UNIVERSITÉ DE MONTRÉAL

Development of a Blood Antigen Molecular Profiling Panel using Genotyping Technologies for Patients Requiring Frequent Transfusions

By

Ian Mongrain

Department of Pharmacology

Faculty of Medicine

Master Thesis presented at the Faculty of Graduate Studies

In preparation for Masters ès Science (M.Sc) level

In Pharmacogenomics

17 Juillet 2009

UNIVERSITÉ DE MONTRÉAL

Graduate Study Faculty

Master Thesis is entitled:

Development of a Blood Antigen Molecular Profiling Panel using Genotyping Technologies for Patients Requiring Frequent Transfusions

By

Ian Mongrain

Pharmacology Department

Faculty of Medicine

Will be evaluated by the following Jury members:

Martin Sirois, PhD, Président Rapporteur

Michael Phillips, PhD, Directeur Scientifique

Simon DeDenus, B.Phar, PhD, Membre du Jury

RÉSUMÉ

Contexte. Les phénotypes ABO et Rh(D) des donneurs de sang ainsi que des patients transfusés sont analysés de façon routinière pour assurer une complète compatibilité. Ces analyses sont accomplies par agglutination suite à une réaction anticorps-antigènes. Cependant, pour des questions de coûts et de temps d'analyses faramineux, les dons de sang ne sont pas testés sur une base routinière pour les antigènes mineurs du sang. Cette lacune peut résulter à une allo-immunisation des patients receveurs contre un ou plusieurs antigènes mineurs et ainsi amener des sévères complications pour de futures transfusions. **Plan d'étude et Méthodes.** Pour ainsi aborder le problème, nous avons produit un panel génétique basé sur la technologie « GenomeLab _SNPstream» de Beckman Coulter, dans l'optique d'analyser simultanément 22 antigènes mineurs du sang. La source d'ADN provient des globules blancs des patients préalablement isolés sur papiers FTA. **Résultats.** Les résultats démontrent que le taux de discordance des génotypes, mesuré par la corrélation des résultats de génotypage venant des deux directions de l'ADN, ainsi que le taux d'échec de génotypage sont très bas (0,1%). Également, la corrélation entre les résultats de phénotypes prédit par génotypage et les phénotypes réels obtenus par sérologie des globules rouges et plaquettes sanguines, varient entre 97% et 100%. Les erreurs expérimentales ou encore de traitement des bases de données ainsi que de rares polymorphismes influençant la conformation des antigènes, pourraient expliquer les différences de résultats. Cependant, compte tenu du fait que les résultats de phénotypages obtenus par génotypes seront toujours co-vérifiés avant toute transfusion sanguine par les technologies standards approuvés par les instances gouvernementales, les taux de corrélation obtenus sont de loin supérieurs aux critères de succès attendus pour le projet. **Conclusion.** Le profilage génétique des antigènes mineurs du sang permettra de créer une banque informatique centralisée des phénotypes des donneurs, permettant ainsi aux banques de sang de rapidement retrouver les profiles compatibles entre les donneurs et les receveurs.

Mots clés : SNPs, groupe sanguin, criblage à haut débit, antigènes mineurs

ABSTRACT

Background. ABO and Rh(D) phenotyping of both blood donors and transfused patients is routinely performed by blood banks to ensure compatibility. These analyses are done by antibody-based agglutination assays. However, blood is not routinely tested for minor blood group antigens on a regular basis because of cost and time constraints. This can result in alloimmunization of the patient against one or more minor antigens and may complicate future transfusions. **Study design and Methods.** To address this problem, we have generated an assay on the GenomeLab SNPstream genotyping system (Beckman Coulter, Fullerton, CA) to simultaneously test polymorphisms linked to 22 different blood antigens using donor's DNA isolated from minute amounts of white blood cells. **Results.** The results showed that both the error rate of the assay, as measured by the strand concordance rate, and the no-call rate were very low (0.1%). The concordance rate with the actual red blood cell and platelet serology data varied from 97 to 100%. Experimental or database errors as well as rare polymorphisms contributing to antigen conformation could explain the observed differences. However, these rates are well above requirements since phenotyping and cross-matching will always be performed prior to transfusion. **Conclusion.** Molecular profiling of blood donors for minor red blood cell and platelet antigens will give blood banks instant access to many different compatible donors through the set-up of a centralized data storage system.

Key words: Genotyping, minor blood antigens, molecular profiling

TABLE OF CONTENT

FIGURE LIST

TABLE LIST

ABBREVIATION LIST

ADR: Adverse drug reaction

AIDS: Acquired Immune Deficiency Syndrome

ASE: Allele specific extension

ASO: Allele specific oligonucleotide

CNV: Copy Number Variation

DNA: deoxyribonucleic acid

FDA: Food and Drug Administration

GBAP: genotyping blood antigen panel

GLP: Good laboratory practice

HDN: Haemolytic disease of the fetus and newborn

HQ: Héma-Québec

MAF: Minor allele frequency

PCR: Polymerase chain reaction

PGX: Pharmacogenomics

RBC: Red blood cell

RFLP: Restriction fragment length polymorphism

SBE: Single base extension

SNP: Single nucleotide polymorphism

SOP: Standard Operating procedure

SSCP: Single-stranded conformation polymorphism

TOF: Time of flight

UCSC: University of California of San-Francisco

UTR: Untranslated regions

CAP : College of American Pathologist

CLIA: Clinical Laboratory Improvement Amendments

AKNOWLEDGEMENT

Je veux remercier tout particulièrement l'équipe de recherche de Héma-Québec, en particulier Maryse St-Louis et Josée Perreault, qui ont travaillé étroitement avec moi pour la mise en place de ce projet de recherche et de son application clinique. Je veux les remercier d'avoir collaboré avec Génome Québec et de nous avoir fait confiance dans le développement de la première banque de sang génotypé clinique au Canada et d'avoir mis notre équipe de recherche en avant plan dans plusieurs présentations qu'ils ont effectuées. Ce fut un plaisir énorme de travailler avec vous. De plus, je ne peux m'empêcher de remercier grandement Alexandre Montpetit qui a initié ce projet à ses tout débuts et d'avoir contribué de façon majeure au développement et à la vulgarisation scientifique de ce projet. Un gros merci à Michal Blazejczyk, Marc Bouffard et Yannick Renaud qui ont contribué de façon généreuse à la mise en place du processus informatisé de ce projet, sans eux l'analyse des résultats aurait été très laborieuse. Également, je suis très reconnaissant à Lizette Mailloux et à Marie-Josée Gaulin qui ont toutes deux effectué un travail de génotypage clinique génial. Bravo! Aussi, je veux remercier Michael Phillips, mon directeur de recherche, de m'avoir fait confiance dans toutes les étapes de ce projet, autant pour le développement que de gestion au jour le jour du projet du début à sa toute fin. Je remercie grandement Andrea Smith, la gestionnaire principale du projet qui a mis de façon grandiose la table entre Génome Québec et Héma-Québec, ce qui a sans aucun doute facilité toutes les collaborations scientifiques et d'affaires présentes et futures. Finalement, je veux remercier Sharon Marsh qui a été une référence sans limites dans la mise en place de ce mémoire, pour ses conseils stratégiques importants et bien sûr, pour la correction détaillée de la langue anglaise.

INTRODUCTION

1. General Introduction

Personalized medicine, Pharmacogenomics, Molecular Profiling, and Targeted Therapeutics are all terms describing the same ideal: identifying individual variations affecting drug therapies and specialized therapies. It's well established in the scientific community as well as in the regulatory instances that modern medicine should target and understand the individual's genomic make-up in order to increase the effectiveness of therapies, as well as decrease the toxicities. Many good examples are now reflected in modern therapies, such as Herceptin© (Trastuzumab), Gleevec© (Imatinib) or Iressa© (Gefitinib). These anti-cancer therapeutic agents are currently prescribed with a molecular profiling assay and they represent the gold standard for the personalised medicine approach. To extend this success to other fields of medicine, many pharmacogenomics studies have to be performed using state-of-the-art genotyping technologies. Many of these technologies are now available to help in the research of SNPs (Single Nucleotide Polymorphisms) that affect or influence a patient's therapeutic outcome. Some of these technologies provide the ability to perform whole genome scanning experiments (e.g. Illumina, Affymetrix) and some are more suitable for focusing on specific genomic regions (e.g. SNPstream, Sequenom). (McLeod et al., 2004).

Since pharmacogenomics is a fairly new field of research, most of the current studies are in the early stages of development. Even if some studies are more mature, most of them are still closely attached to the bench and more effort is required to transfer it into the clinic or at the bed-side. In the last two years, two projects from the Génome Québec and Montreal Heart Institute Pharmacogenomics Centre were successfully transferred from development into the clinical application. These projects entail the development of a molecular profiling panel for minor blood antigens that was one of our first projects that we transferred into a clinical process and the other project that is currently progressing towards a clinical application is the CVD (Cardiovascular

Disease) risk panel. Both projects are using cutting edge technologies available for such applications, and the more important thing is that the projects are developed and optimized for clinical grade usage. This work is already contributing to their use in modern therapy (minor blood antigen panel) as well as to contribute to future pharmacogenomics guidance.

2. Genomic variations: An overview

There is a real interest in documenting the amount of genetic variation in the human species. This information is required by the biomedical community, who require detailed maps of genetic variations (i.e. SNPs: Single Nucleotide Polymorphisms, insertions and deletions (indels), variable number of tandem repeats (VNTRs), Microsatellites, Minisatellites and CNV: Copy Number Variation) in order to identify genes and loci associated with diseases. The ultimate goal is to characterize these variations to identify disease-related polymorphisms among patients and populations. This information is also desired by anthropologists to reconstruct our "Human History" and understand the role of culture and geography in the global distribution of human variation and migration (Weiss, 1998). These genomic variations are also valuable for the investigation of molecular events that underline evolution, genetic drift, mutation, recombination and selection (Nachman, 2001).

2.1. Single Nucleotide Polymorphism

The Human Genome Project has identified that the most common form of variation in the human genome is the single nucleotide polymorphism (SNP). A SNP, by definition, is a stable substitution of a single base with a frequency of 1% in at least one population. Single nucleotide polymorphisms are distributed throughout the human genome at an estimated frequency of one every 1900 bp. Up to now, more than six million SNPs have been mapped and are accessible in public databases. At the chromosome level, SNP densities appear to be constant across the human genome

with the only exception being sex chromosomes (Sachidanandam et al., 2001). SNPs can occur in different places within the genome and this can result in multiple effects on gene function and expression. For example, SNPs can occur in the non-coding 5'UTR, non-coding 3'UTR, intergenic regions, introns and exons. SNPs that occur in exonic regions may be nonsynonymous, coding for an amino acid change (eg. Glu➔Asp) as well as synonymous, coding for no amino change (eg. Glu➔Glu).

2.2. Satellites, Microsatellites, Minisatellites and CNVs

In 1993, Daniel Tautz was the first researcher to distinguish between satellites, microsatellites and minisatellites. He concluded that satellites are composed of repeats of several thousand base pairs with frequencies of 10^3 to 10^7 at each genomic locus and they are located in heterochromatin, mainly in the centromeres. Minisatellites and microsatellites are more broadly dispersed throughout the genome and they have a moderate degree of repetition; microsatellites typically made up of <10 base pair repeats whereas the minisatellites usually contain between 10 to 100 base pairs (Debrauwere et al., 1997). Since the mini and microsatellites are highly polymorphic, these genomic variations have been used to support genomic fine mapping as well as to support the forensic investigations. More recently it has been discovered that the structural diversity of humans is much greater than previously expected. This new finding combined with affordable cutting-edge technologies, has led to the creation of a new research area, copy number variation (CNV) mapping (Goidts et al., 2006). CNVs define regions of copy number polymorphisms in the genome and are responsible for a significant amount of structural variation. One of the most examined CNVs is the segmental duplication, these are low-copy number repeats of DNA blocks ranging from a few kilobases to several hundred kilobases in length (Goidts et al., 2006).

2.3. Genetic variation and human diseases

Generally SNPs have been studied for their impact on defects in biological processes; consequently research has concentrated on SNPs that alter the function or the expression of the genes. From many studies of complex diseases it has been suggested that multiple variants confer the susceptibility, but it's not clear yet whether rare (<1%) or common polymorphisms are determinants that are the most responsible. Based on SNP maps, it has been estimated that there could be between 50,000 and 250,000 functional SNPs (Risch, 2001). Over time, many less common single nucleotide variants have been discovered, these do not occur at a high enough frequency to be considered as a SNPs (at least 1% frequency), however, they nonetheless have been associated with biological defects and other clinical importance (Pritchard, 2001). It is well established that an amino acid substitution caused by a SNP may interfere with the protein function. Likewise, gene expression can also be affected by SNPs positioned in a critical regulatory site. The famous example of the CFTR (Cystic Fibrosis) gene illustrates perfectly the deleterious effect of a variation that changes an amino acid which results in altered gene function, so much so that more than 60 different exonic variants are routinely screened in the clinic as a phenotype screening test (Moskowitz et al., 2008).

It is also important to take note that SNPs located in promoter regions can adversely affect gene expression and infer deleterious phenotypes, such as SNPs located in the promoter region of the genes that coordinate the immune response, namely variations in TNFα, IL4, IL6 and IL10 have been associated with a range of autoimmune and infectious diseases (McGuire et al., 1994). Additionally, it has been demonstrated that the -439T>C SNP located in the promoter region of the Duffy antigen receptor gene encoding the chemokines FY can completely abolish expression when it is present, the same study also demonstrated that this polymorphism is mainly present within African populations and leads to a protective phenotype against the malaria parasite Plasmodium Vivax (Iwamoto et al., 1996).

Historically the intergenic and intronic regions have been less well examined and are infrequently associated with changes in phenotype. Recently however, intergenic and intronic SNPs have been of interest and are no longer associated with "Junk DNA" (Zuckerkandl, et al., 2007). To illustrate this, a recent study involving the CYP2D6 cytochrome enzyme has demonstrated that the single nucleotide polymorphism G>A located in the intron 6 at the position 2988 has a direct impact on the mRNA level and also on protein expression. Further mechanistic experiments have demonstrated that conversion from G to A destabilizes the splicing events and therefore reduces the amount of mature CYP2D6 containing exon 6 (Toscano et al., 2006). This recently described polymorphism is known as *41 and represents one of the main deleterious components of CYP2D6 gene and therefore affects greatly human body detoxification.

Variations in regulatory regions are typically associated with mRNA expression levels, and consequently the resulting protein levels. Polymorphisms found in the promoter regions can also be predictive for human diseases. The UGT1A1 gene illustrates the importance of such polymorphisms in a non-coding region. The variant UGT1A1*28 is well described in literature to affect the downstream regulation of the enzyme consequently leading to the impaired elimination of the bilirubin and subsequently causing Gilbert Syndrome (Roden et al., 2006).

Since 1993, microsatellites and minisatellites have been well characterized and are associated with human diseases that have dominant mode of inheritance. The diseases are classified into two groups, the first represents the size of repeats in the coding or regulatory genomic regions affecting gene expression and leading to aberrant or toxic proteins (Sutherland et al., 1995). The second group is associated with familial heredity defects and cancers. These types of variation are often located in very unstable repeat regions. Taken together more than a dozen well known and characterized diseases have been recorded. Huntingdon disease, muscular dystrophy and fragile X syndrome are some of the most examined pathologies related to minisatellites. Typically, the numbers of repeats are directly connected with the age of the disease onset (Fu et al., 1991). With new research in the CNV area, many

different associations have been made regarding disease susceptibility. A great example of the utility of such mapping is the case of the HIV, where CNVs have been highly associated CCL3L1 gene duplications and the susceptibility to acquire AIDS (Gonzalez et al., 2005).

2.4. The Promises of Pharmacogenomics

Between 1909 and 1923, the physiologist Archibald Garrod, was the first person to propose that common genetic factors might underline the error of metabolism and the variability in drug response (Roden et al., 2006). Today, the outcome of drug therapy is often well understood, but still sometimes unpredictable, varying from beneficial effects to lack of efficacy to very serious adverse drug reactions. Recent statistics have demonstrated that more than 100,000 deaths annually in the US are attributable to adverse drug reactions. Up to 7% of all hospital admission in North America and Europe are related to ADRs (adverse drug reactions) (Manolopoulos, 2007).

The concept that genetic variation contributes to disease phenotypes and drug response is commonly accepted in the scientific community. In 2003, the International Human Genome Sequencing consortium declared the Human project had been completed, giving expectation for clinical applications in the near future. The field of pharmacogenomics promised the end of the "one drug fits all" trial and error drug selection system, and was predicted to be part of the first clinical application of the Human Genome Project (Swen et al., 2007). Understanding the genetic contribution to the variability in drug response provides a new tool in

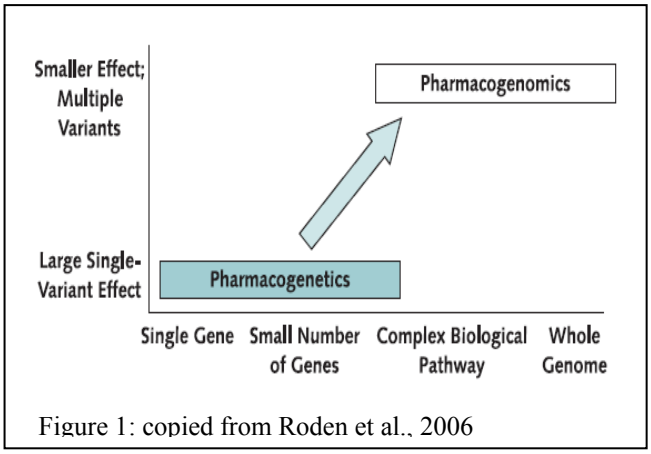


Figure 1: copied from Roden et al., 2006

drug development with the hope of decreasing the risk of toxicity and identifying the patients that will beneficially response to the medication (Roses, 2004). The terms of "Pharmacogenomics" and "Pharmacogenetics" are broadly used in different scientific discussions as well in many literatures and reviews. Both terms are intimately connected since they are referring to the exact same endpoint. Pharmacogenetics focuses on the large clinical effect of single gene variants in a small number of patients, whereas pharmacogenomics examines many genomic loci including large biological pathways such as lipid-lowering agents (Liao, 2002) and hypertension (Dedenus, 2004), as well as the whole genome (Iakoubova et al., 2008) to identify variants that together determine the variability in therapeutic outcomes (See figure 1). Even if pharmacogenomics and pharmacogenetics are quite recent research areas, the literature abounds with studies that involve different genes and different therapeutic fields. In the current clinical practice at least five known drugs are prescribed with pharmacogenetic tests (Sheffield et al. 2009). In early 1990, some studies demonstrated that women with breast cancer having HER2 over expression had very poor prognosis. In the same decade Genetech developed a monoclonal antibody directed drug against the HER2 protein conferring for those women real improvement in clinical outcomes. Pharmacogenomics gained in popularity when Herceptin© (Trastuzumab) was prescribed with a pharmacogenomics pre-screen test. Today, testing for HER2 over expression is a standard practice to detect candidates for Herceptin© (Trastuzumab) treatment among women suffering from breast cancer (Swen et al., 2007). Also, one of the most studied pharmacogenomics tests is for Warfarin. Since this drug is the most widely prescribed oral anticoagulant drug and there is greater than 10-fold interindividual variability in the dose required to attain the therapeutic response, the need of a pharmacogenomics clinical test to determine the optimum starting dose was urgent. The genes CYP2C9 and VKORC1 have been identified as the two major contributors for the interindividual response to warfarin. Today, many retrospective and prospective studied tend to demonstrate that the genomics markers *2 and *3 of CYP2C9 gene and -1639 G>A of VKORC1 gene, contribute at least at 40% of the interindividual variability in dose requirement (Van Schie et al., 2009). Other current examples of pharmacogenomics being applied in

the clinic include irinotecan, a colorectal anticancer agent, Abacavir, an antiviral and Tamoxifen a breast cancer treatment, are routinely prescreened for UGT1A1, HLA-B and CYP2D6 respectively (Roden et al., 2006 and Swen et al., 2007).

