.34%
G1842	825,845	570,005	69.02%	522,076	63.22%	379,403	45.94%
13-093	934,667	665,186	71.17%	610,475	65.31%	457,193	48.92%

Table 17: Allele counts per haplotyping sample.

	G1342					G1343				
Loci	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	379	200,735	201,114	0.19%	99.81%	134,588	365	134,953	99.73%	0.27%
rs113993960 CTT/-	113,579	6	113,585	99.99%	0.01%	47,547	3	47,550	99.99%	0.01%
rs1896887 G/C	26,635	48,535	75,170	35.43%	64.57%	25,324	20	25,344	99.92%	0.08%
rs718829 C/A	268	157,260	157,528	0.17%	99.83%	141,474	329	141,803	99.77%	0.23%
rs2082056 T/C	0	83	83	0.00%	100.00%	85	0	85	100.00%	0.00%
rs213934 T/C	257	150,544	150,801	0.17%	99.83%	107,689	202	107,891	99.81%	0.19%
rs2518873 C/T	23	19,734	19,757	0.12%	99.88%	31,548	25	31,573	99.92%	0.08%
rs213955 A/G	47	13,047	13,094	0.36%	99.64%	3,225	11	3,236	99.66%	0.34%
rs6949974 C/T	24,955	22,059	47,014	53.08%	46.92%	43,559	49	43,608	99.89%	0.11%
rs17547853 G/A	76,538	71	76,609	99.91%	0.09%	79,037	60	79,097	99.92%	0.08%
rs213965 T/A	0	43	43	0.00%	100.00%	27	0	27	100.00%	0.00%
rs213971 G/A	2	1,309	1,311	0.15%	99.85%	1,271	0	1,271	100.00%	0.00%
rs1042077 T/G	54,662	46,391	101,053	54.09%	45.91%	97,484	118	97,602	99.88%	0.12%
	12_032					**G1688**				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	47,384	45,309	92,693	51.12%	48.88%	60,383	50,690	111,073	54.36%	45.64%
rs113993960 CTT/-	33,447	7	33,454	99.98%	0.02%	17,851	10,971	28,822	61.94%	38.06%
rs1896887 G/C	35,520	442	35,962	98.77%	1.23%	19,584	111	19,695	99.44%	0.56%
rs718829 C/A	46,283	41,831	88,114	52.53%	47.47%	71,085	63,412	134,497	52.85%	47.15%
rs2082056 T/C	55	71	126	43.65%	56.35%	34	36	70	48.57%	51.43%
rs213934 T/C	49,033	42,177	91,210	53.76%	46.24%	45,275	42,057	87,332	51.84%	48.16%
rs2518873 C/T	14,743	9,378	24,121	61.12%	38.88%	18,756	2,529	21,285	88.12%	11.88%
rs213955 A/G	7,269	6,966	14,235	51.06%	48.94%	8,702	7,998	16,700	52.11%	47.89%
rs6949974 C/T	14,153	23,122	37,275	37.97%	62.03%	30,114	52	30,166	99.83%	0.17%
rs17547853 G/A	53,001	40	53,041	99.92%	0.08%	74,742	64	74,806	99.91%	0.09%
rs213965 T/A	5	6	11	45.45%	54.55%	7	8	15	46.67%	53.33%
rs213971 G/A	635	539	1,174	54.09%	45.91%	342	282	624	54.81%	45.19%
rs1042077 T/G	24,615	38,441	63,056	39.04%	60.96%	63,100	73	63,173	99.88%	0.12%