2.4.1. Gene selection: Direct and Indirect approaches

With these numerous successes, many other therapeutic areas are on the way to identifying key genomics-based elements correlated with either drug efficiency or toxicity. Among these therapeutic areas, the antihypertensive, lipid-lowering, anticancer and neurologic disorder drugs are the most advanced in the identification of potential pathways that underline genes and markers that could infer specific risk of drug toxicity or poor drug response. The current effort put in these different studies is identifying more and more complex pathways and polygenic problems. For instance, lipid lowering therapies implicates the very complex Mevalonate pathways and indirectly related pathways such as the isopranoids and glycosylation pathways as well as the intracellular signaling pathways (Hanai et al., 2007). This has led to the identification of more than 500 related and less related genes and more than 50,000 potential clinically relevant markers. To investigate this, two genotyping strategies can be employed; the candidate gene approach (direct approach) and the whole genome scan approach (indirect approach). The two approaches have specific advantages and disadvantages. The candidate gene approach refers to the selection of genes that are part of known pathological pathways as well as part of a genetic association hypothesis. This strategy can be extremely powerful for studying the genetic make-up

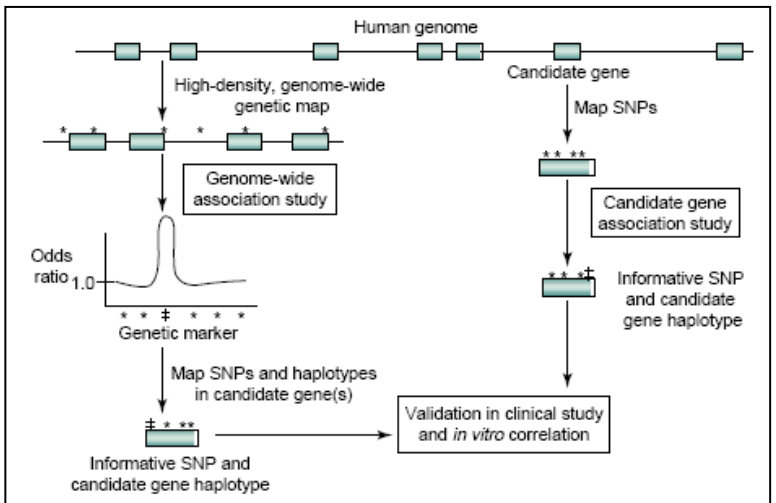


Figure 2: Copied from: Taylor et al., 2001

of complex traits and pin-point the rare variants. Nevertheless, the candidate approach is greatly limited as it relies on existing knowledge about the presumed biological pathways and biological concept, and unfortunately most of the biological attributes remain unknown (Zhu et al., 2007). In contrast, the whole genome scan approach proceeds without any presupposition regarding biological knowledge, which leads to an unbiased approach. The main advantage of this approach is a greater coverage of all pathways including unknown loci that possibly underline the drug toxicity or inefficiency. The biggest disadvantages of this technique are the high cost related to such experiments, which needs expensive technologies and reagents and the relative high number of individual needed to reach statistical endpoint value. Also, since the whole-genome scan approach relies mainly on Hapmap data, it can only identify pre-described loci, so a candidate approach will be necessary to further map and isolate the causative markers (The International HapMap consortium, 2003). Figure 2 illustrates the concept of candidate genes and whole genome scan approaches and the downstream processes involved in defining the clinical application (Taylor et al., 2001). Over the past few years the whole genome approach has held great interest for the pharmacogenomics studies. Since the clinical drug response is not only reflecting the intrinsic properties of target cells or pathways, but also metabolic properties, drug-drug interactions, pharmacokinetic, pharmacodynamic as well as the population and gender. Consequently the whole-genome scan approach enables the highest degree of possible associations to be identified. Also, with the completion of the Hapmap project in 2003, the availability of Hapmap data for different populations have permitted a better evaluation of population and gender effects on drug toxicity (Zhang et al., 2008).

2.4.2. Selection of markers: Candidate genes and HapMap resources

Traditionally when a pharmacogenomics study design uses a candidate gene approach, most candidate markers are selected from both literature and public database (eg dbSNP and OMIM) search. For the candidate gene approach, the

markers are selected based on their known association with specific diseases as well as their pharmacogenetic clinical importance. Most of the time the list of markers is complemented with markers that have unknown function but could potentially modify the target protein (e.g. variants located in the exons, promoter and intronic splicing sites). The HapMap data resource has greatly improved SNP selection. Briefly the HapMap resources have elucidated the common patterns of DNA sequence variation in the human genome, leading to well characterized SNPs, their frequencies in multiple populations, and the extent of linkage disequilibrium between them. Linkage disequilibrium is explained by the fact that each disease-causing mutation arises on a particular copy of the human genome and bears a specific set of common alleles in cis at nearby loci, termed a haplotype (figure 3). Because the recombination rate is low [~1 crossover per 100 megabases (Mb) per generation], disease alleles in the population typically show association with nearby marker alleles for many generations, a phenomenon termed linkage disequilibrium (LD) (Altshuler et al., 2008). The linkage disequilibrium has the great practical advantage to allow selection of only the most

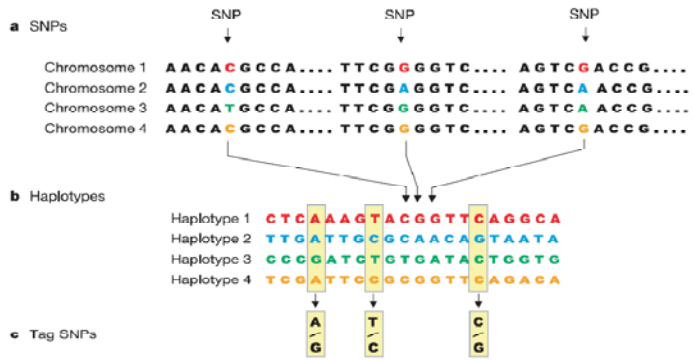


Figure 3: Copied from HapMap project

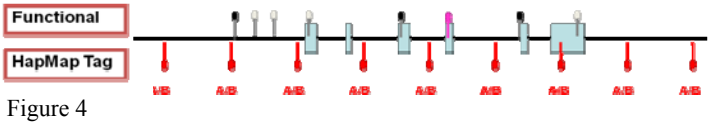


Figure 4

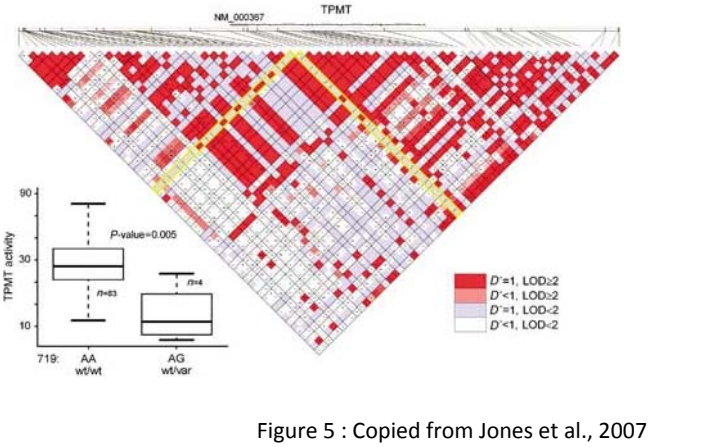


Figure 5 : Copied from Jones et al., 2007

relevant SNPs (tagSNPs) that will provide enough information to predict the frequency and incidence of the remaining common SNPs, thus only "tag SNPs" are required to be part of a genotyping panel and for subsequent analysis in further experiments (figure 3) (The International HapMap consortium, 2003). The combination of tagSNPs and well characterized functional markers permit the highest possible gene coverage, making possible the functional marker-phenotype association as well as a haplotype based association (see figure 4). A recent study (Jones et al., 2007) has demonstrated the use of a combined strategy, which is the use of candidate genes complemented with HapMap resources, giving more strength for finding the associated markers. They demonstrate that the use of HapMap markers and a candidate gene approach have led to identification of the well established TPMT polymorphism (719 A>G), which is linked with the clinical loss of TPMT activity and consequently higher drug related toxicity (see figure 5).

2.4.3. Towards clinical translation

Despite the accumulation of publications in pharmacogenomics showing the clinical benefits for many biological processes, the translation towards the clinic is not in its optimal phase. The lead reason that explain the slow translation of pharmacogenomics to clinical practice is the absence of important randomized studies using pharmacogenetics markers that show positive clinical outcomes (lest negative clinical events, better drug response and less drug toxicities). The main steps for clinical implementation of pharmacogenomics include (See figure 6): A) execution of large prospective clinical studies showing the benefit of pharmacogenomics; B) developing

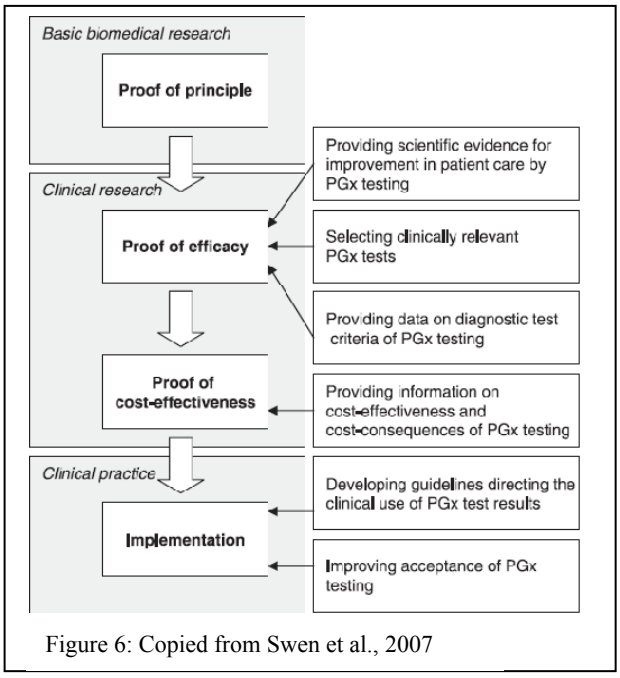


Figure 6: Copied from Swen et al., 2007

robust and validated tests to support the specific molecular profiling assays; C) development of specific clinical guidelines; D) education of clinicians and other health care professionals who interact with patients; E) creation of ethical guidelines (Manolopoulos, 2007). The lack of approved diagnostic tests and the lack of robust genetic assays are the two most critical technical issues today. However, recent improvements in the technologies that support the analysis of variant alleles that underline the ADR or drug inefficiency have resulted in a number of potentially commercial assays. With the recent advancement of available robust and accurate technologies combined with the recent introduction of FDA-approved clinical assays, such as the Amplichip from Roche and the Invader UGT1A1*28 assay, the translation to the clinic is on the way to common use. However, one of the major non-technical issues for the integration of pharmacogenomics into the clinic is still education of health professionals. A common perception of pharmacogenomics is that it represents an additional burden and complication regarding the therapeutic decision-making process. The implementation of pharmacogenomics training at undergraduate and postgraduate levels as well as continuing medical education will help inform professionals how to apply pharmacogenomics in modern medicine. In order to facilitate the integration of pharmacogenomic results into clinical application, some researchers have started to initiate the development of a pharmacogenomic clinical informatics systems (see figure 7). This application will bridge the gap between the current pharmacogenomic research and the reality of clinic application. The clinical translation will facilitate the incorporation of pharmacogenomics testing by informing the clinician about available pharmacogenomics tests, facilitate the transfer of relevant pharmacogenomics

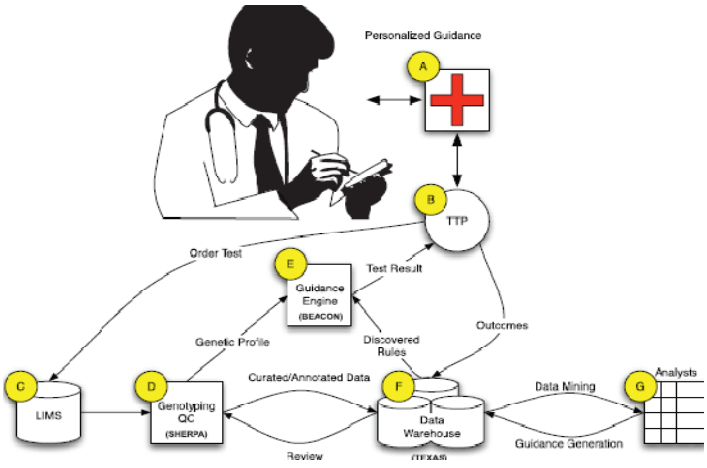


Figure 7: Copied from Van Rooij et al., 2008

information between the lab and clinical site and provide clinical guidance on drug prescriptions and relevant dose based on lab results (Van Rooij et al., 2008).

3. Genotyping and cutting edge technologies

The detection of causative polymorphisms that have the potential to impact protein expression or activity has always been of great interest among the scientific community. Before the arrival of PCR (polymerase chain reaction) in the mid 1980s, the first genotyping technique to be used to detect a single base mismatch was the hybridization of allele-specific oligonucleotides (ASO) and was first described in 1979 (Wallace et al., 1979). The technique was then used for the first time in 1983 to detect the sickle-cell mutation in the β-globin gene (Conner et al., 1983). With the invention of PCR, a broad spectrum of different genotyping techniques was developed. One of the first PCR-based genotyping techniques that was developed was SSCP (single-stranded conformation polymorphism) detection, where both alleles are separated on a non-denaturing polyacrylamide gel and detected following their migration patterns (Shi, 2001). This technique was relatively popular excluding the fact that it is labor intensive and low throughput. Another widely employed PCR-based genotyping technique was called PCR-RFLP. This involved PCR amplification followed by restriction enzyme digestion and became one of the most popular techniques in the 1980s and is still used today. This technique utilizes highly specific restriction enzymes that cleave specific DNA sequences. If the polymorphism creates or removes a restriction enzyme site present in the PCR fragments then the polymorphism will result in a specific digestion pattern that can be visualised using gel electrophoresis (Shi, 2001). The PCR-RFLP based genotyping technique, despite being low-throughput, remains today extremely popular and has been utilized in many recent studies and clinical applications. For instance, this technique is currently used in the clinic to detect variability in the HLA region (Human leukocyte antigen) (Doxiadis et al., 2003) as well as some cytochrome P450 variations (Jannetto et al., 2004). However these PCR-gel based strategies are mainly used for single gene

analysis in a low throughput context, and with the growing interest for polymorphism detection in polygenic diseases, such strategies are not viable.

Over the past 10 years, an impressive number of different non-gel-based high-throughput technologies have been developed in order to support large diseases-based genetic polymorphism association studies as well as for supporting large-scale pharmacogenomics studies. Two main categories of technologies have been developed, the medium-high-throughput and very high-throughput technologies. Each category of technologies has been developed to support different needs. In general, the medium-high-throughput technologies are most suitable for a large number of samples (>384) to be screened for a few genetic polymorphisms (~12 to 50), whereas the very-high-throughput technologies are most suitable for a large amount of genetic polymorphisms (~1536 to 1,000,000) for a low number of samples (<96).

3.1. Medium-High-Throughput technologies

This category of technology primarily utilizes a PCR-based strategy, but instead of identifying the genetic polymorphisms using gel electrophoresis, the detection is performed using either the incorporation of fluorescence or mass-based changes in the PCR amplicons. These are then analysed using different types of DNA array matrices. Typically, this category of technology can multiplex between 12 and 50 different genomic regions in a single PCR reaction and can analyse 384 samples simultaneously. In general, the size of the PCR amplification containing the genetic variation is limited to ~100 to 200 bp in order to amplify uniformly all of the

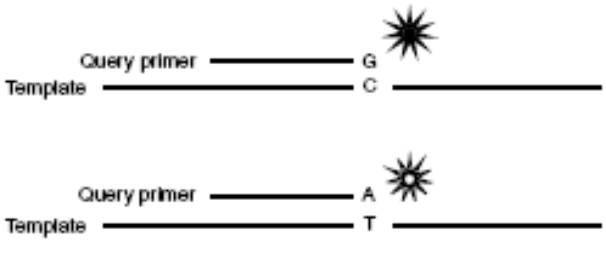


Figure 8A: Copied from Brennan et al., 2001

genomic regions and reduce amplification competition among the PCR reactions within the same tube.

From the multiplex PCR reaction, depending on which technology the downstream genotyping process is processed; the detection of the polymorphic sites is typically achieved using either SBE (Single base extension) or ASE (Allele specific extension). Single base extension tends to be more robust and accurate than the ASE technique since the SNP detection is based on the high accuracy of nucleotide incorporation by the DNA polymerase (Syvanen et al., 2001) (see figure 8A). Basically the same extension probe is hybridized to the pre-amplified PCR template and then the SNPs are extended by one base using a modified DNA polymerase and nucleotides. Since the extension probe for a specific polymorphism has the same affinity for all different alleles, the reaction conditions remain similar giving robust and reliable genotyping results. On the other hand, the ASE technique (see figure 8B) utilizes two different extension probes to detect a specific genetic polymorphism, then the genotyping specificity is driven by the affinity of both extension probes for the PCR template. Since precise and accurate hybridization is required for specific allele extension of the probe, most of the time this technique requires rigorous optimisation to allow for uniform and similar annealing hybridization to accurately discriminate both alleles. This is more complex when multiplexing is involved. However, both genotyping techniques are represented among the most cited technologies. SBE is the method of choice for many known medium-throughput technologies as such as, SNaPshot from Applied Biosystems (Foster City, CA), SNPstream from Beckman Coulter (Fullerton, CA) and MassArray from Sequenom (La Jolla, CA), whereas the ASE detection is utilized by Taqman from Applied Biosystems (Foster City, CA), Tag-It from Luminex Molecular Diagnostic (Austin, TX) and Infinity from Autogenomics

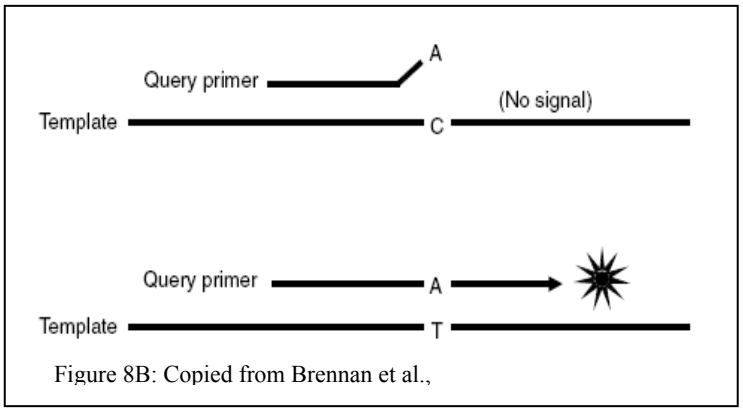


Figure 8B: Copied from Brennan et al.,

(Carlsbad, CA) (Syvanen, 2001 and King et al., 2008). For all these technologies, the detection method depends on specific PCR amplification of the genomic region encompassing the genetic polymorphism. Specific PCR amplification is required to precisely and quantitatively amplify the desired genomic region to a degree where the detection method can adequately distinguish between homozygotes and heterozygotes for the SNPs of interest. The difficulty of designing and performing multiplexed PCR reactions is an important factor that limits the throughput of SNPs within each reaction.

These technologies have been used for many medium sized studies as well as to support clinical assays. In the context of fundamental research applications, these technologies have been utilized to support the fine SNP mapping of the associated genomic loci or to validate associated SNPs in a replicate cohort (Sladek et al., 2007). Since these technologies are completely customizable and give accurate results, diagnostic and clinical applications utilizing them are possible.

4. Blood antigens: Red blood cells and platelet antigens

Blood group antigens are proteins, glycoproteins or glycolipids that are located on the exofacial surface of the RBC (red blood cell) and platelet membranes. They are known to be genetically polymorphic. Inheritance of these variations results in variability in the structural characteristics of an individual's blood cells. For red blood cells, there are 30 blood group antigen systems whereas for platelets there are 24 antigens. The genetic bases of the carbohydrate–dependent blood groups are established by variation in glycosyl-transferase DNA sequences expressed by the erythroid and platelet cells, which are known to affect the enzyme specificity or the efficacy of polysaccharide synthesis occurring in the Golgi compartment (Avent et al., 2007). The most clinically relevant RBC antigen systems are related to the structural erythrocyte membrane proteins: ABO (ABO), RH (D, cC, eE), Duffy-DARC (Fyab), Kidd-JK (Jkab), Kel and XK (kK and Kapab), MNS- GPYA-GPYB

(MN, sS), Colton-AQP1 (CO) and Dombrock-ART4 (DO) (Avent et al., 2007). These proteins have different functions in the cell, such as; transporters and channels (RH, JK, CO), structural functions (GPYA and GPYB), chemokine receptors (FY), and membrane-bound proteins (KEL and DO). The most relevant platelet antigens are HPA1-ab, HPA2-ab, HPA3-ab, HPA4-ab, HPA5-ab and HPA15-ab.