Table 17 (continued)	G1687					13-014				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	70,731	57,977	128,708	54.95%	45.05%	68,779	63,855	132,634	51.86%	48.14%
rs113993960 CTT/-	13,893	10,799	24,692	56.27%	43.73%	22,560	17,404	39,964	56.45%	43.55%
rs1896887 G/C	17,718	179	17,897	99.00%	1.00%	39,394	465	39,859	98.83%	1.17%
rs718829 C/A	62,553	52,897	115,450	54.18%	45.82%	67,721	51,809	119,530	56.66%	43.34%
rs2082056 T/C	15	36	51	29.41%	70.59%	30	41	71	42.25%	57.75%
rs213934 T/C	19,743	18,721	38,464	51.33%	48.67%	65,334	79,618	144,952	45.07%	54.93%
rs2518873 C/T	10,290	1,775	12,065	85.29%	14.71%	14,841	2,327	17,168	86.45%	13.55%
rs213955 A/G	2,297	2,315	4,612	49.80%	50.20%	15,251	11,944	27,195	56.08%	43.92%
rs6949974 C/T	45,017	52	45,069	99.88%	0.12%	47,495	77	47,572	99.84%	0.16%
rs17547853 G/A	42,108	32	42,140	99.92%	0.08%	92,027	66	92,093	99.93%	0.07%
rs213965 T/A	12	12	24	50.00%	50.00%	8	4	12	66.67%	33.33%
rs213971 G/A	420	343	763	55.05%	44.95%	354	443	797	44.42%	55.58%
rs1042077 T/G	65,069	73	65,142	99.89%	0.11%	94,229	163	94,392	99.83%	0.17%
	G1700					G1701				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	54,679	40,742	95,421	57.30%	42.70%	69,076	52,947	122,023	56.61%	43.39%
rs113993960 CTT/-	26,084	15,256	41,340	63.10%	36.90%	67,650	21,599	89,249	75.80%	24.20%
rs1896887 G/C	23,870	115	23,985	99.52%	0.48%	85,267	228	85,495	99.73%	0.27%
rs718829 C/A	51,352	40,814	92,166	55.72%	44.28%	68,371	53,726	122,097	56.00%	44.00%
rs2082056 T/C	26	35	61	42.62%	57.38%	25	40	65	38.46%	61.54%
rs213934 T/C	47,059	45,645	92,704	50.76%	49.24%	56,618	60,078	116,696	48.52%	51.48%
rs2518873 C/T	21,998	2,875	24,873	88.44%	11.56%	10,702	1,369	12,071	88.66%	11.34%
rs213955 A/G	1,389	1,290	2,679	51.85%	48.15%	20,156	20,420	40,576	49.67%	50.33%
rs6949974 C/T	37,097	41	37,138	99.89%	0.11%	37,339	40	37,379	99.89%	0.11%
rs17547853 G/A	73,118	45	73,163	99.94%	0.06%	95,530	81	95,611	99.92%	0.08%
rs213965 T/A	26	10	36	72.22%	27.78%	8	7	15	53.33%	46.67%
rs213971 G/A	410	306	716	57.26%	42.74%	393	314	707	55.59%	44.41%
rs1042077 T/G	79,375	74	79,449	99.91%	0.09%	89,918	108	90,026	99.88%	0.12%

Table 17 (continued)	13-019					G1747				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	100,412	202	100,614	99.80%	0.20%	62,103	55,684	117,787	52.72%	47.28%
rs113993960 CTT/-	32,781	3	32,784	99.99%	0.01%	50,549	7	50,556	99.99%	0.01%
rs1896887 G/C	61,649	2	61,651	100.00%	0.00%	15,787	2	15,789	99.99%	0.01%
rs718829 C/A	104,823	147	104,970	99.86%	0.14%	55,179	47,443	102,622	53.77%	46.23%
rs2082056 T/C	983	0	983	100.00%	0.00%	23	32	55	41.82%	58.18%
rs213934 T/C	126,227	140	126,367	99.89%	0.11%	63,566	62,179	125,745	50.55%	49.45%
rs2518873 C/T	26,991	10	27,001	99.96%	0.04%	8,764	3,650	12,414	70.60%	29.40%
rs213955 A/G	40,481	16	40,497	99.96%	0.04%	57,197	31	57,228	99.95%	0.05%
rs6949974 C/T	30,086	25	30,111	99.92%	0.08%	17,555	15,750	33,305	52.71%	47.29%
rs17547853 G/A	19,598	19	19,617	99.90%	0.10%	52,080	60	52,140	99.88%	0.12%
rs213965 T/A	2,142	0	2,142	100.00%	0.00%	55	46	101	54.46%	45.54%
rs213971 G/A	4,739	3	4,742	99.94%	0.06%	397	381	778	51.03%	48.97%
rs1042077 T/G	28,641	34	28,675	99.88%	0.12%	34,873	24,195	59,068	59.04%	40.96%
	G1748					13-049				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	55,610	44,107	99,717	55.77%	44.23%	46,830	41,769	88,599	52.86%	47.14%
rs113993960 CTT/-	67,647	30	67,677	99.96%	0.04%	52,417	11	52,428	99.98%	0.02%
rs1896887 G/C	35,500	2	35,502	99.99%	0.01%	43,336	2	43,338	100.00%	0.00%
rs718829 C/A	46,144	36,052	82,196	56.14%	43.86%	52,050	43,770	95,820	54.32%	45.68%
rs2082056 T/C	45	126	171	26.32%	73.68%	69	115	184	37.50%	62.50%
rs213934 T/C	52,750	58,508	111,258	47.41%	52.59%	57,002	59,775	116,777	48.81%	51.19%
rs2518873 C/T	6,970	3,419	10,389	67.09%	32.91%	9,993	5,057	15,050	66.40%	33.60%
rs213955 A/G	43,034	16	43,050	99.96%	0.04%	35,703	32	35,735	99.91%	0.09%
rs6949974 C/T	16,394	15,895	32,289	50.77%	49.23%	17,067	10,136	27,203	62.74%	37.26%
rs17547853 G/A	13,240	15,235	28,475	46.50%	53.50%	16,357	9,521	25,878	63.21%	36.79%
rs213965 T/A	79	62	141	56.03%	43.97%	64	53	117	54.70%	45.30%
rs213971 G/A	487	307	794	61.34%	38.66%	254	189	443	57.34%	42.66%
rs1042077 T/G	15,403	16,420	31,823	48.40%	51.60%	18,612	12,456	31,068	59.91%	40.09%