The cells carrying a particular antigen can, if introduced into the circulation of patient with difference minor blood cell antigens, produce an immune response. The antibodies produced by the patient against the foreign antigens can result in problems in transfusion incompatibility, maternal-fetal incompatibility and autoimmune hemolytic anemia (Reid, 2003). In many countries, only major blood antigens ABO and D are routinely tested for in patients receiving infrequent blood transfusions, however chronically transfused patients that are suffering from sickle cell diseases, thalassaemia, autoimmune hemolytic anemia and plastic anemia need closer monitoring of their RBC and platelet antigen profiles (Polin et al., 2008). Additionally, it has been described by the National Heart, Lung and Blood institute that the incidence rate for alloimmunization for patients suffering from sickle cell disease and receiving blood, is 20 percent or higher compared to 5 percent in the other transfusion-dependent patients (Reid, 2007). The major blood antigens, ABO, are rarely involved in HDN (hemolytic disease of the fetus and newborn), in contrast the minor antigens Rh, Kell, Duffy and Kidd are seriously implicated in neonatal alloimmunization. In the case of platelets, limited numbers of platelet transfusion-dependent patients develop HPA antibodies; however the HPA profile is often required in the case of neonatal alloimmune-mediated thrombocytopenic purpura (Beiboer et al., 2005).

4.1. Goodbye hemagglutination, welcome DNA based assays

Traditionally, hemagglutination has been the principal analytical tool of immunohematology to determine blood antigens. This technique was used for the

first time by Karl Landsteiner in 1901 when he discovered the major ABO blood antigens and was then modified by Coombs, Mourant and Race when they discovered many other minor blood antigens. The principal of hemagglutination is the antibody-mediated clumping of particles that express antigens on their surface. Basically, serum containing a known antibody is added to a suspension of RBCs and if the cells carry the equivalent antigen they are agglutinated, if no agglutination appears the cells are lacking the antigen (Anstee, 2005). More recently, 29 of the 30 blood group genes have been cloned that encode more than 300 blood group antigens. This has paved the way toward the development of DNA molecular based testing, especially for the minor blood group antigens where the majority of genetic bases are associated with only one SNP. However the genetic basis of the major blood group (ABO) antigens are quite complex, since more than 15 SNPs have to be typed to correlate the current hemagglutination test. Robust, reliable tests and reagents have been optimized for the major blood antigens (ABO and D) over the last twenty years, but at the same time the hemagglutination technique has shown a lack of robustness and reliability for the minor blood antigens. Even if the hemagglutination tests, when appropriately performed, have the sensitivity and specificity to support clinical needs, the technique has demonstrated limitations for many important clinical aspects such as unreliable prediction of a fetus at risk to develop hemolytic disease of newborn (HDN), difficulty to correctly type the RBCs of patient who recently received a blood transfusion, as their blood cells are coated with antibodies, and it does not predict D zygosity in D+ patients (Reid, 2003). In addition, only relatively small numbers of donors can be typed because of reagent shortages in reliable antisera due to limited availability of antigen-specific antisera inventories. Also it has been reported that many antibodies are not well characterized and some antibodies are weakly reactive or simply unavailable (Reid, 2003). This is also true for the platelet antigens where only the HPA1 assay has a specific monoclonal antibody available, whereas the other important HPA antigens can only by typed using fastidious human antisera-based assays (Reid, 2007). Because of the poor availability and poor performance for some of these reagents, the hemagglutination technique has been severely limited while at the same time being the accepted standard for worldwide blood-typing for the entire

20th century. With the challenges and inefficiencies of the hemagglutination assay coming to light and the advent of new genomic technologies, now is the time to change the paradigm and to evaluate the effectiveness of antigen specific genotyping as a tool for clinical blood banking.

5. Scientific interest and global objectives

Since it is widely accepted that genetic variation contributes to variability in disease phenotypes and drug responsiveness, the development of DNA based molecular profiling assays has become more established and has led to the promise of personalized medicine. From the drive to develop many meaningful genetic assays, many different techniques and technologies have been developed to support the identification of genetic variants and therefore contribute to the association between DNA makeup, diseases and xenobiotics responses. Historically, because of technical issues and statistical power, the interaction of diseases or drug response with DNA composition has only evaluated one gene at the time for a limited number of genetic markers. With the recent advancement in genotyping technologies, reduction in costs, the capacity to multiplex many genes and the ability to screen many markers in a single reaction, the use of genetic biomarkers in clinical research has been greatly accelerated. With the numerous possible applications brought on by these cutting-edge technologies, we have set out to develop many different DNA based panels to support pharmacogenomic needs in the context of the clinical environment. The actual worldwide interest for personalized medicine has increased significantly since the development of new DNA-based technologies but, up to now, only a few clinically approved DNA-based assays are available and used for specific applications. In that context, along with the development of clinical grade pharmacogenomics assays, we have created a unique Canadian Clinical Centre environment dedicated to pharmacogenomics research. One of the goals of our Centre to is to set the standards for pharmacogenomics panel development and

operations. Our Centre is presently working to become recognized as supporting clinical research under GLP (good laboratory practice) standards and to become accredited by the College of American Pathologist (CAP and CLIA). Until we reach this accreditation, we are developing and validating all DNA-based assays following rigorous clinical guidelines in order to support pharmaceutical development projects or clinical trials, as well as biotechnology companies and hospital network projects.

Since using the traditional hemagglutination techniques to screen blood banking donors for compatible blood can be very problematic and sometimes impossible, and because, as described earlier, hemagglutination is the main technique used to type the worldwide blood reserve. As part of my thesis research, we wanted to develop a clinical grade genetic panel that could improve on the efficiency and cost of the hemagglutination assay and that would support the creation of a minor blood group antigens electronic database for donors from Héma-Québec. Using knowledge of DNA-based panel development previously acquired at the Pharmacogenomics Centre, we plan to create a unique pharmacogenomics panel of 11 DNA markers that could infer the phenotypes of 22 of most clinically-relevant minor blood group antigens using a medium-high-throughput technology. The development process will consist of a development phase followed by a clinical validation phase. Once the validation phase is completed and accepted by Héma-Québec, we will develop a set of clinical laboratory standard operational procedures (SOPs) that will permit the integration of the pharmacogenomics minor blood group antigen genotyping panel into Héma-Québec's clinical processes. We will then conduct a proof of concept project to evaluate how the use of this genotyping panel and the creation of a database of molecularly profiled donors can improve Héma-Québec's operations. The proof of concept study will consist of the screening of 22,000 patients in order to create the first North America blood bank electronic register for minor blood antigens.

The molecular profiling data in the database will be used to:

- Help assign appropriate donors to patients requiring frequent blood transfusions

- Identify a fetus at risk for HDN
- Detect weakly expressed antigens (e.g., Fyb with the FyX phenotype); where the patient is unlikely to make antibodies to transfused antigen-positive RBCs
- Replace antibody hemagglutination assays that are weak or not available,
- Distinguish an alloantibody from an autoantibody
- Identify molecular basis of unusual serologic results, especially Rh variants
- Permit large scale screening of antigen-negative donors to create an inventory of donors whose RBCs lack a high-prevalence antigen
- Detect genes that encode weak antigens

MATERIALS AND METHODS

1. Gene and marker selection

The nine genes and eleven markers presented in table 1 were selected by the clinical and the development departments of Héma-Québec (HQ). Briefly these variations were selected on the basis of the clinical requirements of regulatory agencies (Canadian, American as well as European) as well as the specific Quebec requirements regarding frequencies (above 5%) and relevance of these antigens in the population.

Table 1: Genes, markers and antigens details

Gene	Antigens	PLT or RBC	Related Genomic marker	Genomic variation	Amino Acid Change
Kel	Kp a/b	RBC	rs8176059	CGG → TGG	Arg → Trp
ITGbIII	HPA1 a/b	PLT	rs5918	CTG → CCG	Leu → Pro
GYPA	M/N	RBC	rs7682260	TTA → TCA	Leu → Ser
GYPB	s/S	RBC	rs7683365	ACG → ATG	Thr → Met
RHCE	c/C	RBC	rs1053344	AAT → AGT	Asn → Ser
GP1BA	HPA2 a/b	PLT	rs6065	ACG → ATG	Thr → Met
DARC	Fy a/b	RBC	rs12075	GGT → GAT	Gly → Asp
SLC14A1	Jk a/b	RBC	rs1058396	GAC → GGC	Asp → Gly
Kel	k/K	RBC	rs8176058	ACG → ATG	Thr → Met
RHCE	e/E	RBC	rs609320	GCT → CCT	Ala → Pro
ITGA2	HPA5 a/b	PLT	rs10471371	GAG → AAG	Glu → Lys

* RBC= Red blood cells antigens, PLT= Platelets antigens

This table describes the genotyping minor blood antigen panel content. For each antigen the table shows their public database IDs (related genomic marker), nucleotide variations and the downstream amino acid changes.

2. DNA sources

Panel Development: DNAs from 96 Caucasian, 96 African-American and 89 Asian unrelated healthy individuals were obtained from the Coriell Institute for Medical Research (Camden, NJ) following completion of the appropriate assurance and research statement forms. All samples were thawed at room temperature and 8.3µL (all original samples concentrations were around 300ng/µL) of all samples were transferred into three 96-well plates. The volumes were completed to 200µL

with PCR grade water to yield a final concentration of ~12.5ng/µL. FTA cards with blood were also received from Héma-Québec for testing the punching and DNA extraction techniques, as well as the genotyping conditions.

Panel Validation: Peripheral blood samples were collected from 618 randomized volunteers from the province of Québec after signature of an informed consent form approved by Héma-Québec's Research Ethics Committee. Samples consisted of a blood drop applied on an FTA card. All samples were anonymized and the FTA cards were sent to Génome Québec in Montreal for genotyping. These samples were used for the genotyping blood antigen panel (GBAP) validation since a good proportion of these blood samples were also phenotyped at Héma-Québec for the typical minor blood group antigens using standard serology techniques.

3. DNA extraction from FTA cards

DNA was extracted from the FTA cards using the GENERATION DNA purification technology (Qiagen, Mississauga, ON). To test which FTA paper punch size generated the highest extracted DNA yield, 2mm, 3mm and 4mm punches were made using a Harris Uni-Core Puncher and distributed in a standard 96-well plate containing 1µL of PCR grade water whereas the 6mm punches were distributed in a deep well plate containing 1µL of PCR grade water. Three blank punches were generated between each sample to clean the puncher. During the validation steps only the 2mm punch size was used. To remove haemoglobin contained in the FTA card, 150 µL of Buffer 1 (DNA purification solution; G1-1000) was added to each well using a 12-channel pipette and then incubated at room temperature for 20 minutes. All samples were mixed three times using a 12-channel pipette with an up and down action and the coloured mixture was discarded. Two more washing cycles were performed with the Buffer 1 and Buffer 2 (DNA elution solution; G2-500). All remaining purification solution was completely removed. To extract DNA from the FTA card punch, 25 µL of Buffer 2 was added to each sample and incubated at 99°C for 15 minutes. After the incubation, the plates were pulse centrifuged and the entire contents of each well (without the FTA card) were transferred to a new 96-well plate.

The extracted DNAs were quantified with the Nanodrop system using 2 µL of undiluted mixture. The DNAs were also quantified using Picogreen (Invitrogen, Carlsbad, CA) diluted from 200X to 1X in TE buffer and the Varioskan (ThermoFisher, Waltham, MA) (Biotechniques, 1996).

3.1. Long PCR amplification on DNAs extracted from FTA cards

CYP2D6 (chromosome 22, positions 40853647 to 40855434) was amplified using PCR oligos 2D6_Prd7-F 5' CTGGAATCCGGTGTCGAAGTGG 3' and 2D6_Prd7-R 5' CGGCCCTGACACTCCTTCTTG 3' at 0.5 µM final concentration. The PCR master mix contained 1X PCR buffer, 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 5% DMSO and 0.3 U/µL of Platinum Taq (Invitrogen, Carlsbad, CA). 2 µL of DNA extracted from FTA cards was mixed with 8 µL of the master mix and amplified following the initial denaturation at 95°C for 2 minutes then 2 cycles of 92°C x 1 min., 59°C x 1 min., 68°C x 6 min. and 35 cycles of 92°C x 30 sec., 59°C x 30 sec. and 68°C x 6 min. 5 µL of the PCR amplification buffer and 3 µL of the 1 kb ladder (Invitrogen, Carlsbad, CA) were mixed with 10 µL of PCR grade water and loaded on a 1% agarose gel for band analysis.

APOE (chromosome 19, positions 50100128 to 50101071) was amplified using PCR oligos rs449647-PCRU3 5' GCATCATACTGTTCCCACCCCTC 3' and rs440446_PCRL2 5' CCGCTCCTCCTCTCCCCAAG 3' at 0.45 µM final concentration. The PCR master mix contained 1X PCR buffer, 1.5 mM $MgCl_2$, 0.1 mM dNTPs, 1X Qsolution and 0.06 U/µL of Qiagen HotStart Taq (Qiagen, Mississauga, ON). The amplification was obtained using 2 µL of DNA and 3 µL of master mix and amplified with the initial denaturation at 95°C for 15 minutes then 12 cylces of 95°C x 30 sec., 66°C x 45 sec. (-0.5°C/cycle), 72°C x 45 sec. and 30 cycles of 95°C x 30 sec., 60°C x 45 sec., and 72°C x 45 sec. 5 µL of the PCR amplification buffer and 3 µL of the 1 kb ladder (Invitrogen, Carlsbad, CA) were mixed with 10 µL of PCR grade water and loaded on a 1% agarose gel for band analysis.

4. Genotyping

4.1. Primer Design

DNA sequences flanking each SNP were tested for the presence of repeats or duplicated regions and masked if needed using the BLAT program (http://www.genome.ucsc.edu). All DNA sequences, except for c/C, e/E, M/N and s/S antigens, were formatted using the appropriate SNPstream and Sequenom oligonucleotide design software programs. For the first phase of development, the PCR and extension primers were designed using the Autoprimer (http://www.autoprimer.com) and Sequenom AssayDesign software for SNPstream and Sequenom technologies respectively. For the SNPstream technology, the program selects PCR primers that will generate products ranging between 80 and 200 bp and an optimized single base-pair extension primer 5' to the SNP site and assembles them into panels of 12 SNPs of the same extension type (i.e. G/A or C/T extension mixes). Also, the Autoprimer program added a unique tag address of 20 bases at the 5' of each extension oligo in order to make the hybridization possible onto the microarray SNPware plate (Bell *et al.* 2002). Similar to the Autoprimer program, the AssayDesign software designs PCR primers that will generate amplicons ranging from 80 to 200 bp. The software also suggests the appropriate single base-pair extensions ranging in size from 18 bp to 35bp in order to separate them from the extended SNPs according to their respective masses. The AssayDesign software also adds 10 extra bases (ACGTTGGATG) at the 5'end of the PCR oligo to increase the total mass above 10,000 Daltons. This brings the detected PCR oligo masses outside of the analyzed window (3500 to 9000 Daltons) in the mass spectrum. Both the autoprimer and AssayDesign software are dedicated to design only one oligo set per SNP at a time (one set meaning two PCR oligos and one extension primer). Since the minor blood group antigen genotyping panel design required assays for both DNA directions, the second extension primers were designed using a modified input file that directed the extension oligo design in the opposite direction of the initial set. The PCR and extension oligos of c/C, e/E, M/N and s/S antigens were manually designed to overcome genomic challenges and SNPstream platform

specific issues respectively. The output design was then submitted to blast analysis (http://www.ncbi.nlm.nih.gov/) and to *in*Silico PCR (http://www.genome.ucsc.edu) to ensure the uniqueness of all PCR oligos and expected amplicon sizes. Since all PCR oligos and extension primers need to work in a unique multiplex reaction, all oligo sequences were assayed for secondary structures and for possible self and inter-oligo priming issues using the software FastPCR (http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm). In subsequent design rounds, modifications were made to the PCR and extension primer compositions to resolve genomic related problems, PCR amplification specificities, and lack of extension performance.

4.2. Multiplex PCR amplification

Twelve-plex PCR reactions were performed in 384-well plates (MJS BioLynx, Brockville, ON) in a 5 µL volume using 2 µL of ~12.5 ng/µL of Coriell Institute of Medical Research (Camden, NJ) DNAs and 2 µL of extracted DNA from the FTA cards. The remainder of the volume consisted of 75 µM dNTPs, 4.5 mM $MgCl_2$, 0.5 U Hotstart DNA polymerase (Qiagen, Mississauga, ON), and the 24 PCR primers at a concentration varying from 50 nM to 100 nM each in 1 X PCR buffer (refer to tables 2 and 3 for PCR oligo details). Thermal cycling was performed using GeneAmp PCR system 9700 thermal cyclers (Applied Biosystems, Foster City, CA) with the following program: initial denaturation at 95°C for 15 min. followed by 45 cycles of 95°C for 30 sec., 55°C for 55 sec., and 72°C for 30 sec. After the last cycle, the reaction was held at 72°C for 7 min. Following multiplex PCR using SNPstream oligos and plates were centrifuged briefly and 3 µL of a mixture containing 0.67 U Exonuclease I (USB Corporation, Cleveland, OH) and 0.33 U Shrimp Alkaline Phosphatase (USB Corporation, Cleveland, OH) were added to each well. Following the multiplex PCR using Sequenom oligos, 2 µL of Sequenom Shrimp Alkaline Phosphatase mixture was added to each well. The plates were sealed and incubated for 30 min. at 37°C and at 95°C for 10 min. During the panel validation only the SNPstream and Sequenom PCR oligo mixes versions 2 and 1 were used (Table 2).

Table 2: SNPstream PCR oligos mix details

	VERSION 1						VERSION 2									
Antigen	Oligo IDs	Sequence (5'3')	Tm	GC%	Nb	Amplicon size (bp)	Oligo IDs	Sequence (5'3')	Tm	GC%	Nb	Amplicon size (bp)	Concentration (µM)	Volume (µL)	Final Concentration (µM)	Final Volume (µL)
K/k	KelEx6K1_2F-s2	CGATCCTTATGCTCAGCCC	57.6	60	19	157	KelEx6K1_2-F-5	CGATCCTTATGCTCAGCCC	57.6	60	19	157	100	10	1	1000 µL
	KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	60.4	56	25		KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	60.4	56	25		100	10	1	
Kap a/b	KelExon8Kpa_bF	AGCAAGGTGCAAGAACACT	54.5	47	19	101	KelExon8Kpa_bF	AGCAAGGTGCAAGAACACT	54.5	47	19	101	100	10	1	
	KelExon8Kpa_bR	AGAGCTTGCCCTGTGCCC	60.3	67	18		KelExon8Kpa_bR	AGAGCTTGCCCTGTGCCC	60.3	67	18		100	10	1	
M/N	285-GYPAs-HD9131	CAACTTCTATGTTATACAGC	45.7	35	20	98	285-GYPAs-HD9131_1	TGAGGGAATTTGTCTTTGCA	53	38	21	98	100	10	1	
	287-GYPAas-HD9133	TCTTTGTGACTGAAGAAGAA	49.3	40	20		GYPAex2M_N-R-4	TCTTTGTGACTGAAGAAGAR	49.3	40	20		100	10	1	
s/S	GYPBEx4s_SF-s2	TTGTCAAATATTAACATACCTGGTACAGT	54	31	29	165	GYPBEx4s_S-F-6	CTGGTACAGTGAAACGATGG	53	50	20	86	100	10	1	
	GYPBEx4s_SR-s2	AATATGATTAAGAAAAGGAAACCCG	51.6	32	25		GYPBEx4s_S-R-6	GCACATGTCTTWCTTATTTGG	50	38	21		100	10	1	
HPA2 a/b	GPIBHPA-2a_bF-s2	CTCAGTCAAGTTGTTGTTAGCCAGAC	57.7	46	26	102	GPIBHPA-2a_bF-s2	CTCAGTCAAGTTGTTGTTAGCCAGAC	57.7	46	26	102	100	10	1	
	GPIBHPA-2a_bR-s2	CTCTACCTGAAAGGCAATGAGCT	56.6	48	23		GPIBHPA-2a_bR-s2	CTCTACCTGAAAGGCAATGAGCT	56.6	48	23		100	10	1	
HPA1 a/b	GPIIIAex3HPA-1a_bF	ATTCTGGGGCACAGTTATCC	54.5	50	20	114	GPIIIAex3HPA-1a_bF	ATTCTGGGGCACAGTTATCC	54.5	50	20	114	100	10	1	
	GPIIIAex3HPA-1a_bR	ATAGTCTGATTGCTGGACTTC	54.3	46	22		GPIII3HPA-1a_b-R-5	ATAGTCTGATTGCTGGACTTC	54.3	46	22		100	10	1	
e/E	RHCE_e-F-s2	TGGATGTTCTGGCCAAGTG	55.1	53	19	107	RHCE_e-F-s2	TGGATGTTCTGGCCAAGTG	55.1	53	19	107	100	10	1	
	RHCE_e-R-s2	CTGTCACCACACTGACTGCTAG	57.1	55	22		RHCE_e-R-s2	CTGTCACCACACTGACTGCTAG	57.1	55	22		100	10	1	
c/C	RHCEc_CF-s3	TCCCCCTCCTCCTTCTCA	56.2	61	18	143	RHCEc_CF-s3	TCCCCCTCCTCCTTCTCA	56.2	61	18	143	100	10	1	
	289-RHCEas-HD9135	GCCAGCATGAAGAGGTTGAA	55.4	50	20		289-RHCEas-HD9135	GCCAGCATGAAGAGGTTGAA	55.4	50	20		100	10	1	
Fy a/b	FYex2Fya_bF	AGTCATCCAGCAGGTTACAGG	56.3	52	21	111	FYex2Fya_b-F-5	AGTCATCCAGCAGGTTACAGG	56.3	52	21	111	100	10	1	
	FYex2Fya_bR	AAGATGTATGGAATTCTTCCTATGG	52.3	36	25		FYex2Fya_bR	AAGATGTATGGAATTCTTCCTATGG	52.3	36	25		100	10	1	
HPA5 a/b	GPIaHPA5_F	GCTCTTGGTAGGTGCACCAATGT	59.5	52	23	148	GPIaHPA5_F	GCTCTTGGTAGGTGCACCAATGT	59.5	52	23	148	100	10	1	
	GPIaHPA5_R	TTCCAAATGCAAGTTAAATTACCAG	52.4	32	25		GPIaHPA5_R	TTCCAAATGCAAGTTAAATTACCAG	52.4	32	25		100	10	1	
Jk a/b	JKa_b_F-s3	CATAGGATCATTGCTGGGCATAG	55.5	48	23	301	JKa_b_F-s4	CCTGCTAACTTTCAATCCCACCCTC	59.6	52	25	194	100	10	1	
	JKa_b_R-s3	TTGAAACCCCAGAGTCCAAAGT	56.2	46	22		JKa_b_R-s4	TGCCAGGTGAGCGCCATGAAC	62	62	21		100	10	1	
c/C-Intron	RHCintron-F	CAGGGCCACCACCATTTGAA	58.1	55	20	150	RHCintron-F	CAGGGCCACCACCATTTGAA	58.1	55	20	150	100	10	1	
	RHCintron-R	TGGTAGCAGGCGTCTGTAAAAAA	56.6	44	23		RHCintron-R	TGGTAGCAGGCGTCTGTAAAAAA	56.6	44	23		100	10	1	

Version one, except for the antigens M/N, s/S, c/C, c/C-intron and e/E where they were manually designed, represents the automatic SNPstream PCR oligo output results obtained after the initial design phase using the Autoprimer software. The nucleotides highlighted in white for version 2 represent the manual changes applied on the initial design to increase the specificity and quality of the genotyping results. In both versions, the oligos were mixed together and volumes increased to 1.0 mL with PCR grade water to reach a PCR oligo mix final concentration of 1 µM .

Table 3: Sequenom PCR oligos mix details

Antigen	Oligo IDs	Sequence (5'3')	Tm	GC%	Nb	Amplicon size (bp)	Concentration (µM)	Volume (µL)	Final Concentration (µM)	Final Volume (µL)
K/k	KelEx6K1-F-4	ACGTTGGATGTGGCGCATCTCTGGTAAATG	55.2	50	30	99	100	10	1	1000 µL
	KelEx6K1-R-4	ACGTTGGATGGGAAATGGCCATACTGACTC	53.2	50	30		100	10	1	
Kap a/b	KelExon8Kpa-F-4	ACGTTGGATGCTTCCTTGTCAATCTCCATC	50.5	45	30	113	100	10	1	
	KelExon8Kpa-R-4	ACGTTGGATGTGACCATCTGGAAGAGCTTG	54.2	50	30		100	10	1	
M/N	GYPAex2M_N-F-4	ACGTTGGATGTGAGGGAATTTGTCTTTTGCA	52.6	38	31	181	100	10	1	
	GYPAex2M_N-R-4	ACGTTGGATGTCTTTGTGACTGAAGAAGAR	48.9	38	30		100	10	1	
s/S	GYPBEx4s_S-F-4	ACGTTGGATGCTGGTACAGTGAAACGATGG	53.2	50	30	150	100	10	1	
	GYPBEx4s_S-R-4	ACGTTGGATGGCACATGTCTTWCTTATTTGG	49.7	38	31		100	10	1	
HPA2 a/b	GPIBex2HPA-2a_b-F-4	ACGTTGGATGACCTGAAAGGCAATGAGCTG	55.5	50	30	102	100	10	1	
	GPIBex2HPA-2a_b-R-4	ACGTTGGATGTTAGCCAGACTGAGCTTCTC	53.6	50	30		100	10	1	
HPA1 a/b	GPIIIAex3HPA1-F-4	ACGTTGGATGTTGCTGGACTTCTCTTTGGG	54.8	50	30	99	100	10	1	
	GPIIIAex3HPA1-R-4	ACGTTGGATGCAGATTCTCCTTCMGGTCAC	53.6	50	30		100	10	1	
e/E	RHCE_e-F-4	ACGTTGGATGTGGATGTTCTGGCCAAGTG	55.1	53	29	126	100	10	1	
	RHCE_e-R-4	ACGTTGGATGCTGTCACCACACTGACTGCTAG	57.1	55	32		100	10	1	
c/C	RHCEc_C-F-4	ACGTTGGATGTCCCCCTCCTCCTTCTCA	56.2	61	28	163	100	10	1	
	RHCEc_C-R-4	ACGTTGGATGGCCAGCATGAAGAGGTTGAA	55.4	50	30		100	10	1	
Fy a/b	Fya_b-F-4	ACGTTGGATGTGATTCCTTCCCAGATGGAG	53.6	50	30	99	100	10	1	
	Fya_b-R-4	ACGTTGGATGAGWCATCCAGCAGGTTACAG	53.9	50	30		100	10	1	
HPA5 a/b	GPIaHPA5-F-4	ACGTTGGATGGGAAGAGTCTACCTGTTTAC	49.5	45	30	116	100	10	1	
	GPIaHPA5-R-4	ACGTTGGATGCCAAATGCAAGTTAAATTACC	48.3	33	31		100	10	1	
Jk a/b	JKa_b-F-4	ACGTTGGATGTGAAACCCCAGAGTCCAAAG	54.6	50	30	109	100	10	1	
	JKa_b-R-4	ACGTTGGATGTTAGTCCTGAGTTCTGACCC	53	50	30		100	10	1	
c/C-Intron	RHCintron-F-4	ACGTTGGATGCAGGGCCACCACCATTTGAA	58.1	55	30	150	100	10	1	
	RHCintron-R-4	ACGTTGGATGTGGTAGCAGGCGTCGTAAAAA	56.4	48	31		100	10	1	

Except for the antigens M/N, s/S, c/C, c/C-intron and e/E where manual designed was performed, the table represents the automatic Sequenom PCR oligo output results obtained after the initial design phase using the AssayDesign software. This version of the PCR oligo design was used to cross-compare and validates the genotyping results with the SNPstream panel. The oligos were mixed together and volumes were increased to 1.0 mL with PCR grade water to reach a PCR oligo mix final concentration of 1 µM.

4.3. Single-Base multiplex extension reaction and genotyping scan

SNPstream technology: Genotyping extension reactions and hybridizations to the SNPware 12-plex microarray plates were carried out as described by Bell *et al.*, 2002. Briefly, the extension mixes were prepared in a 7 µL volume using 3.77 µL of SNPware extension mix diluents (Beckman Coulter, Fullerton, CA), 0.2 µL of the SNPware 12-Plex extension mixes GA and TC, containing two dideoxynuclotides labelled with either BODIPY-Fluorescein or TAMRA dye (Beckman Coulter, Fullerton, CA), 2.95 µL of PCR grade water, 0.017 µL of SNPware DNA polymerase (Beckman Coulter, Fullerton, CA) and 0.06 µL of GA and TC multiplex extension oligo mixes (refer to table 4). These mixes were added to each well of the previous multiplex PCR reactions. Thermal cycling was performed in GeneAmp PCR system 9700 thermal cyclers (Applied Biosystems, Forster City, CA) using the following program: initial denaturation at 96°C for 3 min. followed by 45 cycles of 94°C for 20 sec. and 40°C for 11 sec. The SNPware 12-plex microarray plates were conditioned three times with buffer 1 (1X of non-stringent) prepared as per the manufacturer recommendations, and partially dried using a vacuum system. 8 µL of hybridization mixture containing 7.56 µL of SNPware hybridization Solution and 0.44 µL of SNPware hybridization additive were added to each well of the extension reaction plates. After mixing up and down each samples ten times, 15 µL of mixture was transferred to the conditioned SNPware 12-plex microarray plate following the same sample-plate layout (A01 to A01...B01 to B01...). The transferred mixture was then uniformly distributed throughout all wells by gently tapping the corner of the microarray plate. The plates were then placed into a hybridization oven at 42°C for two hours in a humid container. After two hours, the microarray plates were washed out three times using ~20 µL of buffer 2 (1X of stringent buffer) prepared as per the manufacturer recommendation (Beckman Coulter, Fullerton, CA). The plates were dried out as much as possible with the vacuum system and then placed upside-down into the centrifuge for 5 min at 3000 RPM. The plates were read using the GenomeLab SNPstream Array Imager (Beckman Coulter, Fullerton, CA) and fluorescence intensity was measured with help of the SNPstream Imager software.

Intensity was plotted and genotypes were called by the GetGenos software. After visual inspection of the clusters, manual adjustments were made for some of the assays.

Sequenom technology: The extension mixes were prepared in 2 µL using 0.2 µL of iPlex buffer 10X (Sequenom, La Jolla, CA), 0.2 µL iPlex terminator mix containing four dideoxynuclotides modified with a unique mass component (Sequenom, La Jolla, CA), 0.76 µL of PCR grade water, 0.041 µL of iPlex DNA polymerase (Sequenom, La Jolla, CA) and finally 0.804 µL of multiplex extension oligos mixes (refer to table 5). Both 2 µL extension mixes were added to each well of the previous multiplex PCR reactions. Thermal cycling was performed in GeneAmp PCR system 9700 thermal cyclers (Applied Biosystems, Foster City, CA) using the following program: initial denaturation at 94°C for 30 sec. followed by 200 cycles of 94°C for 5 sec., 52°C for 5 sec., and 80°C for 5 sec. followed by a final extension of 3 min. at 72°C. 15 µL of PCR grade water and 6 mg of chelating resin were added to each well of the extension reaction plates. The plates were sealed, incubated for 15 min. by rotation at room temperature and spun down. The plates were then placed onto the nanodispenser deck (Sequenom, La Jolla, CA) and ~ 15 nL of the extension reactions and calibrator were spotted on the SpectroChip (Sequenom, La Jolla, CA). The chips were read using the Sequenom MassArray Maldi-TOF system (Sequenom, La Jolla, CA) and the different extended mass intensities were measured with the help of the Sequenom MassArray Imager software. Intensities were plotted and genotypes were called by the SequenomTyper software. After visual inspection of the clusters, manual adjustments were made for some of the assays.

Table 4: SNPstream GA and TC Extension oligo version details

	VERSION 1							VERSION 2								
Antigen	Oligo IDs	Sequence (5'3')	Extension Type	Hybridization position	Tm	GC%	Nb	Oligo IDs	Sequence (5'3')	Extension Type (Panel)	Tm	GC%	Nb	Stock Conc. µM	Vol. µL	Final Conc. µM	Final volume
K/k	KelEx6K1_2-GA-U5	GCGGTAGGTTCCCGACATATACTCATCAGAAGTCTCAGC	GA	U5	57.6	60	39	KelEx6K1_2-GA-U5	GCGGTAGGTTCCCGACATATACTCATCAGAAGTCTCAGC	GA	57.6	60	39	100	100	10	1000
	KelEx6K1_2U7	AGGGTCTCTACGCTGACGATTGGACTTCCTTAAACTTTAACCGAA	CT	U7	55	36	45	KelEx6K1_2U7	AGGGTCTCTACGCTGACGATTGGACTTCCTTAAACTTTAACCGAA	CT	55	36	45	100	50	5	
Kap a/b	KelExon8KGA_U1	ACGCACGTCCACGGTGATTTGCCTCAGAAACTGGAACAGCC	GA	U1	54.5	47	41	KelExon8KGA_U1	ACGCACGTCCACGGTGATTTGCCTCAGAAACTGGAACAGCC	GA	54.5	47	41	100	50	5	
	KelExon8Kpa_bU12	CGACTGTAGGTGCGTAACTCTTCCTTGTCAATCTCCATCACTTCA	CT	U12	56	40	45	KelExon8Kpa_bU12	CGACTGTAGGTGCGTAACTCTTCCTTGTCAATCTCCATCACTTCA	CT	56	40	45	100	50	5	
M/N	GYPAex2M_NGA-U2	GGATGGCGTTCCGTCCTATTGTGCATTGCCAC\YCAGTGGTACTT	GA	U2	45.7	35	45	GYPAex2M_NGA-U2_2	GGATGGCGTTCCGTCCTATTGTGCATTGCCAC\YCAGTGGTACTT	GA	45.7	35	45	100	50	5	
	GYPAex2M_NCTU11-s3_1	AGAGCGAGTGACGCATACTAACAG\AATTGTGAGCATATCAGCAT	CT	U11	55	36	45	GYPAex2M_NCTU11-s3_2	AGAGCGAGTGACGCATACTAACAG\AATTGTGAGCATATCAGCAT	CT	55	36	45	100	50	5	
s/S	GYPBs_S-GA-U7	AGGGTCTCTACGCTGACGATAAACGATGGACAAGTTGTCCC	GA	U7	54	31	41	GYPBs_S-GA-U7	AGGGTCTCTACGCTGACGATAAACGATGGACAAGTTGTCCC	GA	54	31	41	100	50	5	
	GYPBEx4s_S-CT-U2-s2	GGATGGCGTTCCGTCCTATTTTGAAATTTTGCTTTATAGGAGAAA	CT	U2	50	23	45	GYPBEx4s_S-CT-U2-s2	GGATGGCGTTCCGTCCTATTTTGAAATTTTGCTTTATAGGAGAAA	CT	50	23	45	100	50	5	
HPA2 a/b	GPIBex2HPA2GA-U10-s3	AGATAGAGTCGATGCCAGCTGAGCTTCTCCAGCTTGGGTGTGGGC	GA	U10	57.7	46	45	GPIBex2HPA2GA-U10-s3	AGATAGAGTCGATGCCAGCTGAGCTTCTCCAGCTTGGGTGTGGGC	GA	57.7	46	45	100	50	5	
	GPIBex2HPA-2a_bCTU9	GACCTGGGTGTCGATACCTAGACCCTGCCCCCAGGGCTCCTGA	CT	U9	68	73	43	GPIBex2HPA-2a_bCTU9	GACCTGGGTGTCGATACCTAGACCCTGCCCCCAGGGCTCCTGA	CT	68	73	43	100	50	5	
HPA1 a/b	GPIIIAex3HPA-1GAU8	GTGATTCTGTACGTGTCGCCCTTCAGGTCACAGCGAGGTGAGCCC	GA	U8	54.5	50	45	GPIIIAex3HPA-1GAU8	GTGATTCTGTACGTGTCGCCCTTCAGGTCACAGCGAGGTGAGCCC	GA	54.5	50	45	100	50	5	
	GPIIIAex3HPA1CTU8-s2	GTGATTCTGTACGTGTCGCCCTCTTTGGGCTCCTGTCTTACAGGCCCTGCCTC	CT	U8	68	60	53	GPIIIAex3HPA1CTU8-s2	GTGATTCTGTACGTGTCGCCCTCTTTGGGCTCCTGTCTTACAGGCCCTGCCTC	CT	68	60	53	100	50	5	
e/E	RHCEex5RbE_eCTU3-s3	CGTGCCGCTCGTGATAGAATCCTTTGGATTGGACTTCTCAGCAGAG	GA	U3	55.1	53	46	RHCEex5RbE_eCTU3-s3_1	CGTGCCGCTCGTGATAGAATCCTTTGGATTGGACTTCTCAGCAGAG	GA	59	50	46	100	50	5	
	RHCEex5RbE_eCTU1-s2_1	ACGCACGTCCACGGTGATTTCCTTTGGATTGGACTTCTCAGCAGAG	CT	U1	59	50	46	RHCEex5RbE_eCTU1-s2_1	ACGCACGTCCACGGTGATTTCCTTTGGATTGGACTTCTCAGCAGAG	CT	59	50	46	100	50	5	
c/C	RHCEEx2Rbc_CGA-U9	GACCTGGGTGTCGATACCTACCCTTGGCTTGGGCTTCCTCACCTC	GA	U9	56.2	61	45	RHCEEx2Rbc_CGA-U9	GACCTGGGTGTCGATACCTACCCTTGGCTTGGGCTTCCTCACCTC	GA	56.2	61	45	100	50	5	
	RHCEEx2Rbc_CctU6-s2	GGCTATGATTCGCAATGCTTTGCTCCAGCTGTGTCTCCGGAAA	CT	U6	63	57	43	RHCEEx2Rbc_CctU6-s2	GGCTATGATTCGCAATGCTTTGCTCCAGCTGTGTCTCCGGAAA	CT	63	57	43	100	50	5	
Fy a/b	FYex2Fya_b-U11	AGAGCGAGTGACGCATACTAGATTCCTTCCCAGATGGAGACTATG	GA	U11	56.3	52	45	FYex2Fya_b-U11	AGAGCGAGTGACGCATACTAGATTCCTTCCCAGATGGAGACTATG	GA	56.3	52	45	100	50	5	
	FYex2Fya_bU4	AGCGATCTGCGAGACCGTATGGGGGCAGCTGCTTCCAGGTTGGCA	CT	U4	69	68	45	FYex2Fya_bU4	AGCGATCTGCGAGACCGTATGGGGGCAGCTGCTTCCAGGTTGGCA	CT	69	68	45	100	50	5	
HPA5 a/b	GPIaHPA5-GA-U6	GGCTATGATTCGCAATGCTTGTCTACCTGTTTACTATCAAA	GA	U6	59.5	52	41	GPIaHPA5-GA-U6	GGCTATGATTCGCAATGCTTGTCTACCTGTTTACTATCAAA	GA	59.5	52	41	100	50	5	
	GPIaHPA5-CT-U3	CGTGCCGCTCGTGATAGAATTTAGTTTATTTTTTTTTTTTACCT	CT	U3	45	12	45	GPIaHPA5-CT-U3	CGTGCCGCTCGTGATAGAATTTAGTTTATTTTTTTTTTTTACCT	CT	45	12	45	100	50	5	
Jk a/b	JKa_b-U4	AGCGATCTGCGAGACCGTATACTCAGTCTTTCAGCCCCATTTG\G	GA	U4	55.5	48	45	JKa_b-U4_1	AGCGATCTGCGAGACCGTAT\TCAGTCTTTCAGCCCCATTTG\G	GA	58	50	44	100	50	5	
	JKa_b-CT-U5-s2_1	GCGGTAGGTTCCCGACATATGAAACCCCAGAGTCCAAAGTAGATGT	CT	U5	56.2	46	46	JKa_b-CT-U5-s2_1	GCGGTAGGTTCCCGACATATGAAACCCCAGAGTCCAAAGTAGATGT	CT	56.2	46	46	100	50	5	
c/C-Intron	RHCintron-Gins-U12	CGACTGTAGGTGCGTAACTCGGTGCCCTTTGTCACTTCCCA	GA	U12	58.1	55	41	RHCintron-Gins-U12	CGACTGTAGGTGCGTAACTCGGTGCCCTTTGTCACTTCCCA	GA	58.1	55	41	100	100	10	
	RHCintron-CT-U10	AGATAGAGTCGATGCCAGCTGGCAGAGGCTGCAATGAGCTATGATTGTAC	CT	U10	63	50	50	RHCintron-CT-U10	AGATAGAGTCGATGCCAGCTGGCAGAGGCTGCAATGAGCTATGATTGTAC	CT	63	50	50	100	100	10	

For version one, except for the antigens M/N, s/S, c/C, c/C-intron and e/E where the oligos were manually designed, the extension oligos were designed using the automatic Autoprimer software. The oligos nucleotides highlighted in white represent the manual changes applied on the initial design to increase the specificity and quality of the genotyping results. In both versions, the extension oligos for both ”GA and CT” reactions were mixed together with PCR grade water in two separate tubes in a final volume of 1.0 mL to reach extension oligo mixes of 5 and 10 µM final concentrations, respectively.