Table 17 (continued)	G1831					G1832				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	77,885	59,419	137,304	56.72%	43.28%	77,321	41,484	118,805	65.08%	34.92%
rs113993960 CTT/-	73,826	9	73,835	99.99%	0.01%	80,862	5	80,867	99.99%	0.01%
rs1896887 G/C	104,006	7	104,013	99.99%	0.01%	68,114	405	68,519	99.41%	0.59%
rs718829 C/A	73,411	58,614	132,025	55.60%	44.40%	62,374	53,300	115,674	53.92%	46.08%
rs2082056 T/C	296	388	684	43.27%	56.73%	191	234	425	44.94%	55.06%
rs213934 T/C	102,205	104,800	207,005	49.37%	50.63%	58,408	63,486	121,894	47.92%	52.08%
rs2518873 C/T	8,131	4,054	12,185	66.73%	33.27%	15,929	7,175	23,104	68.94%	31.06%
rs213955 A/G	43,328	19	43,347	99.96%	0.04%	3,241	2,977	6,218	52.12%	47.88%
rs6949974 C/T	20,699	17,302	38,001	54.47%	45.53%	23,415	14,695	38,110	61.44%	38.56%
rs17547853 G/A	106,653	79	106,732	99.93%	0.07%	49,383	51	49,434	99.90%	0.10%
rs213965 T/A	183	145	328	55.79%	44.21%	116	94	210	55.24%	44.76%
rs213971 G/A	511	385	896	57.03%	42.97%	576	433	1,009	57.09%	42.91%
rs1042077 T/G	77,307	52,679	129,986	59.47%	40.53%	39,709	20,602	60,311	65.84%	34.16%
	13-089					G1841				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	48,179	44,191	92,370	52.16%	47.84%	367	120,653	121,020	0.30%	99.70%
rs113993960 CTT/-	45,739	5	45,744	99.99%	0.01%	86,268	12	86,280	99.99%	0.01%
rs1896887 G/C	51,021	5	51,026	99.99%	0.01%	70,920	11	70,931	99.98%	0.02%
rs718829 C/A	32,088	27,455	59,543	53.89%	46.11%	230	97,350	97,580	0.24%	99.76%
rs2082056 T/C	25	52	77	32.47%	67.53%	0	119	119	0.00%	100.00%
rs213934 T/C	54,013	50,129	104,142	51.86%	48.14%	133	90,653	90,786	0.15%	99.85%
rs2518873 C/T	15,013	8,161	23,174	64.78%	35.22%	32	24,684	24,716	0.13%	99.87%
rs213955 A/G	35,541	17	35,558	99.95%	0.05%	54,919	42	54,961	99.92%	0.08%
rs6949974 C/T	13,063	7,098	20,161	64.79%	35.21%	95	24,111	24,206	0.39%	99.61%
rs17547853 G/A	30,134	36	30,170	99.88%	0.12%	196	86,987	87,183	0.22%	99.78%
rs213965 T/A	90	66	156	57.69%	42.31%	0	197	197	0.00%	100.00%
rs213971 G/A	414	266	680	60.88%	39.12%	1	780	781	0.13%	99.87%
rs1042077 T/G	23,971	14,281	38,252	62.67%	37.33%	194	60,832	61,026	0.32%	99.68%