Table 5: Sequenom Forward-Reverse Extension oligos details

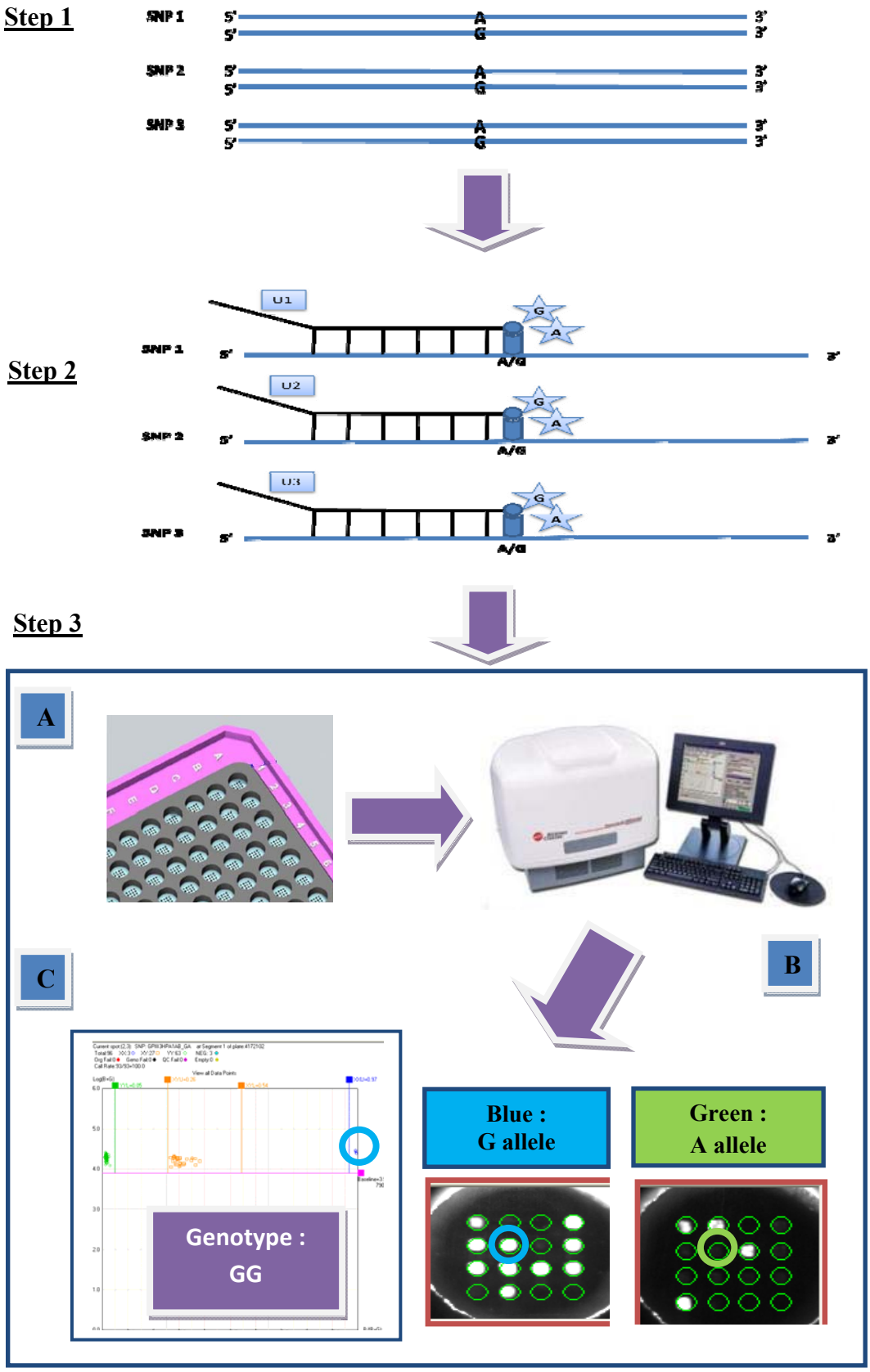
	Sequenom Extension oligos													
Antigen	Oligo IDs	Sequence (5'3')	Extension Type	Panel ID	Probe mass	Extension mass 1	Extension mass 2	Tm	GC%	Nb	Concentration (µM)	Volume (µL)	Final Concentration (µM)	Final Volume (µL)
K/k	KELEX6K1_R-2	TCATCAGAAGTCTCAGC	GA	F	5154.42	5441.62	5425.62	47.8	47	17	200	58.8	11.8	1000 µL
	KelEx6K1_F(CT)	ttcaTTAAACTTTAACCGAA	CT	R	6059.05	6306.25	6386.15	44.9	25	20	200	100	20.2	
Kap a/b	KELEXON8KPA_R-2	cGGCCTCAGAAACTGGAACAGCC	GA	F	7027.61	7314.81	7298.81	62.5	61	23	200	43	8.6	
	KelExon8Kpa_F(CT)	AATCTCCATCACTTCA	CT	R	4760.17	5007.37	5087.27	43.4	38	16	1000	8.6	8.6	
M/N	GYPAEX2M_N_R-2	GTGCATTGCCACCTCAGTGGTACTT	GA	F	7624	7911.2	7895.2	61.2	52	25	200	57.5	11.5	
	GYPAex2M_N_F(CT)	GTGAGCATATCAGCAT	CT	R	4905.27	5152.47	5232.37	45.5	44	16	200	114.2	22.8	
s/S	GYPBEX4S_S_R-2	AAACGATGGACAAGTTGTCCC	GA	F	6439.26	6710.46	6726.46	55.4	48	21	200	98.4	19.7	
	GYPBEx4s_S_F(CT)	TGCTTTATAGGAGAAA	CT	R	4944.32	5191.52	5271.42	40.2	31	16	200	64.7	12.9	
HPA2 a/b	GPIBEX2HPA-2A_B_R-4(GA)	TCCAGCTTGGGTGTGGGC	GA	F	5562.65	5849.85	5833.85	60.3	67	18	1000	14.9	14.9	
	GPIBex2HPA-2a_b_F(CT)	TGCCCCCAGGGCTCCTGA	CT	R	5436.55	5683.75	5763.65	63.3	72	18	200	79.1	15.8	
HPA1 a/b	GPIIIAEX3HPA1_R-2	ACAGCGAGGTGAGCCC	GA	F	4916.24	5203.44	5187.44	57	69	16	200	50.2	10.1	
	GPIIIAex3HPA1_F(CT)	TTACAGGCCCTGCCTC	CT	R	4793.15	5040.35	5120.25	53.8	63	16	200	50.9	10.2	
e/E	RHCEEX5RHE_E_R-2	aTGGATTGGACTTCTCAGCAGAG	GC	F	7103.69	7390.89	7350.89	56.3	48	23	200	106.3	21.4	
	RHCEex5RhE_e_F(CG)	cccCTGGCCAAGTGTCAACTCT	CG	R	6631.34	6878.54	6918.54	61.3	59	22	200	108.5	21.6	
c/C	RHCEEx2Rhc_C_F(AG)	ccTGGGCTTCCTCACCTC	GA	F	5362.5	5633.7	5649.7	57.4	67	18	200	71.9	14.4	
	RHCEEX2RHC_C_R-2	CCAGCTGTGTCTCCGGAAA	CT	R	5788.81	6115.91	6036.01	56.9	58	19	200	82.5	16.5	
Fy a/b	Fya_F(AG)	TCCCAGATGGAGACTATG	GA	F	5523.66	5794.86	5810.86	49.7	50	18	200	86.3	17.3	
	FYA_R-2	aCTGCTTCCAGGTTGGCA	CT	R	5490.62	5817.72	5737.82	56.6	56	18	200	66.8	13.4	
HPA5 a/b	GPIaHPA5_F(AG)_2	AGTCTACCTGTTTACTATCAAA	GA	F	6668.43	6939.63	6955.63	48.1	32	22	200	115.3	23.1	
	GPIAHPA5_R-2	ATTAGTTTATTTTTTTTTTTTTACCT	CT	R	7878.25	8205.35	8125.45	45	12	26	200	130	26	
Jk a/b	JKa_b_F(AG)	TCTTTCAGCCCCATTTGAG	GA	F	5729.78	6000.98	6016.98	52.4	47	19	200	93.5	18.7	
	JKA_B_R-2	aCAGAGTCCAAAGTAGATGT	CT	R	6174.11	6501.21	6421.31	50.1	40	20	200	90.5	18.1	
c/C-Intron	RHCintron_R(CINS)_3	gGGCTGCAATGAGCTATGATTGTAC	GA	F	7737.1	7984.3	8008.3	58.1	48	25	200	122.1	24.4	
	RHCintron_F_5	AGGGCCACCACCATTTGAAATCC	CT	R	6977.59	7224.79	7264.79	60.1	52	23	200	122.5	24.5	

Except for the antigens M/N, s/S, c/C, c/C-intron and e/E where manual designs were obtained, the table represents the automatic Sequenom Extension oligo output results obtained after the initial design phase using the AssayDesign software. This version of the PCR oligo design was used to cross-compare the genotyping results with the SNPstream panel. The extension oligos of "F and R" were mixed together with PCR grade water in two separate tubes to a final volume of 1.0 mL.

4.4. Detailed Genotyping procedure using the Beckman Coulter SNPstream technology

As illustrated in figure 9, the SNPstream platform is a medium-high-throughput technology multiplexing up to 12 or 48 SNPs and for up to 384 samples at the time. The technology is only capable of multiplexing, in a single tube, the same variant types at any one time (GA or CT or CA or GC or TA or GT) and it combines single base extension (SBE) and the quantification of specific fluorescent-labelled probes captured on 384-well format array plate. Each multiplexed PCR amplicon has been designed to amply the SNPs of interest (Figure 9, Step 1), unincorporated nucleotides and primers are removed enzymatically and a mixture of extension mixes and extension probes are added to the cleaned-up PCR reactions. The extension probe is hybridized to specific amplicons in the multiplex reaction, one base 3' of the SNP site (Figure 9, Step 2). The extension probes are single-base extended in a patented two-dye system by incorporation of a fluorescent-labelled chain terminating acyclonucleotide. The 5' end of the extended probes contain an extra unique 10 bases (Tag) and are specifically hybridized to one of the twelve complementary Anti-Tag sequences arrayed in each of the 384-wells (Figure 9, Step 3-A). The two-color detection allows determination of the genotype by comparing the signals from the two fluorescent dyes. The arrayed Anti-Tag captures the extended products and allows laser-based detection of each of the SNP allele signals (blue and green) (Figure 9, Step 3-B) and final SNPs calls are automatically and computationally generated (Figure 9, Step 3-C). Two extra self-extending control oligonucleotides are included in each extension mix and are extended during the probe extension. The Anti-Tags arrayed in the 384-wells include three positive controls (XX, XY and YY) and one negative control.

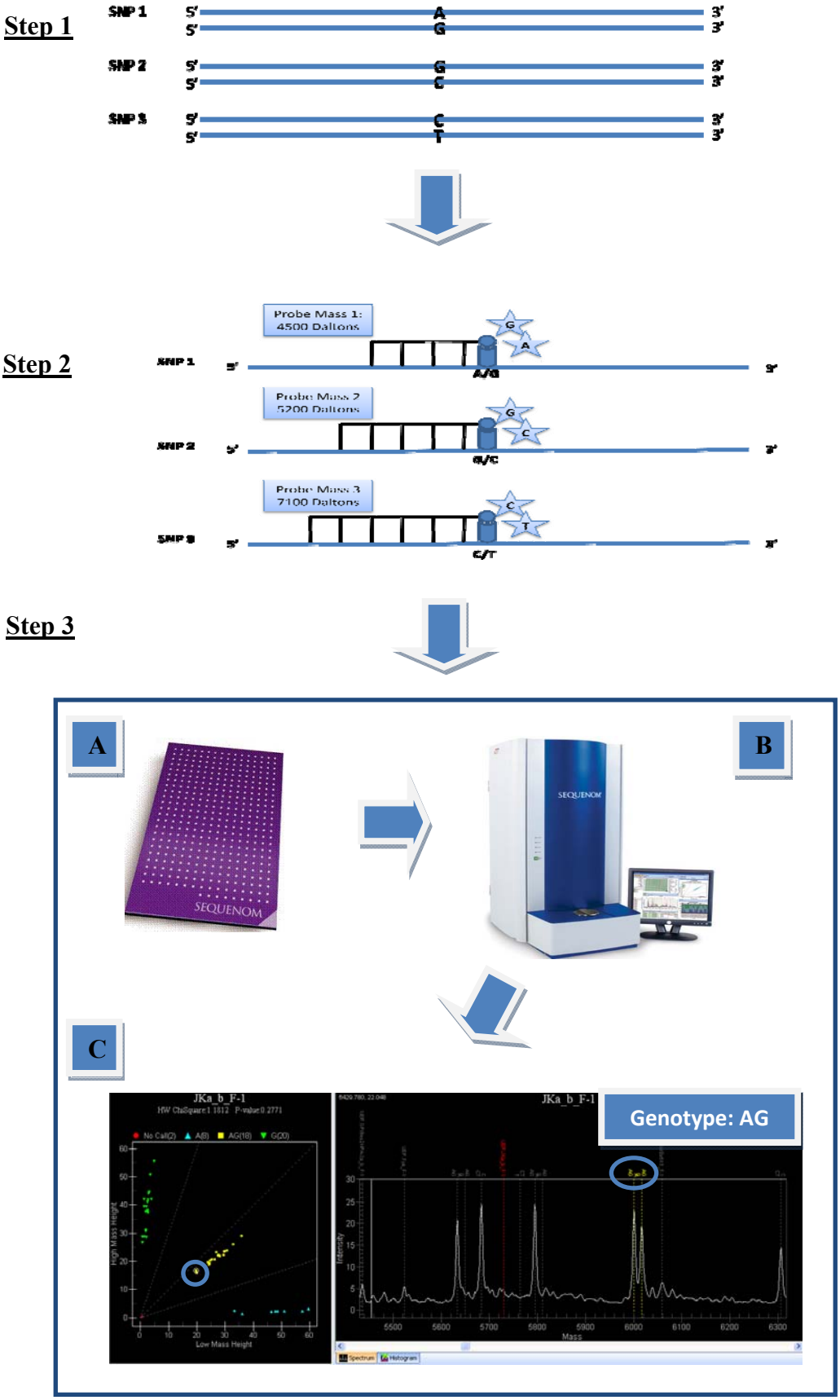
Figure 9: Detailed genotyping flow using Beckman Coulter SNPstream platform



4.5. Detailed Genotyping procedure using Sequenom MassArray system

As seen in figure 10, the MassArray platform is a medium-high-throughput technology multiplexing up to 40 SNPs on up to 384 samples at the time. The technology allows multiplexing in a single tube, and all variant types can be queried in a single reaction (GA, CT, CA, GC, TA, GT). This platform technology combines single base extension (SBE) and quantification of specific MassExtend probes detected and analysed by a time-of-flight mass spectrometer. Briefly, after multiplex PCR, the amplicons contain the SNPs of interest (Figure 10, Step 1), unincorporated nucleotides are removed enzymatically, a mixture of extension bases and extension probes are added to cleaned-up PCR reactions and the extension probes are hybridized to a specific template in the multiplex reaction one base 3' of the SNP site (Figure 10, Step 2). The extension probes are single-base extended by the incorporation of a specific mass terminating acyclonucleotides, resulting in an allele-specific difference in mass between extension products. The final extended mixture is desalted using a chelating based resin and then spotted onto a maldi-SpectroCHIP (Figure 10, Step 3-A). The SpectroCHIPS are placed into the Maldi-TOF mass spectrometer and the mass difference correlating to the genotype is determined in real time (Figure 10, Step 3-B). This mass difference allows computational data analysis by interpreting the difference in mass for each of test SNPs alleles present in the multiplex reaction (Figure 10, Step 3-C).

Figure 10: Detailed genotyping flow using Sequenom MassArray system



4.6. Sequencing

M/N antigen: Simplex PCR reactions were performed in 384-well plates (MJS BioLynx, Brockville, ON) in a 5 µL volume using 2 µL of ~12.5 ng/µL of selected Coriell Institute of Medical Research (Camden, NJ) DNAs, 140 µM dNTPs, 1.5 mM of $MgCl_2$, 0.3 U of Hotstart DNA polymerase (Qiagen), and 0.1 µL of PCR primers (forward oligo GYPA_PCRU2: 5' CACGCTTTATCTGTAAACCTCTGCTATG 3' and reverse oligo GYPA_PCRL1: 5' TGGCTGCATATGTGTCCCGTTTGT 3') at a concentration of 50 nM each in 1 X PCR buffer. Thermal cycling was performed in GeneAmp PCR system 9700 thermal cyclers (Applied Biosystems, Foster City, CA) using the following program: initial denaturation at 95°C for 15 min. followed by 45 cycles of 95°C for 30 sec., 60°C for 55 sec., 72°C for 45 sec. with a final extension of 72°C for 3 min. Following simplex PCR, 3 µL of a mixture containing 0.67 U Exonuclease I (USB Corporation, Cleveland, OH) and 0.33 U Shrimp Alkaline Phosphatase (USB Corporation, Cleveland, OH) was added to each well. The plates were sealed and incubated for 30 min. at 37°C and at 95°C for 10 min. To each well, containing cleaned-up PCR product, 1 µL of sequencing oligos 5 µM (GYPA_SEQU-L_1: 5' CTGTGAGATATTTCAGAGGCAAG 3') was added and mixed. The plate was then sent to the sequencing platform of the Génome Québec Innovation Centre (Montreal, QC) and sequencing was performed on a sequencing instrument 3730XL (Applied Biosystems, Foster City, CA) according to the manufacturers instructions. The sequencing results were then analyzed using PhredPhrap software (http://www.phrap.org/phredphrapconsed.html).

Whole GBAP content: Samples NA17175, NA17275, NA17223, NA17284 NA17294 and NA19007 from the Coriell Institute of Medical Research (Camden, NJ) were sequenced to validate the genotyping results obtained during the development phase. These samples will now be used as internal controls during the future clinical genotyping processes. The PCR and sequencing method is the same as the one described above for the *M/N antigen* section except for the PCR and sequencing oligos (table 6).

Table 6: PCR and sequencing oligos used to sequence the whole genotyping blood antigen panel contents.

Antigen	Oligo Type	Oligo IDs	Sequence (5'3')	Tm	GC%	Nb	Amplicon size (bp)	Concentration (µM)	Volume (µL)	Final Concentration (µM)	Final Volume (µL)
K/k	PCR_F	KelEx6K1_2-F-5	GRATCCTTATGCTCAGCCCC	57.6	60	20	163	100	5	0.05	1000
	PCR_R	KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	60.4	56	25		100	5	0.05	
	Sequencing	KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	60.4	56	25		100	5	0.05	1000
Kap a/b	PCR_F	KelExon8Kpa_bF-2	CTGACATTTCCTTCCTCCAGATC	54.5	47	23	180	100	5	0.05	1000
	PCR_R	KelExon8Kpa_bR	AGAGCTTGCCCTGTGCCC	60.3	67	18		100	5	0.05	
	Sequencing	KelExon8Kpa_bF-2	CTGACATTTCCTTCCTCCAGATC	60.3	67	23		100	5	0.05	1000
M/N	PCR_F	GYPA_PCRU2	CACGCTTTATCTGTAAACCTCTGCTATG	53	38	28	1100	100	5	0.05	1000
	PCR_R	GYPA_PCRL2	AACTTCATGAGCTCTAGGAGTGG	49.3	40	23		100	5	0.05	
	Sequencing	GYPA_SEQU-L_1	CTGTGAGATATTTCAGAGGCAAG	53	44	23		100	5	0.05	1000
s/S	PCR_F	GYPBEx4s_SF-s2	TTGTCAAATATTAACATACCTGGTACAGT	53	50	29	165	100	5	0.05	1000
	PCR_R	GYPBEx4s_SR-s2	AATATGATTAAGAAAAGGAAACCCG	50	38	25		100	5	0.05	
	Sequencing	GYPBEx4s_SR-s2	AATATGATTAAGAAAAGGAAACCCG	50	38	25		100	5	0.05	1000
HPA2 a/b	PCR_F	GPIBHPA-2a_bF-s2	CTCAGTCAAGTTGTTGTTAGCCAGAC	58	46	26	229	100	5	0.05	1000
	PCR_R	GPIBHPA-2a_bR-s2	CTCTACCTGAAAGGCAATGAGCT	56.6	48	23		100	5	0.05	
	Sequencing	GPIBHPA-2a_bF-s2-2	CAATCAGCTGCAAAGCCTGCC	60	57	21		100	5	0.05	1000
HPA1 a/b	PCR_F	GPIIIAex3HPA-1a_bF	ATTCTGGGGCACAGTTATCC	54.5	50	20	183	100	5	0.05	1000
	PCR_R	GPIIIAex3HPA-1a_bR-2	GCTCCAATGTACGGGGTAAACTC	54.3	46	23		100	5	0.05	
	Sequencing	GPIIIAex3HPA-1a_bR-2	GCTCCAATGTACGGGGTAAACTC	54.3	46	23		100	5	0.05	1000
e/E	PCR_F	RHCE_e-F-s2	TGGATGTTCTGGCCAAGTG	55.1	53	19	106	100	5	0.05	1000
	PCR_R	RHCE_e-R-s2	CTGTCACCACACTGACTGCTAG	57.1	55	22		100	5	0.05	
	Sequencing	RHCE_e-R-s2	CTGTCACCACACTGACTGCTAG	57.1	55	22		100	5	0.05	1000
c/C	PCR_F	RHCEc_CF-s3	TCCCCCTCCTCCTTCTCA	56.2	61	18	143	100	5	0.05	1000
	PCR_R	289-RHCEas-HD9135	GCCAGCATGAAGAGGTTGAA	55.4	50	20		100	5	0.05	
	Sequencing	RHCEc_CF-s3	TCCCCCTCCTCCTTCTCA	56.2	61	18		100	5	0.05	1000
Fy a/b	PCR_F	FYex2Fya_b-F-5	AGWCATCCAGCAGGTTACAGG	56.3	52	21	196	100	5	0.05	1000
	PCR_R	FYex2Fya_bR	AAGATGTATGGAATTCTTCCTATGG	52.3	36	25		100	5	0.05	
	Sequencing	FYex2Fya_bF-2	GTGTAACTCTGATGGCCTCCTCTG	58	54	24		100	5	0.05	1000
HPA5 a/b	PCR_F	GPIaHPA5_F	GCTCTTGGTAGGTGCACCAATGT	59.5	52	23	147	100	5	0.05	1000
	PCR_R	GPIaHPA5_R	TTCCAAATGCAAGTTAAATTACCAG	52.4	32	25		100	5	0.05	
	Sequencing	GPIaHPA5_F	GCTCTTGGTAGGTGCACCAATGT	59.5	52	23		100	5	0.05	1000
Jk a/b	PCR_F	JKa_b_F-s3_1	CCTGCTAACTTTCAATCCCACCCTC	59.6	52	25	253	100	5	0.05	1000
	PCR_R	JKa_b_R-s3_1	TGCCAGGTGAGCGCCATGAAC	62	62	21		100	5	0.05	
	Sequencing	JKa_b_F-s3_1	CCTGCTAACTTTCAATCCCACCCTC	59.6	52	25		100	5	0.05	1000

The table represents the PCR oligos and the sequencing oligos used to sequence six control DNAs from the Coriell Institute of Medical Research (Camden, NJ) (NA17175, NA17275, NA17223, NA17284, NA17294 and NA19007) to cross-validate the genotyping results obtained during the development phases. These validated samples will now be used subsequently as internal quality controls in all clinical genotyping experiments.