Table 17 (continued)	G1842					13-093				
	ref	alt	total	%ref	%alt	ref	alt	total	%ref	%alt
rs213950 G/A	320	133,867	134,187	0.24%	99.76%	292	169,555	169,847	0.17%	99.83%
rs113993960 CTT/-	70,906	14	70,920	99.98%	0.02%	111,115	5	111,120	100.00%	0.00%
rs1896887 G/C	80,136	1,027	81,163	98.73%	1.27%	139,370	9	139,379	99.99%	0.01%
rs718829 C/A	149	81,151	81,300	0.18%	99.82%	117	84,473	84,590	0.14%	99.86%
rs2082056 T/C	0	130	130	0.00%	100.00%	0	94	94	0.00%	100.00%
rs213934 T/C	136	121,794	121,930	0.11%	99.89%	107	125,400	125,507	0.09%	99.91%
rs2518873 C/T	21	18,158	18,179	0.12%	99.88%	18	18,300	18,318	0.10%	99.90%
rs213955 A/G	24,646	43,412	68,058	36.21%	63.79%	59,223	34	59,257	99.94%	0.06%
rs6949974 C/T	59	35,834	35,893	0.16%	99.84%	50	43,878	43,928	0.11%	99.89%
rs17547853 G/A	58,515	36,934	95,449	61.30%	38.70%	136	101,949	102,085	0.13%	99.87%
rs213965 T/A	0	164	164	0.00%	100.00%	0	110	110	0.00%	100.00%
rs213971 G/A	2	624	626	0.32%	99.68%	1	609	610	0.16%	99.84%
rs1042077 T/G	129	48,491	48,620	0.27%	99.73%	90	57,378	57,468	0.16%	99.84%

Table 18: Haplotype category counts at heterozygous loci for trio 1.

SNP Pair	Category	G1342		G1343		12-032	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213950-rs718829	ref-ref	sample homozygous		sample homozygous		13039	24.44%
	ref-alt					11899	22.30%
	alt-ref					11705	21.94%
	alt-alt					16717	31.33%
rs213950-rs213934	ref-ref	sample homozygous		sample homozygous		7342	29.29%
	ref-alt					6880	27.45%
	alt-ref					4776	19.06%
	alt-alt					6066	24.20%
rs718829_rs213934	ref-ref	sample homozygous		sample homozygous		4001	26.57%
	ref-alt					4507	29.93%
	alt-ref					3057	20.30%
	alt-alt					3495	23.21%
rs718829_rs2518873	ref-ref	sample homozygous		sample homozygous		909	42.02%
	ref-alt					332	15.35%
	alt-ref					568	26.26%
	alt-alt					354	16.37%
rs718829_rs213955	ref-ref	sample homozygous		sample homozygous		383	29.67%
	ref-alt					324	25.10%
	alt-ref					266	20.60%
	alt-alt					318	24.63%
rs213934_rs2518873	ref-ref	sample homozygous		sample homozygous		5546	29.63%
	ref-alt					2621	14.00%
	alt-ref					5888	31.45%
	alt-alt					4664	24.92%
rs213934_rs213955	ref-ref	sample homozygous		sample homozygous		659	25.14%
	ref-alt					582	22.21%
	alt-ref					732	27.93%
	alt-alt					648	24.72%
rs2518873_rs213955	ref-ref	sample homozygous		sample homozygous		105	40.08%
	ref-alt					86	32.82%
	alt-ref					35	13.36%
	alt-alt					36	13.74%
rs213955_rs213971	ref-ref	sample homozygous		sample homozygous		32	37.21%
	ref-alt					17	19.77%
	alt-ref					16	18.60%
	alt-alt					21	24.42%

Table 18 (continued)

SNP Pair	Category	G1342		G1343		12-032	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs6949974_rs213971	ref-ref	sample homozygous		sample homozygous		12	23.08%
	ref-alt					7	13.46%
	alt-ref					13	25.00%
	alt-alt					20	38.46%
rs6949974_rs1042077	ref-ref	13901	33.11%	sample homozygous		5075	18.81%
	ref-alt	8351	19.89%			3581	13.27%
	alt-ref	8645	20.59%			3758	13.93%
	alt-alt	11093	26.42%			14564	53.98%

Table 19: Haplotype category counts at heterozygous loci for trio 2.

SNP Pair	Category	G1687		G1688		13-014	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213950-rs113993960	ref-ref	5338	35.48%	3497	35.16%	1834	33.07%
	ref-alt	1851	12.30%	1285	12.92%	187	3.37%
	alt-ref	1710	11.37%	994	9.99%	230	4.15%
	alt-alt	6145	40.85%	4170	41.93%	3295	59.41%
rs213950-rs718829	ref-ref	32481	32.16%	26319	30.55%	22425	32.65%
	ref-alt	23394	23.16%	19986	23.20%	14523	21.14%
	alt-ref	21837	21.62%	19099	22.17%	14128	20.57%
	alt-alt	23291	23.06%	20738	24.07%	17614	25.64%
rs213950-rs213934	ref-ref	2395	26.01%	2742	25.87%	13147	23.36%
	ref-alt	3010	32.69%	2986	28.17%	15659	27.83%
	alt-ref	1652	17.94%	1843	17.39%	9794	17.41%
	alt-alt	2152	23.37%	3030	28.58%	17669	31.40%
rs113993960-rs718829	ref-ref	42	34.43%	234	40.41%	73	58.40%
	ref-alt	47	38.52%	96	16.58%	36	28.80%
	alt-ref	17	13.93%	83	14.34%	9	7.20%
	alt-alt	16	13.11%	166	28.67%	7	5.60%
rs113993960_rs213934	ref-ref	3154	50.27%	6925	51.76%	5046	24.97%
	ref-alt	627	9.99%	1341	10.02%	1536	7.60%
	alt-ref	505	8.05%	1134	8.48%	1259	6.23%
	alt-alt	1988	31.69%	3980	29.75%	12367	61.20%