5. Clinical blood antigen genotyping assay procedure

Once the minor blood group antigen genotyping panel was optimised for robustness and accuracy levels needed for clinical applications, two SOPs (Standard Operating Procedure) and two LIMS (Laboratory Information Management System) protocols were developed to transfer the blood antigen genotyping panel from R&D into clinical application. All steps in the SOPs were mediated and tracked by a LIMS (Ocimum Bioscience, Hyderabad, India). Please refer to the Annexes A and B for the complete details on the SOPs PGx-PR012-V2.0 (Automated DNA extraction of blood spots from FTA cards using Generation DNA purification technology) and PGx-PR-013-V2.0 (Héma-Québec Genotyping Assay panel#1 on SNPstream Platform). Briefly these SOPs describe in detail all steps that encompass the entire process flow from the blood sampling on FTA micro cards to DNA extraction and genotyping on the SNPstream instrument, along the process of reporting the results back to Héma-Québec (Figure 11).

5.1. Blood sample collection at Héma-Québec

Frequent blood donors (≥ 3times per year) who had recently donated were selected through a new application from the Héma-Québec IT department. In total 22,000 frequent donors, having pre-determined interesting blood antigen combination profiles were requested to join the Héma-Québec blood antigen database registry. Peripheral blood samples were collected from these donors after signature of an informed consent form approved by Héma-Québec's Research Ethics Committee.

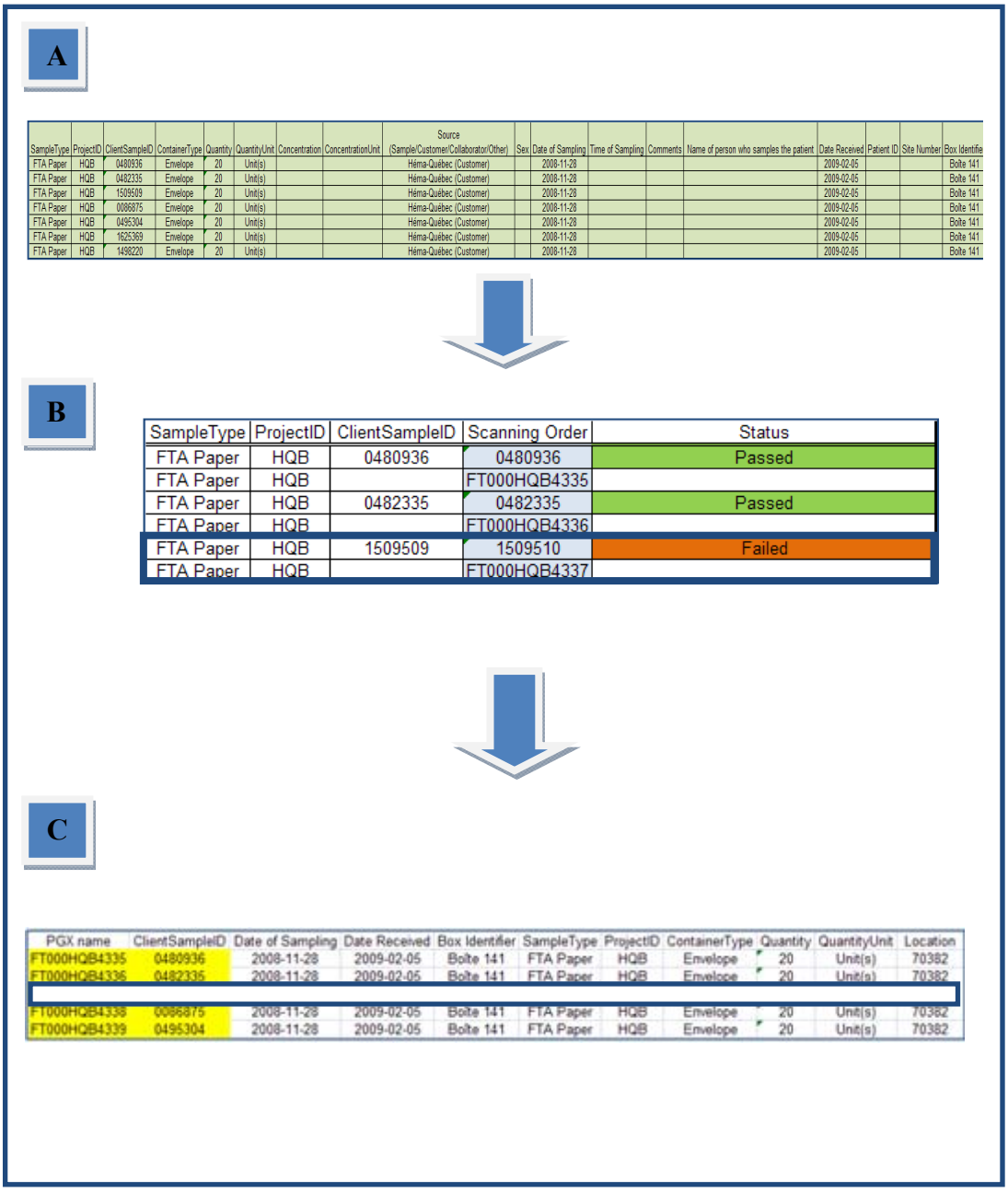
5.2. Sample management

Refer to the Annexe A sections 9.1 and 9.2 for the complete sample management procedure. Briefly, every month ~ 2,000 barcodes were pre-generated from our LIMS and sent to Héma-Québec. 120 µL of blood from selected donors were spotted onto FTA micro cards (GE Health Care/Whatman, Piscataway, NJ) and samples were stored at room temperature for a minimum of three hours in order to fully dry the cards (Figure 11-A). Both barcodes, the one generated by Héma-Québec and the one generated from our LIMS, were placed on every FTA card (Figure 11-B). The FTA

cards were then organized as a batch of 100 samples per storage box and 500 samples every two weeks were sent at room temperature to the Pharmacogenomics Centre. Upon arrival, the sample list provided by Héma-Québec was downloaded from our secure share website folder and pasted into the "sample checker" tool (Figure 12-A). To match the FTA cards samples against the list sent by Héma-Québec, the HQ barcode is scanned first and immediately a notification appears as either "Passed or Failed" based on the original submitted sample list. Then the LIMS barcode is scanned to make the final link between the HQ and LIMS barcodes (Figure 12-B). The final file containing all sample-related information is created only if the previous step was completed with 100% "Passed" comments. The file is then uploaded into the LIMS (Figure 12-C).

Figure 11: Process flow overview of the clinical blood antigen genotyping assay

Figure 12: Clinical FTA blood sample management flow



5.3. DNA plate management and DNA extraction

Refer to Annexe A sections 9.3 and 9.4 for complete details on the DNA management and extraction procedure. Briefly, 88 Héma-Québec FTA samples and 1 Blank FTA were retrieved and pooled together to create one 96-well DNA extraction plate (the remaining positions H6 to H12 were reserved for 1 water negative control and 6 known genotype positive controls DNAs). All remaining FTA samples were retrieved and pooled together following the same extraction plate layout.

Using a Duet 600 (BSD Robotics, Brisbane, Australia), 2mm punches were automatically punched out of each FTA card sample and deposited into a standard full-skirt 96-well plate (Bio-Rad, Hurcules, CA) containing 1 µL of PCR grade water (Figure 11-C). Prior to performing the punch, the Duet 600 robot scanned each sample barcode in order to match the DNA plate extraction layout previously created and only perfect matches allowed the system to move forward. The DNAs were extracted using exactly the same procedure described in section 3 above but were automated and adapted onto the BiomekFX liquid handler (Beckman Coulter, Fullerton, CA) to maximize the sample extraction throughput. A total of 6 plates of 89 samples were extracted in a single run (Figure 11-D). Once the DNAs were extracted the plates were stored at -20°C until needed. The DNA samples were single use and ready-to-use since no quantification is required.

5.4. PCR oligos and extension oligo mixes batches preparation

Please refer to Annexe B, appendices PGx_PR-012_APP-3 and PGx_PR-012_APP-5 for complete oligo mix preparation methodologies.

5.5. Multiplex PCR master mix batch preparation

Refer to Annexe B, appendices PGx_PR-012_APP-4 for details on the multiplex PCR master mix preparation.

5.6. Genotyping process

The Annexe B section 8.1 to 8.8, describes in detail the complete genotyping process of one assay plate for 192 different samples in both DNA directions, for a total of 384 total reactions (Figure 11-E). The same flow and SOPs are applied to genotype two or more assay plates at the same time.

The known DNA control preparation is detailed in the Annexe B appendix PGx_PR-012_APP-1.

The PCR, extension as well as the hybridization plate layouts are described in the Annexe B appendix PGx_PR-012_APP-2.

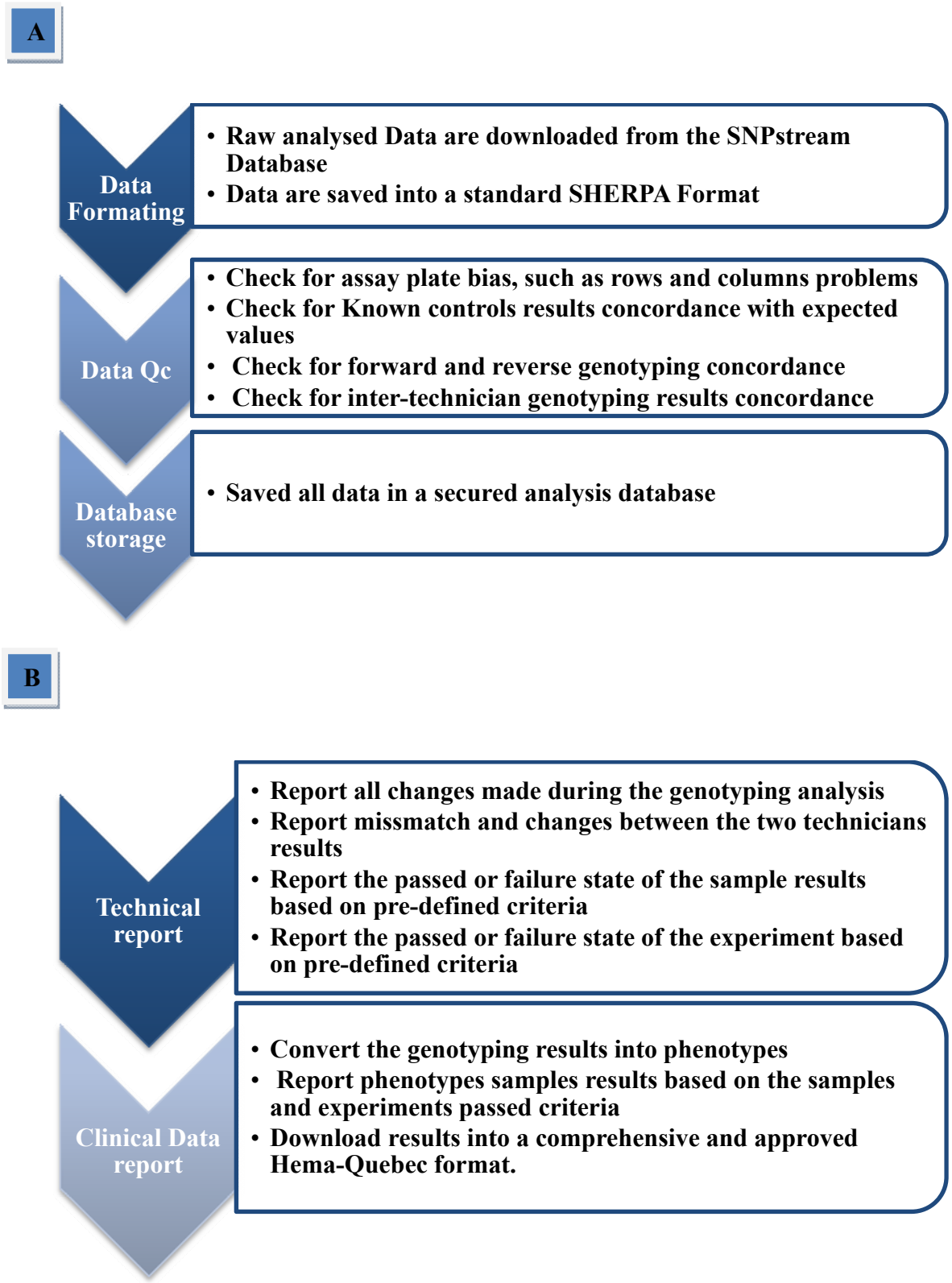
5.7. Data acquisition

Please refer to the Annexe B appendix PGx_PR-012_APP-7 for complete details on the Raw Data acquisition using the SNPstream platform.

5.8. Data Analysis and reporting

Please see Annexe B appendix PGx_PR-012_APP-8 for data analysis and reporting methodologies. Briefly, the data analysis and data reporting are supported by two different tools, SHERPA and Area51. Both software tools were developed at the Pharmacogenomics Centre, led by Michal Blazejczyk and Marc Bouffard respectively. Both applications have different roles but are complementary to each other (See Figures 13-A and B).

Figure 13: Data analysis and reporting flow



RESULTS

1. Assay development and optimisation

1.1. Bioinformatics analysis of the genes and markers of interest

A major factor affecting genotyping assay design efficiency, especially in the context of multiplex PCR, is lack of knowledge of the genomic complexities that surrounding the regions of interest. These complexities will ultimately alter the quality and accuracy of the PCR and extension oligo designs. Consequently, prior to primer design, each target gene and marker's genomic sequence was analysed bioinformatically for possible gene homology, repeat structures, and neighbouring polymorphisms. In addition, all homologous genes were aligned using ClustalW alignment software.

For example, the gene ***Kel*** encompassed two markers of interest, **Kap a/b (rs8176059) and k/K (rs8176058)**. The gene is located on chromosome 7 (7q34), and it is positioned from between 142348323 to 142369625 (genome sequence Build 35). Using the University of California, Santa Cruz (UCSC) genome browser view showed that the ***Kel*** gene was not located in a highly repetitive region (Figure 14-A), some localized regions, particularly within the introns were repeated. Since both antigen SNPs of interest were located in the exons 7 and 5 respectively the gene was initially considered not problematic for assay design. The gene was then subjected to a BLAT analysis and no homologous genes were identified (Figure 14-B). The BLAT analysis found more than 50 matches but all of them matched only a few hundred bases out of 23,000 total bases. Only one perfect match was found and it was associated with the gene of interest. According to the NCBI dbSNP build 126, more than 450 SNPs were found within the gene and the majority of them were mainly located in the introns (Figure 14-A). In order to narrow down the analysis of the markers of interest, approximately 200 bp upstream and downstream of the antigens

Kap a/b (Chr7:142361476) and k/K (Chr7:142364880) were submitted to BLAT and neighbouring SNP analysis. Following the BLAT analysis, no homology was found for homologous chromosome regions or additional genes, and only the regions of interest matched perfectly (Figures 14-C and D; highlighted lines). The genomic regions around both SNPs were also analyzed for repeat elements and neighbouring SNPs. No repeats elements were found in the surrounding sequences, whereas only 1 and 3 neighbouring SNPs were found for rs8176059 and rs8176058 respectively (Figures 14-E and F). The same exercise was performed for all test markers and four of them **(rs7682260-M/N, rs7683365-s/S, rs1053344-c/C and rs609320-e/E)** were identified as highly problematic (Table 7). The problems were specifically identified as regions of high gene homologies around the test marker DNA sequences. Following these findings, the gene *GYPA* was aligned with *GYPB* and *GYPE* and the gene *RHCE* were aligned with *RHD* (Figures 15-A to D). The highly homologous regions found for these four genes led to the need for manual PCR and extension oligo designs for the markers within these gene regions, whereas the designs of the other markers were produced automatically using the Autoprimer software (Beckman Coulter, Fullerton, CA) and the AssayDesign software (Sequenom, La Jolla, CA).

Figure 14. Bioinformatics analysis of the gene *Kel*

A) Graphical representation of the gene *Kel* with its respective exon and intron structure (1). The middle part of the picture represents the polymorphism content for the whole gene based on dbSNP build 126 (2). The blue rectangle highlights the repeat elements throughout the whole gene (3) and the arrows are pointing the respective gene location for each repeat section.

B) BLAT results of the entire *Kel* gene (23,300 bases) using the UCSC BLAT tool. The blue rectangle highlights the main homology results. The "Score" represents the total number of bases that match the genome, the "Start" and "End" represent the start and end base positions that match the genome, the "Qsize" is the total size submitted as query, the "Identity" represents the % of identity between the query and the match results, the "Chr" represents on which chromosome the query matches, and the "Span" represents how many consecutive bases are found within the genomic regions.

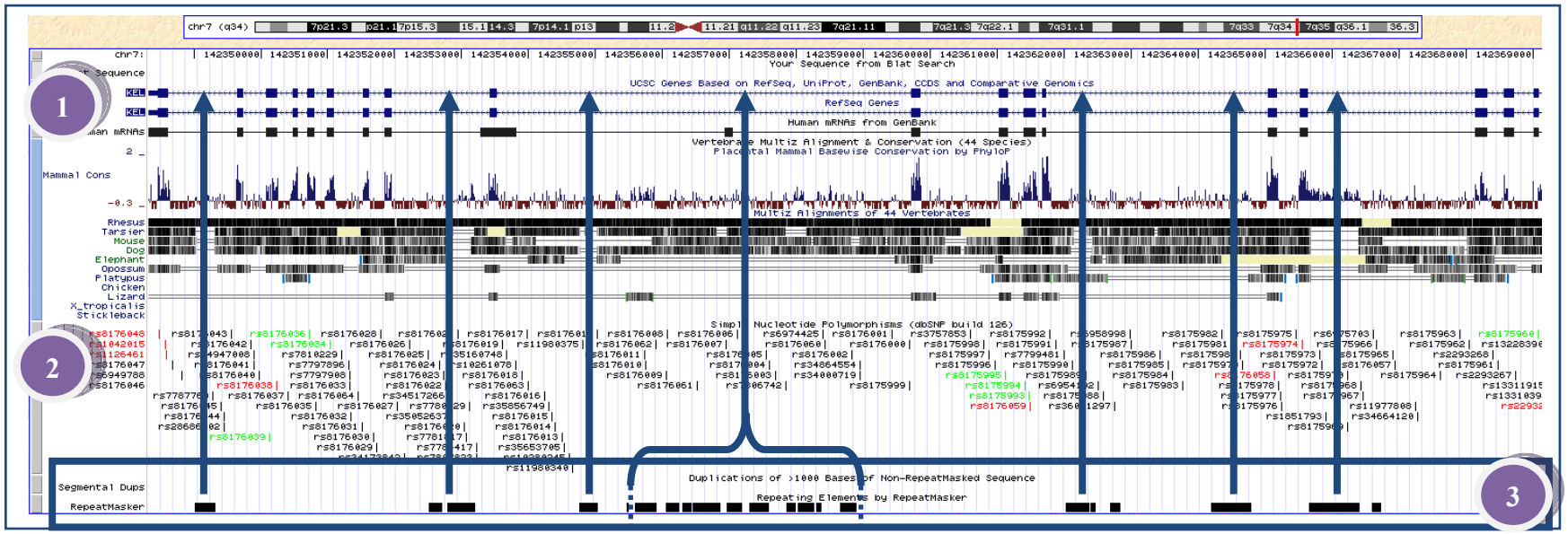
C) BLAT results of a 356 bp DNA sequence that surrounds the marker rs8176059 using the UCSC BLAT tool.