Table 19 (continued)

SNP Pair	Category	G1687		G1688		13-014	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs718829_rs213934	ref-ref	19	31.15%	8876	28.48%	10923	40.85%
	ref-alt	10	16.39%	7653	24.55%	4509	16.86%
	alt-ref	7	11.48%	6012	19.29%	3473	12.99%
	alt-alt	25	40.98%	8627	27.68%	7836	29.30%
rs718829_rs213955	ref-ref	26	30.23%	243	16.91%	893	33.36%
	ref-alt	28	32.56%	577	40.15%	647	24.17%
	alt-ref	18	20.93%	473	32.92%	506	18.90%
	alt-alt	14	16.28%	144	10.02%	631	23.57%
rs213934_rs213955	ref-ref	64	26.12%	41	25.31%	2769	31.82%
	ref-alt	52	21.22%	41	25.31%	2002	23.01%
	alt-ref	66	26.94%	48	29.63%	2273	26.12%
	alt-alt	63	25.71%	32	19.75%	1657	19.04%

Table 20: Haplotype category counts at heterozygous loci for trio 3.

SNP Pair	Category	G1700		G1701		13-019	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213950_rs113993960	ref-ref	8210	43.26%	7569	36.91%	sample homozygous	
	ref-alt	2952	15.55%	2663	12.99%		
	alt-ref	2797	14.74%	2471	12.05%		
	alt-alt	5020	26.45%	7802	38.05%		
rs213950_rs718829	ref-ref	13915	28.63%	26733	35.22%	sample homozygous	
	ref-alt	12063	24.82%	17923	23.62%		
	alt-ref	11534	23.73%	16581	21.85%		
	alt-alt	11086	22.81%	14657	19.31%		
rs213950_rs213934	ref-ref	4742	22.45%	3509	19.97%	sample homozygous	
	ref-alt	6738	31.91%	5662	32.23%		
	alt-ref	3682	17.44%	2980	16.96%		
	alt-alt	5956	28.20%	5417	30.83%		
rs113993960_rs718829	ref-ref	47	40.52%	467	39.54%	sample homozygous	
	ref-alt	37	31.90%	315	26.67%		
	alt-ref	15	12.93%	182	15.41%		
	alt-alt	17	14.66%	217	18.37%		

Table 20 (continued)

SNP Pair	Category	G1700		G1701		13-019	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs113993960_rs213934	ref-ref	6337	39.75%	7906	45.54%	sample homozygous	
	ref-alt	2553	16.01%	1257	7.24%		
	alt-ref	2141	13.43%	830	4.78%		
	alt-alt	4912	30.81%	7367	42.44%		
rs718829_rs213934	ref-ref	4819	25.36%	5569	23.54%	sample homozygous	
	ref-alt	5780	30.42%	5967	25.23%		
	alt-ref	3811	20.06%	5825	24.63%		
	alt-alt	4592	24.17%	6293	26.60%		
rs718829_rs213955	ref-ref	11	24.44%	450	26.10%	sample homozygous	
	ref-alt	11	24.44%	495	28.71%		
	alt-ref	16	35.56%	383	22.22%		
	alt-alt	7	15.56%	396	22.97%		
rs213934_rs213955	ref-ref	144	21.79%	7582	29.96%	sample homozygous	
	ref-alt	178	26.93%	6188	24.46%		
	alt-ref	155	23.45%	6290	24.86%		
	alt-alt	184	27.84%	5243	20.72%		
rs213955_rs213965	ref-ref	10	55.56%	0	0.00%	sample homozygous	
	ref-alt	5	27.78%	0	0.00%		
	alt-ref	0	0.00%	0	0.00%		
	alt-alt	3	16.67%	0	0.00%		
rs213955_rs213971	ref-ref	46	47.42%	37	28.24%	sample homozygous	
	ref-alt	18	18.56%	31	23.66%		
	alt-ref	21	21.65%	35	26.72%		
	alt-alt	12	12.37%	28	21.37%		

Table 21: Haplotype category counts at heterozygous loci for trio 5.

SNP Pair	Category	G1831		G1832		13-089	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213950_rs718829	ref-ref	20253	34.86%	10686	26.57%	9283	30.29%
	ref-alt	13547	23.32%	9928	24.69%	7050	23.00%
	alt-ref	12661	21.79%	9494	23.61%	6864	22.39%
	alt-alt	11637	20.03%	10104	25.13%	7454	24.32%
rs213950_rs213934	ref-ref	6060	27.79%	5574	19.78%	6030	29.92%
	ref-alt	7040	32.28%	16088	57.09%	4769	23.67%
	alt-ref	3919	17.97%	3119	11.07%	3345	16.60%
	alt-alt	4790	21.96%	3400	12.06%	6008	29.81%

Table 21 (continued)

SNP Pair	Category	G1831		G1832		13-089	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs718829_rs213934	ref-ref	12063	26.45%	12221	29.61%	3690	26.63%
	ref-alt	13250	29.05%	12069	29.24%	3958	28.57%
	alt-ref	9466	20.76%	8416	20.39%	2946	21.26%
	alt-alt	10825	23.74%	8570	20.76%	3261	23.54%
rs718829_rs2518873	ref-ref	767	35.54%	2801	39.42%	1615	38.16%
	ref-alt	375	17.38%	1178	16.58%	660	15.60%
	alt-ref	678	31.42%	2173	30.58%	1361	32.16%
	alt-alt	338	15.66%	954	13.43%	596	14.08%
rs718829_rs213955	ref-ref	sample homozygous		205	55.71%	sample homozygous	
	ref-alt			25	6.79%		
	alt-ref			14	3.80%		
	alt-alt			124	33.70%		
rs2082056_rs213934	ref-ref	121	22.45%	57	21.27%	6	16.67%
	ref-alt	118	21.89%	67	25.00%	5	13.89%
	alt-ref	143	26.53%	67	25.00%	7	19.44%
	alt-alt	157	29.13%	77	28.73%	18	50.00%
rs213934_rs2518873	ref-ref	2378	31.95%	4889	38.47%	4764	30.84%
	ref-alt	1267	17.02%	1954	15.37%	2505	16.22%
	alt-ref	2415	32.44%	3802	29.92%	4405	28.52%
	alt-alt	1384	18.59%	2064	16.24%	3772	24.42%
rs213934_rs213955	ref-ref	sample homozygous		51	28.81%	sample homozygous	
	ref-alt			42	23.73%		
	alt-ref			52	29.38%		
	alt-alt			32	18.08%		
rs213934_rs6949974	ref-ref	sample homozygous		11	23.91%	sample homozygous	
	ref-alt			10	21.74%		
	alt-ref			12	26.09%		
	alt-alt			13	28.26%		
rs2518873_rs213955	ref-ref	sample homozygous		245	34.90%	sample homozygous	
	ref-alt			248	35.33%		
	alt-ref			85	12.11%		
	alt-alt			124	17.66%		
rs213955_rs213971	ref-ref	sample homozygous		46	26.14%	sample homozygous	
	ref-alt			32	18.18%		
	alt-ref			57	32.39%		
	alt-alt			41	23.30%		

Table 21 (continued)

SNP Pair	Category	G1831		G1832		13-089	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs6949974_rs213971	ref-ref	20	32.26%	48	42.48%	9	22.50%
	ref-alt	8	12.90%	24	21.24%	6	15.00%
	alt-ref	30	48.39%	19	16.81%	13	32.50%
	alt-alt	4	6.45%	22	19.47%	12	30.00%
rs6949974_rs1042077	ref-ref	10929	33.57%	13103	47.00%	7631	46.17%
	ref-alt	6934	21.30%	5021	18.01%	3442	20.82%
	alt-ref	6802	20.89%	5220	18.72%	3379	20.44%
	alt-alt	7895	24.25%	4536	16.27%	2077	12.57%
rs213965_rs1042077	ref-ref	26	25.74%	sample homozygous		9	23.08%
	ref-alt	28	27.72%			12	30.77%
	alt-ref	21	20.79%			10	25.64%
	alt-alt	26	25.74%			8	20.51%
rs213971_rs1042077	ref-ref	143	27.45%	158	32.51%	146	35.27%
	ref-alt	136	26.10%	121	24.90%	113	27.29%
	alt-ref	123	23.61%	117	24.07%	86	20.77%
	alt-alt	119	22.84%	90	18.52%	69	16.67%

Table 22: Haplotype category counts at heterozygous loci for trio 5.