D) BLAT results of 356 bp DNA sequences that surround the marker rs8176058 using the UCSC BLAT tool.

E) Neighbouring SNP annotation of 200 bp downstream and upstream of the marker rs8176059. The base highlighted in yellow represents the marker rs8176059 (C>T) and the bases highlighted in pink represent the flanking SNPs.

F) Neighbouring SNP annotation of 200 bp downstream and upstream of the marker rs8176058. The base highlighted in yellow represents the marker rs8176058 (G>A) and the bases highlighted in pink represent the flanking SNPs.

A


1
2
3
chr7 (q34)
Your Sequence from Blat Search
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs from GenBank
Vertebrate Multiz Alignment & Conservation (44 Species)
Placental Mammal Basewise Conservation by PhyloP
Mammal Cons
Multiz Alignments of 44 Vertebrates
Rhesus
Tarsier
Mouse
Dog
Elephant
Opossum
Platypus
Chicken
Lizard
X_tropicalis
Stickleback
Simple Nucleotide Polymorphisms (dbSNP build 126)
Duplications of >1000 Bases Of Non-RepeatMasked Sequence
Segmental Dups
Repeating Elements by RepeatMasker
RepeatMasker

B

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	23302	1	23302	23302	100.0%	7	-	142347324	142370625	23302
browser details	YourSeq	585	4227	4970	23302	91.7%	X	-	153775048	153775795	748
browser details	YourSeq	575	4227	4966	23302	90.7%	14	+	53018477	53019219	743
browser details	YourSeq	568	4241	4968	23302	91.3%	X	+	72645176	72920047	274872
browser details	YourSeq	566	4226	4968	23302	91.3%	4	-	104515245	104515994	750
browser details	YourSeq	564	4227	4969	23302	90.7%	3	+	149432106	149432869	764
browser details	YourSeq	564	4231	4965	23302	91.5%	18	+	57715624	57716354	731
browser details	YourSeq	561	4227	4968	23302	90.9%	17	+	9251491	9252256	766
browser details	YourSeq	560	4025	4965	23302	88.4%	17	+	31256089	31256868	780
browser details	YourSeq	558	3946	4970	23302	90.1%	14	-	40403311	40404482	1172
browser details	YourSeq	555	4227	5023	23302	91.3%	X	+	135517162	135518360	1199
browser details	YourSeq	554	4227	4969	23302	91.6%	12	-	91882051	91882788	738
browser details	YourSeq	553	4227	4966	23302	91.6%	8	+	41049445	41058839	9395
browser details	YourSeq	550	4241	4965	23302	90.6%	5	-	114721680	114722413	734
browser details	YourSeq	545	4226	4968	23302	89.8%	13	-	45864191	45864940	750
browser details	YourSeq	544	4230	4968	23302	91.7%	14	-	20304868	20305609	742

C

Home Genomes Tables Gene Sorter PCR Session FAQ Help

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	356	1	356	356	100.0%	7	-	142361276	142361631	356
browser details	YourSeq	33	246	317	356	94.5%	12	-	114558015	114558302	288
browser details	YourSeq	32	7	40	356	100.0%	17	+	22641454	22641507	54
browser details	YourSeq	30	83	137	356	94.2%	1	-	17581134	17581263	130
browser details	YourSeq	26	281	312	356	90.7%	6	-	121953330	121953361	32
browser details	YourSeq	25	159	186	356	96.5%	12	-	50968976	50969016	41
browser details	YourSeq	20	70	89	356	100.0%	17	-	68542655	68542674	20
browser details	YourSeq	20	299	318	356	100.0%	15	-	58377843	58377862	20
browser details	YourSeq	20	296	315	356	100.0%	12	-	87125899	87125918	20
browser details	YourSeq	20	294	313	356	100.0%	10	-	113916389	113916408	20

D

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	356	1	356	356	100.0%	7	-	142364930	142365285	356
browser details	YourSeq	22	182	203	356	100.0%	17	-	43466110	43466131	22
browser details	YourSeq	22	224	247	356	87.0%	14	+	103133889	103133911	23
browser details	YourSeq	21	11	31	356	100.0%	5	+	36723732	36723752	21
browser details	YourSeq	20	240	259	356	100.0%	18	-	67527600	67527619	20
browser details	YourSeq	20	184	203	356	100.0%	14	-	91070918	91070937	20
browser details	YourSeq	20	160	179	356	100.0%	17	+	42699841	42699860	20

E

```
>chr7:142361276-142361676
TGGAGTCAGAGACAGAATGAAAAGGTATTAAGGGCACTAGGAGGAAGAAGGAAAGGAGGT
AATGTTTGAGAGGAAGATCCCCATGCCCACAGTCTTCTGGCCCCCAGTTCCAGGCACCTT
GAGCTGGTCgATAGTGACCATCTGGAAGAGCTTGCCCTGTGCCCGCCGCTGCTCCAGGGG
CCTCAGAAACTGGAAcAGCCgTGAAGTGATGGAGATTGACAAGGAAGAGTGTTCTTGCAC
CTTGCTTGGGTCTCCTCCCAGCAAGGTTCCCAGCTGATTCAGGTAAGTCAGGTATTCCCG
AAAGATCTGGAGGAAGGAAATGTCAGTCACAGGTGCCAGAGGTCCCTGCTTTTCTCCTGG
CCTCAGAGGGCAGGGCCTTTGTCCATGTGCCATCTTACCTG
```

F

```
>chr7:142364900-142365330
TTCCCTTAAGGTAGGGTTGTTTCCTATATCACACAGGTGTCCTCTCTTCCCAACCTGCAA
CCTTCCTCTAAGGGATGATTGGCTAGAGTTGGAGGCAGGCCAGGTCCAACTGTGTCTTCG
CCAGTGCATCCCTCACCTGGATGACTGGTGTGTGTGGAGAGGCAGGATGAGGTCCTAGGT
AGGCTCTGAAGAAAGGGAAATGGCCATACTGACTCATCAGAAGTCTCAGCgTTCGGTTAA
AGTTTAAGGAAGTCCATTTACCAGAGATGCGCCAGCCTCCAAGCTTTAAAGGAGAGAGAG
gGGGCTGAGCATAAGGATCcGTGGAGCCCATCCCCcATTGTCTGGATCGGCTCTTACACA
GCTGCTACTGCCTGACCTCTGCCCTGCGCCCTTCGATTCCTCTAGGAAGCCACATCTATC
CTTTTTTATCT
```

Figure 15. Bioinformatics analysis of the genes *GYPA, GYPB* and *RHCE.*

A) Alignment of 400 bp of *GYPA, GYPB* and *GYPE* genes using the ClustalW software tool. The bases in turquoise are the mismatched nucleotides and the base in green represents the marker **rs7682260-M/N.**

B) Alignment of 400 bp of *GYPA, GYPB* and *GYPE* genes using ClustalW. The bases in turquoise are the mismatched nucleotides and the base in green represents the marker **rs7683365-s/S.**

C) Alignment of 400 bp of *RHCE* and *RHD* genes using ClustalW. *RHCE-c* and *RHCE-C* represent the sequences of both *RHCE* small c and *RHCE* big C alleles. The bases in turquoise are the mismatched nucleotides and the nucleotide in green represents the marker **rs1053344-c/C.**

D) Alignment of 400 bp of *RHCE* and *RHD* genes using ClustalW. *RHCE-e* and *RHCE-E* represent the sequences of both RHCE small e and *RHCE* big E alleles. The bases in turquoise are the mismatched nucleotides and the nucleotide in green represents the marker **rs609320-e/E.**

A

```
GYPA       TGGAGGGATGTGAG-GGAATTTGTCTTTTGCAATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
GYPB       TGGAGGGATGTGAGAGGAATTTGTCTTTTGTGATATGCTTTATGGTCTGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
GYPE       TGGAGGGATGTGAG-GGAATTTGTCTTTTGTGATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
consensus  TGGAGGGATGTGAG-GGAATTTGTCTTTTGtgATATGCTTTATGGTCcGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
           28481....28490.....28500.....28510.....28520.....28530.....28540.....28550......

GYPA       ACTTCTATTTTATACAGAAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACTTCTTCTT
GYPB       ACTTCTATTTTATACAGAAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACCTCTTCTT
GYPE       ACTTCTATTTTATACAGGAATTGTGAGCATATCAGCATCAAGTACCACTGGTGTGGCAATGCACACTTCAACCTCTTCTT
consensus  ACTTCTATTTTATACAGaAATTGTGAGCATATCAGCATtAAGTACCACTGagGTGGCAATGCACACTTCAACcTCTTCTT
          28561....28570.....28580.....28590.....28600.....28610.....28620.....28630......

GYPA       CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCATT
GYPB       CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATGATT
GYPE       CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCGTT
consensus  CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATcaTT
           28641....28650.....28660.....28670.....28680.....28690.....28700.....28710......
```

B

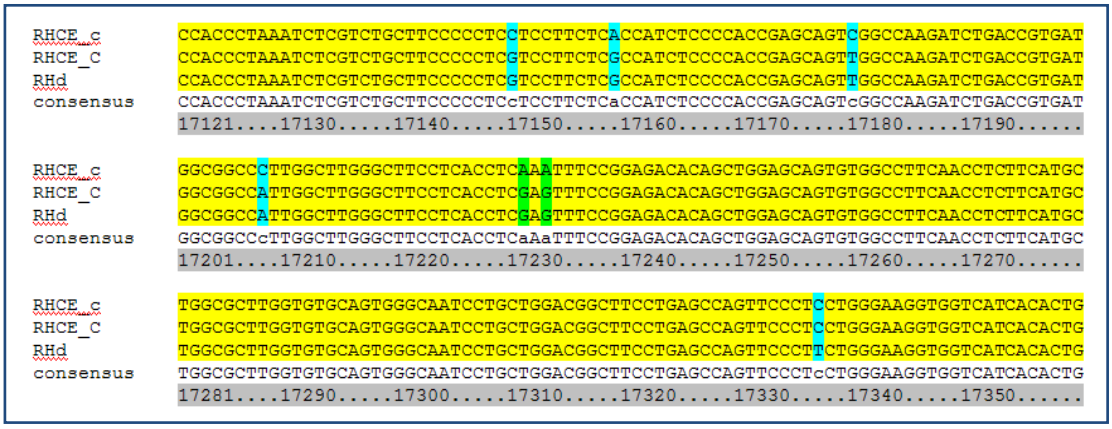
```
GYPA       TGGTACGACTGACATATTACCTCATAAATGTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTAA
GYPB       TGGTAAGACTGACACATTACCTCATAAATGTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTAA
GYPE       TGGTAGGACTGACACATTCCCTCATAAATTTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTAA
consensus  TGGTA-GACTGACAcATTaCCTCATAAATgTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTAA
           30241....30250.....30260.....30270.....30280.....30290.....30300.....30310......

GYPA       GAAAAGGAAACCCACAGAACAGTTTGATTCCAAAATGATTTTTTTCTTTGCACATGCCTTACTTATTTGGACTTACATTG
GYPB       GAAAAGGAAACCCGCAGAACAGTTTGATTCCAAAATGATTTTTTTCTTTGCACATGTCTTTCTTATTTGGACTTACATTG
GYPE       GAAAAGGAAACCCACAGAACAGTTCGATTCCAAAATGATTTTTTTCTTTGCAC--GTCTTACTTATTTGGACTTACATTG
consensus  GAAAAGGAAACCCaCAGAACAGTTtGATTCCAAAATGATTTTTTTCTTTGCACatGtCTTaCTTATTTGGACTTACATTG
           30321....30330.....30340.....30350.....30360.....30370.....30380.....30390......

GYPA       AAATTTTGCTTTATAGGAGAAAGGGTACAACTTGCCCATCATTTCTCTGAACCAGGTATGTTAATATTTGACAAAGAATA
GYPB       AAATTTTGCTTTATAGGAGAAACGGGACAACTTGTCCATCGTTTCACTGTACCAGGTATGTTAATATTTGACAAAGAATA
GYPE       AAATTTCGCTTTATAGGAGAAAGGGGACAACTTGTCCATCGTTTCCCTGAAGCAGATATGTTAATATTTGACAAAGAATC
consensus  AAATTTtGCTTTATAGGAGAAAgGGgACAACTTGtCCATCgTTTC-CTGaAcCAGgTATGTTAATATTTGACAAAGAATa
           30401....30410.....30420.....30430.....30440.....30450.....30460.....30470......

GYPA       AAAGTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCTATCTT-----AGAAGAGCAT
GYPB       AAAGTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCTATCTTAAAACAGAAGAGCAT
GYPE       AAAATCATTCCATTTTAAACTATCCCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCAATCTTAAAACAGAAGAGCAT
consensus  AAAgTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCtATCTTaaaacAGAAGAGCAT
           30481....30490.....30500.....30510.....30520.....30530.....30540.....30550......
```

C


RHCE_c CCACCCTAAATCTCGTCTGCTTCCCCCTCCTCCTTCTCACCATCTCCCCACCGAGCAGTCGGCCAAGATCTGACCGTGAT
RHCE_C CCACCCTAAATCTCGTCTGCTTCCCCCTCGTCCTTCTCGCCATCTCCCCACCGAGCAGTTGGCCAAGATCTGACCGTGAT
RHd CCACCCTAAATCTCGTCTGCTTCCCCCTCGTCCTTCTCGCCATCTCCCCACCGAGCAGTTGGCCAAGATCTGACCGTGAT
consensus CCACCCTAAATCTCGTCTGCTTCCCCCTCcTCCTTCTCaCCATCTCCCCACCGAGCAGTcGGCCAAGATCTGACCGTGAT
17121....17130.....17140.....17150.....17160.....17170.....17180.....17190.....
RHCE_c GGCGGCCCTTGGCTTGGGCTTCCTCACCTCAAATTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
RHCE_C GGCGGCCATTGGCTTGGGCTTCCTCACCTCGAGTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
RHd GGCGGCCATTGGCTTGGGCTTCCTCACCTCGAGTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
consensus GGCGGCCcTTGGCTTGGGCTTCCTCACCTCaAaTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
17201....17210.....17220.....17230.....17240.....17250.....17260.....17270.....
RHCE_c TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTCCTGGGAAGGTGGTCATCACACTG
RHCE_C TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTCCTGGGAAGGTGGTCATCACACTG
RHd TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTTCTGGGAAGGTGGTCATCACACTG
consensus TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTcCTGGGAAGGTGGTCATCACACTG
17281....17290.....17300.....17310.....17320.....17330.....17340.....17350.....

D

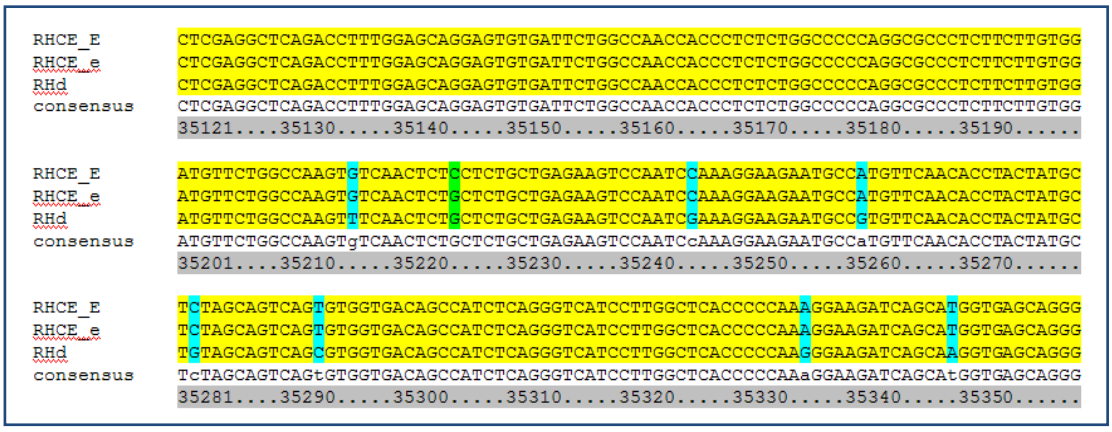

RHCE_E CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
RHCE_e CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
RHd CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
consensus CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
35121....35130.....35140.....35150.....35160.....35170.....35180.....35190.....
RHCE_E ATGTTCTGGCCAAGTGTCAACTCTGCTCTGCTGAGAAGTCCAATCCAAAGGAAGAATGCCATGTTCAACACCTACTATGC
RHCE_e ATGTTCTGGCCAAGTGTCAACTCTGCTCTGCTGAGAAGTCCAATCCAAAGGAAGAATGCCATGTTCAACACCTACTATGC
RHd ATGTTCTGGCCAAGTTTCAACTCTGCTCTGCTGAGAAGTCCAATCGAAAGGAAGAATGCCGTGTTCAACACCTACTATGC
consensus ATGTTCTGGCCAAGTgTCAACTCTGCTCTGCTGAGAAGTCCAATCcAAAGGAAGAATGCCaTGTTCAACACCTACTATGC
35201....35210.....35220.....35230.....35240.....35250.....35260.....35270.....
RHCE_E TCTAGCAGTCAGTGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAAGGAAGATCAGCATGGTGAGCAGGG
RHCE_e TCTAGCAGTCAGTGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAAGGAAGATCAGCATGGTGAGCAGGG
RHd TGTAGCAGTCAGCGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAGGGAAGATCAGCAAGGTGAGCAGGG
consensus TcTAGCAGTCAGtGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAaGGAAGATCAGCAtGGTGAGCAGGG
35281....35290.....35300.....35310.....35320.....35330.....35340.....35350.....

Table 7: Qualitative representation of the DNA sequence complexity at the gene and marker levels

				At the Gene level							At the marker region level		
Gene	Chromosome Pos	Gene Start	Gene End	Repeat element	Homology	Polymorphic level	Antigens	Marker ID	Genomic variation	Amino Acid Change	Repeat element	Homology	Polymorphic level
Kel	Chr 7	142348323	142369625	Repeated in Introns	NO	Highly	Kap a/b	rs8176059	CGG → TGG	Arg → Trp	NO	NO	Low
ITGbIII	Chr17	42686207	42745076	Repeated in Introns	NO	Highly	HPA1 a/b	rs5918	CTG → CCG	Leu → Pro	NO	NO	Low
GYPA	chr4	145249906	145281354	Low	Yes (GYPB, GYPE)	Highly	M/N	rs7682260	TAA → TCA	Leu → Ser	NO	Yes (GYPE,GYPB)	Low
GYPB	chr4	145136707	145159946	Low	Yes (GYPA,GYPE)	Highly	s/S	rs7683365	ACG → ATG	Thr → Met	NO	Yes (GYPE,GYPA)	Low
RHCE	chr1	25561327	25619950	Highly repeated	Yes (RHD)	Highly	c/C	rs1053344	AAT → AGT	Asn → Ser	NO	Yes (RHD)	Low
GP1BA	chr17	4776372	4779067	Low	NO	Low	HPA2 a/b	rs6065	ACG → ATG	Thr → Met	NO	NO	Low
DARC	chr1	157441134	157442914	Low	NO	Low	Fy a/b	rs12075	GGT → GAT	Gly → Asp	NO	NO	Low
SLC14A1	chr18	41560916	41586482	Repeated in Introns	NO	Highly	Jk a/b	rs1058396	GAC → GGC	Asp → Gly	NO	NO	Low
Kel	Chr 7	142348323	142369625	Repeated in Introns	NO	Highly	k/K	rs8176058	ACG → ATG	Thr → Met	NO	NO	Low
RHCE	chr1	25561327	25619950	Highly repeated	Yes (RHD)	Highly	e/E	rs609320	GCT → CCT	Ala → Pro	NO	Yes (RHD)	Low
ITGA2	chr5	52320913	52426366	Repeated in Introns	NO	Highly	HPA5 a/b	rs10471371	GAG → AAG	Glu → Lys	NO	NO	Low

The table represents a summary of qualitative result of complexity of the test genes and markers DNA sequences. These results were obtained after performing the same process flow described for the gene *Kel* and its related markers. The evaluation of the repeat elements and the polymorphic sites was achieved using the UCSC genome browser view. The homology was evaluated using the BLAT tool from UCSC and confirmed by the alignment of DNA sequences using ClustalW.