SNP Pair	Category	G1841		G1842		13-093	
		Counts	Percentage of Total Counts	Counts	Percentage of Total Counts	Counts	Percentage of Total Counts
rs213955_rs17547853	ref-ref	sample homozygous		7885	17.77%	sample homozygous	
	ref-alt			4917	11.08%		
	alt-ref			20742	46.74%		
	alt-alt			10832	24.41%		

BIBLIOGRAPHY

1. Hospital, U.C.f.M.G.a.M.-W., *Prenatal Testing Options Information Sheet*, M.-W. Hospital, Editor.
2. Cartier, L., et al., *Counselling considerations for prenatal genetic screening.* J Obstet Gynaecol Can, 2012. **34**(5): p. 489-93.
3. Wilson, K.L., et al., *NSGC practice guideline: prenatal screening and diagnostic testing options for chromosome aneuploidy.* J Genet Couns, 2013. **22**(1): p. 4-15.
4. Bulletins, A.C.o.P., *ACOG Practice Bulletin No. 77: screening for fetal chromosomal abnormalities.* Obstet Gynecol, 2007. **109**(1): p. 217-27.
5. American College of, O. and Gynecologists, *ACOG Practice Bulletin No. 88, December 2007. Invasive prenatal testing for aneuploidy.* Obstet Gynecol, 2007. **110**(6): p. 1459-67.
6. Lippman, A., et al., *Canadian multicentre randomized clinical trial of chorion villus sampling and amniocentesis. Final report.* Prenat Diagn, 1992. **12**(5): p. 385-408.
7. *CDC publishes report on chorionic villus sampling and amniocentesis.* Am Fam Physician, 1995. **52**(4): p. 1210.
8. Jackson, L.G., et al., *A randomized comparison of transcervical and transabdominal chorionic-villus sampling. The U.S. National Institute of Child Health and Human Development Chorionic-Villus Sampling and Amniocentesis Study Group.* N Engl J Med, 1992. **327**(9): p. 594-8.
9. Herzenberg, L.A., et al., *Fetal cells in the blood of pregnant women: detection and enrichment by fluorescence-activated cell sorting.* Proc Natl Acad Sci U S A, 1979. **76**(3): p. 1453-5.
10. Bianchi, D.W., et al., *Isolation of fetal DNA from nucleated erythrocytes in maternal blood.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3279-83.
11. Bianchi, D.W., et al., *PCR quantitation of fetal cells in maternal blood in normal and aneuploid pregnancies.* Am J Hum Genet, 1997. **61**(4): p. 822-9.
12. Cheung, M.C., J.D. Goldberg, and Y.W. Kan, *Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood.* Nat Genet, 1996. **14**(3): p. 264-8.
13. Bianchi, D.W., et al., *Fetal gender and aneuploidy detection using fetal cells in maternal blood: analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study.* Prenat Diagn, 2002. **22**(7): p. 609-15.
14. Malone, F.D., et al., *First-trimester or second-trimester screening, or both, for Down's syndrome.* N Engl J Med, 2005. **353**(19): p. 2001-11.
15. Lo, Y.M., et al., *Presence of fetal DNA in maternal plasma and serum.* Lancet, 1997. **350**(9076): p. 485-7.

16. Lun, F.M., et al., *Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma.* Clin Chem, 2008. **54**(10): p. 1664-72.
17. Chiu, R.W., et al., *Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20458-63.
18. Fan, H.C., et al., *Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood.* Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16266-71.
19. Finning, K.M., et al., *Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RHD genotyping service.* Transfusion, 2002. **42**(8): p. 1079-85.
20. Li, Y., et al., *Non-invasive prenatal detection of achondroplasia in size-fractionated cell-free DNA by MALDI-TOF MS assay.* Prenat Diagn, 2007. **27**(1): p. 11-7.
21. Li, Y., et al., *Cell-free DNA in maternal plasma: is it all a question of size?* Ann N Y Acad Sci, 2006. **1075**: p. 81-7.
22. Li, Y., et al., *Detection of paternally inherited fetal point mutations for beta-thalassemia using size-fractionated cell-free DNA in maternal plasma.* JAMA, 2005. **293**(7): p. 843-9.
23. Gonzalez-Gonzalez, M.C., et al., *Huntington disease-unaffected fetus diagnosed from maternal plasma using QF-PCR.* Prenat Diagn, 2003. **23**(3): p. 232-4.
24. Rodriguez de Alba, M., et al., *Noninvasive prenatal diagnosis of monogenic disorders.* Expert Opin Biol Ther, 2012. **12 Suppl 1**: p. S171-9.
25. Lun, F.M., et al., *Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19920-5.
26. Gu, W., et al., *Noninvasive prenatal diagnosis in a fetus at risk for methylmalonic acidemia.* Genet Med, 2014. **16**(7): p. 564-7.
27. Fan, H.C., et al., *Non-invasive prenatal measurement of the fetal genome.* Nature, 2012. **487**(7407): p. 320-4.
28. Kitzman, J.O., et al., *Noninvasive whole-genome sequencing of a human fetus.* Sci Transl Med, 2012. **4**(137): p. 137ra76.
29. Lam, K.W., et al., *Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to beta-thalassemia.* Clin Chem, 2012. **58**(10): p. 1467-75.
30. Organization, W.H. *Genes and human disease*. March 25, 2016]; Available from: http://www.who.int/genomics/public/geneticdiseases/en/index2.html.
31. Moskowitz, S.M., et al., *CFTR-Related Disorders*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).
32. (U.S.), N.L.o.M. *Genetics Home Reference: Cystic fibrosis*. 2015 [cited 2015 May 27]; Available from: http://ghr.nlm.nih.gov/condition/cystic-fibrosis.
33. Boyle, M.P. and K. De Boeck, *A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect.* Lancet Respir Med, 2013. **1**(2): p. 158-63.
34. Sosnay, P.R., et al., *Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene.* Nat Genet, 2013. **45**(10): p. 1160-7.
35. Rosenstein, B.J. and G.R. Cutting, *The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel.* J Pediatr, 1998. **132**(4): p. 589-95.
36. Hamosh, A., et al., *Comparison of the clinical manifestations of cystic fibrosis in black and white patients.* J Pediatr, 1998. **132**(2): p. 255-9.

37. Kerem, E., *Pharmacological induction of CFTR function in patients with cystic fibrosis: mutation-specific therapy.* Pediatr Pulmonol, 2005. **40**(3): p. 183-96.
38. Watson, M.S., et al., *Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel.* Genet Med, 2004. **6**(5): p. 387-91.
39. Strom, C.M., et al., *Extensive sequencing of the cystic fibrosis transmembrane regulator gene: assay validation and unexpected benefits of developing a comprehensive test.* Genet Med, 2003. **5**(1): p. 9-14.
40. Saleheen, D. and P.M. Frossard, *The cradle of the deltaF508 mutation.* J Ayub Med Coll Abbottabad, 2008. **20**(4): p. 157-60.
41. Morral, N., et al., *The origin of the major cystic fibrosis mutation (delta F508) in European populations.* Nat Genet, 1994. **7**(2): p. 169-75.
42. Hulten, M., *Chiasma distribution at diakinesis in the normal human male.* Hereditas, 1974. **76**(1): p. 55-78.
43. International HapMap, C., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
44. Foundation, C.F. *Newborn Screening for CF*. [cited 2016 April 1]; Available from: https://www.cff.org/What-is-CF/Testing/Newborn-Screening-for-CF.
45. Vernooij-van Langen, A.M., et al., *Novel strategies in newborn screening for cystic fibrosis: a prospective controlled study.* Thorax, 2012. **67**(4): p. 289-95.
46. van Gool, K., et al., *Understanding the costs of care for cystic fibrosis: an analysis by age and health state.* Value Health, 2013. **16**(2): p. 345-55.
47. Bioinformatics, B. *Trim Galore!* ; Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
48. Lab, H. *FASTX-Toolkit*. Available from: http://hannonlab.cshl.edu/fastx_toolkit/.
49. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
50. Dugaiczyk, A., H.W. Boyer, and H.M. Goodman, *Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures.* J Mol Biol, 1975. **96**(1): p. 171-84.
51. Shoaib, M., et al., *Multiple displacement amplification for complex mixtures of DNA fragments.* BMC Genomics, 2008. **9**: p. 415.
52. Hayashi, K., et al., *Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol.* Nucleic Acids Res, 1986. **14**(19): p. 7617-31.
53. Zeevi, D.A., et al., *Proof-of-principle rapid noninvasive prenatal diagnosis of autosomal recessive founder mutations.* J Clin Invest, 2015. **125**(10): p. 3757-65.
54. Colombo, R., *Age estimate of the N370S mutation causing Gaucher disease in Ashkenazi Jews and European populations: A reappraisal of haplotype data.* Am J Hum Genet, 2000. **66**(2): p. 692-7.
55. Yang, W.Y., et al., *Leveraging reads that span multiple single nucleotide polymorphisms for haplotype inference from sequencing data.* Bioinformatics, 2013. **29**(18): p. 2245-52.
56. Korlach, J., et al., *Real-time DNA sequencing from single polymerase molecules.* Methods Enzymol, 2010. **472**: p. 431-55.