1.2. SNPstream and Sequenom minor blood group antigen genotyping panel assay design

After the completion of the bioinformatics analysis, 200 bp sequences upstream and downstream of the non-problematic markers (**Kap a/b-rs8176059, HPA1 a/b-rs5918, HPA2 a/b-rs6065, Fy a/b-rs12075, Jk a/b-rs1058396, k/K-rs8176058 and HPA5 a/b-rs10471371)** were submitted for an automatic "Standard" assay design to the Autoprimer and the AssayDesign software for the SNPstream and Sequenom platforms respectively. In general, the software programs more than adequately designed the genotyping assays using the same pipeline process, where for each marker a pair of PCR oligos that encompassed the target markers and an extension oligo ending one base (5') next to the test marker were generated. In order to reduce the amplification competition within the multiplex PCR reaction, all output PCR oligos were designed to generate PCR amplicons ranging from 80 bp to 200 bp. The software also generated oligos with minimal self-priming and dimer-dimer interactions within the multiplex pools (PCR as well as the extension pools). Since both software programs designed only one "set of designs" per marker (two PCR oligos and one Extension oligo), the submitted DNA sequences were subsequently modified to force the output results to also design assays for the opposite DNA strand in the reverse direction. The automatic output oligo designs are represented in Tables 2 to 5 (Version 1 column) in the Materials and Methods.

The problematic markers (**s/S-rs7683365 and M/N- rs7682260)** were manually designed from the aligned sequences using the standard procedure for both platforms. The markers (**c/C-rs1053344 and e/E-rs609320)** were subjected to a non-standard oligo design using the aligned sequences. For both platforms the manual PCR oligo design for **s/S-rs7683365** and **M/N- rs7682260** were meticulously performed to get the highest specificity in the 3' end regions whilst maintaining a GC% content between 40 and 60%, and a Tm of 50 to 60°C (Figure 16-A and B).

On the SNPstream, the assay designs for the markers **c/C-rs1053344 and e/E-rs609320** were developed using the "nonstandard" design approach since on this

platform only the same allele type (C/T or G/A or C/A or T/G or T/A) can be analysed as a single multiplex extension reaction. So one of the obstacles in the "Standard" creation of this panel was the e/E antigen pair, which is encoded by a G/C variation, whereas all other antigen pairs (Jk a/b, Fy a/b etc.) were G/A or C/T markers. To address this problem and since the genotyping blood antigen panel was expected to be developed in both DNA directions (G/A and C/T), we decided to interrogate one allele at the time, where the G nucleotide was detected in the C/T genotyping panel and the C nucleotide was detected in the G/A panel (Figure 17-A). By combining both genotyping results, the final genotyping calls (GG, GC or CC) can be inferred (Figure 17-B and C).

Another problem encountered during the creation of the "Standard" SNPstream minor blood group antigen genotyping panel design was the antigen c/C, where very high homology was found between the genes *RHD* and *RHCE,* and especially near the **c/C-rs1053344** region (Figure 15-C)**.** In addition, the alleles *RHCE_C* and the gene *RHD* have close to 100% homology to each other but are completely different from the *RHCE_c* allele. The high homology of the *RHCE_C* and *RHD* alleles considerably increased the complexity of these oligo designs and the requirement for highly specific PCR. As previously reported by (Poulter et *al.* and Tax *et al.*), the insertion of 109 bp into intron 2 of the *RHCE* gene is strongly associated with the presence of the *RHCE_C* allele. Consequently, we decided to interrogate the allele ***RHCE_C*** and ***RHCE_c*** independently in two different PCR and extension reactions (c/C for *RHCE_c* and c/C-intron for *RHCE_C*). The assay c/C discriminated for the presence of *RHCE_c* antigen over *RHCE_C/C* individuals where the PCR oligos were designed to uniquely amplify the *RHCE_c* allele (Figure 18-A), and the assay for the c/C-Intron looked for the presence of the insertion of 109 bp which is associated with the presence of *RHCE_C* (Figures 18-B and C). The results obtained using the assay for the c/C-intron facilitated the isolation of the C/C carriers from the c/C carriers, and when both c/C and c/C-intron assays are combined the heterozygote individuals and c/c carriers are easily identified (Figures 18-D and E).

For the Sequenom assay design, the platform can support the detection of all allele types in a single assay reaction. Only the antigen **c/C-rs1053344** assay required a "Non-Standard" design. As this assay was considerably complex in the context of multiplex genotyping, the same assay developed for the SNPstream was transferred to the Sequenom panel. The PCR and extension oligos of the antigen **e/E-rs609320** were then manually developed similarly to the technique describe in Figure 16 with respect to the specificities of the *RHCE* gene.

After completing the manual designs for the standard and non-standard SNPs, the PCR oligo sequences, except for the c/C-Intron design, were submitted to the *inSilico* PCR tool (UCSC) to ensure the specificity and uniqueness of the amplification. Also, since the manually designed oligos were not evaluated for dimer-dimer and self-priming potentials, the manually designed oligos were combined with the ones obtained automatically from both Autoprimer and AssayDesign and were submitted to FastPCR software for analysis. No conflicting oligo interactions were identified from these two quick analyses where all amplicons were compatible with the UCSC *inSilico* tool, and no major interaction was found within the PCR and extension oligo pools. The measured self-priming and dimer-dimer Tm ranged from 10 to 35 °C with minimal interactions from the 3' end of the primers.

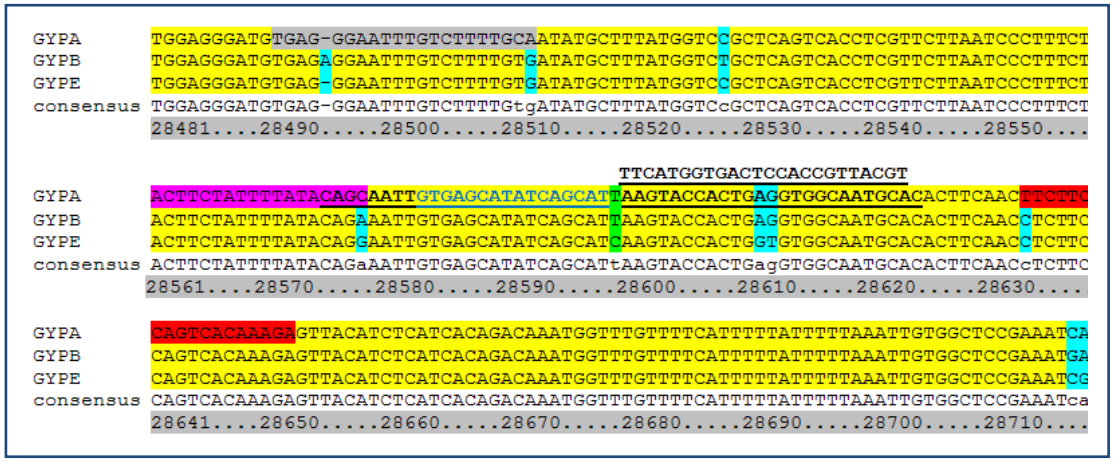
Following the assay designs and quick quality control tests, the PCR oligos and extension oligos shown in Tables 2 and 4 (version 1 column) were the oligos used in the first step of the SNPstream panel optimisation. The oligos designed for Sequenom were not optimized, since this assay is used essentially to validate the SNPstream panel. All further results will focus on the optimisation and validation of the SNPstream assay.

Figure 16: Standard manual PCR and extension oligo designs for SNPstream and Sequenom platforms.

A) Details of the manual design of M/N antigens marker (nucleotide highlighted in green) using the aligned DNA sequences. The turquoise nucleotides represent the variation between homologous genes. The pink highlighted sequence represents the forward PCR oligo design for the SNPstream platform and the red highlighted sequence represents the reverse PCR oligo for both SNPstream and Sequenom platforms. The forward oligo for Sequenom is represented in gray. The extension oligos for both the forward and reverse allele detection are represented in bold-underlined format. The gray letters are specific to the Sequenom platform since shorter sequences are needed to fit into the mass spectrum window. Both designs interrogate a standard C/T or G/A (in the reverse direction) marker.

B) Details of the manual design of the s/S antigen marker (nucleotide highlighted in green). The turquoise nucleotides represent the variation between homologous genes. The sequences highlighted in pink represent the forward and reverse oligo designs for the SNPstream platform, whereas the sequences in dark orange and bold-italic-underline purple represent the forward and reverse oligos respectively for Sequenom. The sequences in red represent the forward and reverse extension oligos for SNPstream and the bold-italic segment represent the forward extension oligo design. Both designs interrogate either a C/T or G/A marker (in the forward or reverse direction).

A


GYPA TGGAGGGATGTGAG-GGAATTTGTCTTTTGCAATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCT
GYPB TGGAGGGATGTGAGAGGAATTTGTCTTTTGTGATATGCTTTATGGTCTGCTCAGTCACCTCGTTCTTAATCCCTTTCT
GYPE TGGAGGGATGTGAG-GGAATTTGTCTTTTGTGATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCT
consensus TGGAGGGATGTGAG-GGAATTTGTCTTTTGtgATATGCTTTATGGTCcGCTCAGTCACCTCGTTCTTAATCCCTTTCT
28481....28490.....28500.....28510.....28520.....28530.....28540.....28550....
TTCATGGTGACTCCACCGTTACGT
GYPA ACTTCTATTTTATACAGCAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACTTCTTC
GYPB ACTTCTATTTTATACAGAAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACCTCTTC
GYPE ACTTCTATTTTATACAGGAATTGTGAGCATATCAGCATCAAGTACCACTGGTGTGGCAATGCACACTTCAACCTCTTC
consensus ACTTCTATTTTATACAGaAATTGTGAGCATATCAGCATtAAGTACCACTGagGTGGCAATGCACACTTCAACcTCTTC
28561....28570.....28580.....28590.....28600.....28610.....28620.....28630....
GYPA CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCA
GYPB CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATGA
GYPE CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCG
consensus CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATca
28641....28650.....28660.....28670.....28680.....28690.....28700.....28710....

B

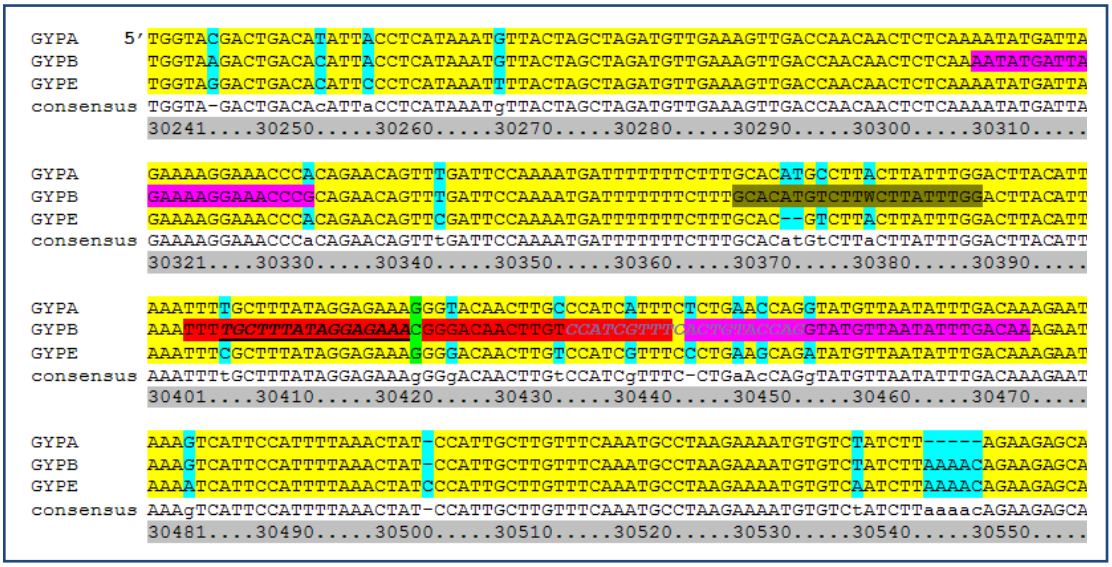

GYPA 5′TGGTACGACTGACATATTACCTCATAAATGTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTA
GYPB TGGTAAGACTGACACATTACCTCATAAATGTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTA
GYPE TGGTAGGACTGACACATTCCCTCATAAATTTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTA
consensus TGGTA-GACTGACAcATTaCCTCATAAATgTTACTAGCTAGATGTTGAAAGTTGACCAACAACTCTCAAAATATGATTA
30241....30250.....30260.....30270.....30280.....30290.....30300.....30310.....
GYPA GAAAAGGAAACCCACAGAACAGTTTGATTCCAAAATGATTTTTTTCTTTGCACATGCCTTACTTATTTGGACTTACATT
GYPB GAAAAGGAAACCCGCAGAACAGTTTGATTCCAAAATGATTTTTTTCTTTGCACATGTCTTWCTTATTTGGACTTACATT
GYPE GAAAAGGAAACCCACAGAACAGTTCGATTCCAAAATGATTTTTTTCTTTGCAC--GTCTTACTTATTTGGACTTACATT
consensus GAAAAGGAAACCCaCAGAACAGTTtGATTCCAAAATGATTTTTTTCTTTGCACatGtCTTaCTTATTTGGACTTACATT
30321....30330.....30340.....30350.....30360.....30370.....30380.....30390.....
GYPA AAATTTTGCTTTATAGGAGAAAGGGTACAACTTGCCCATCATTTCTCTGAACCAGGTATGTTAATATTTGACAAAGAAT
GYPB AAATTTTGCTTTATAGGAGAAACGGGACAACTTGTCCATCGTTTCCCTGAAGCAGATATGTTAATATTTGACAAAGAAT
GYPE AAATTTCGCTTTATAGGAGAAAGGGGACAACTTGTCCATCGTTTCCCTGAAGCAGATATGTTAATATTTGACAAAGAAT
consensus AAATTTtGCTTTATAGGAGAAAgGGgACAACTTGtCCATCgTTTC-CTGaAcCAGgTATGTTAATATTTGACAAAGAAT
30401....30410.....30420.....30430.....30440.....30450.....30460.....30470.....
GYPA AAAGTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCTATCTT-----AGAAGAGCA
GYPB AAAGTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCTATCTTAAAACAGAAGAGCA
GYPE AAAATCATTCCATTTTAAACTATCCCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCAATCTTAAAACAGAAGAGCA
consensus AAAgTCATTCCATTTTAAACTAT-CCATTGCTTGTTTCAAATGCCTAAGAAAATGTGTCtATCTTaaaacAGAAGAGCA
30481....30490.....30500.....30510.....30520.....30530.....30540.....30550.....

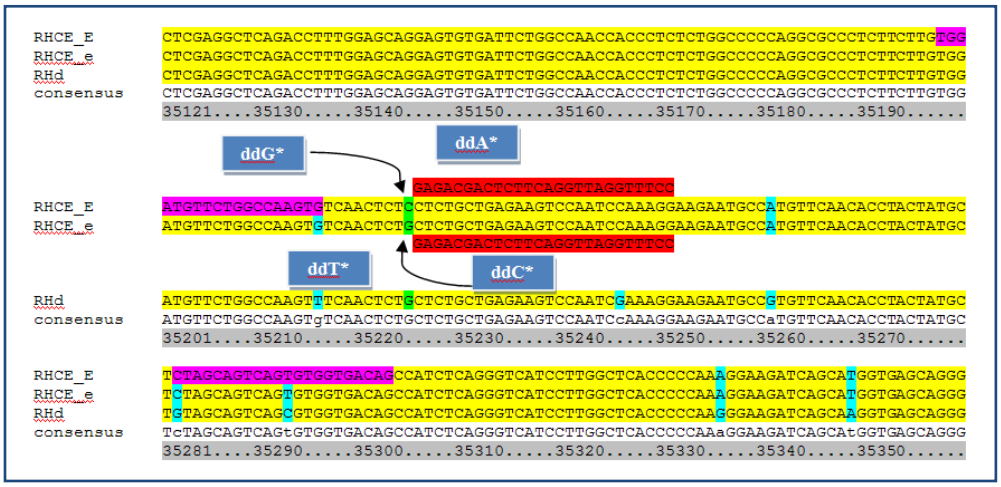
Figure 17: SNPstream specific non-standard PCR and extension oligo design for the antigen e/E.

A) Detailed oligo design strategy for the non-standard polymorphism **e/E-rs609320** (nucleotide in green). The sequences highlighted in pink represent the manual PCR oligo designs that are specific to *RHCE*. The sequences highlighted in red represent the extension oligos used in the extension pools GA and CT. The same extension oligo is used in both the **G**A and **C**T single-base extension reactions, but only the dd**G**-BODIPY-Fluorescein and the dd**C**-TAMRA are respectively incorporated, leading to a non-standard analysis of a G/C variation.

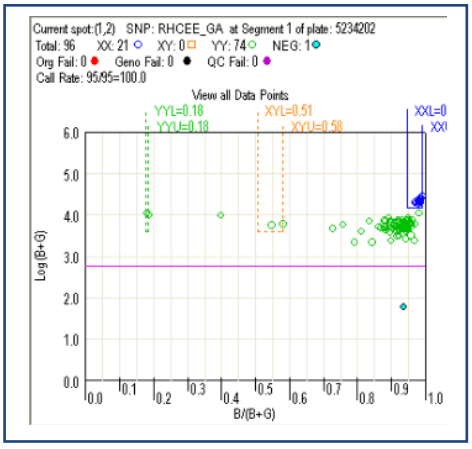
B) Graphical representation of the blue and green fluorescence in logarithmic mode (Polar View). The genotyping cluster illustrates the ddG* detection (GA extension pool) of the non-standard design of the polymorphism **e/E-rs609320.** The cluster in blue represents the positive ddG* extension, therefore it reflects the presence of G allele. Individuals located in this cluster are either G/G (E/E) or G/C (E/e), whereas the individuals in the green cluster are all C/C (e/e).

C) Genotyping cluster that illustrates the ddC* detection (CT extension pool) of the nonstandard design of the polymorphism **e/E-rs609320.** The cluster in green represents the positive ddC* extension, therefore it reflects the presence of the C allele. Individuals located in this cluster are either G/C (E/e) or C/C (e/e), whereas the individuals in the blue cluster are all G/G (E/E).

A


RHCE_E CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGGG
RHCE_e CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
RHd CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
consensus CTCGAGGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCCTCTCTGGCCCCCAGGCGCCCTCTTCTTGTGG
35121....35130.....35140.....35150.....35160.....35170.....35180.....35190.....
ddG*
ddA*
GAGACGACTCTTCAGGTTAGGTTTCC
RHCE_E ATGTTCTGGCCAAGTGTCAACTCTGCTCTGCTGAGAAGTCCAATCCAAAGGAAGAATGCCATGTTCAACACCTACTATGC
RHCE_e ATGTTCTGGCCAAGTGTCAACTCTGCTCTGCTGAGAAGTCCAATCCAAAGGAAGAATGCCATGTTCAACACCTACTATGC
GAGACGACTCTTCAGGTTAGGTTTCC
ddT*
ddC*
RHd ATGTTCTGGCCAAGTTTCAACTCTGCTCTGCTGAGAAGTCCAATCGAAAGGAAGAATGCCGTGTTCAACACCTACTATGC
consensus ATGTTCTGGCCAAGTgTCAACTCTGCTCTGCTGAGAAGTCCAATCcAAAGGAAGAATGCCaTGTTCAACACCTACTATGC
35201....35210.....35220.....35230.....35240.....35250.....35260.....35270.....
RHCE_E TCTAGCAGTCAGTGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAAGGAAGATCAGCATGGTGAGCAGGG
RHCE_e TCTAGCAGTCAGTGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAAGGAAGATCAGCATGGTGAGCAGGG
RHd TGTAGCAGTCAGCGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAGGGAAGATCAGCAAGGTGAGCAGGG
consensus TcTAGCAGTCAGtGTGGTGACAGCCATCTCAGGGTCATCCTTGGCTCACCCCCAAaGGAAGATCAGCAtGGTGAGCAGGG
35281....35290.....35300.....35310.....35320.....35330.....35340.....35350.....

B


Current spot:(1,2) SNP: RHCEE_GA at Segment 1 of plate: 5234202
Total: 96 XX: 21 XY: 0 YY: 74 NEG: 1
Org Fail: 0 Geno Fail: 0 QC Fail: 0
Call Rate: 95/95=100.0
View all Data Points
YYL=0.18
YYU=0.18
XYL=0.51
XYU=0.58
XXL=0
Log (B+G)
B/(B+G)

C

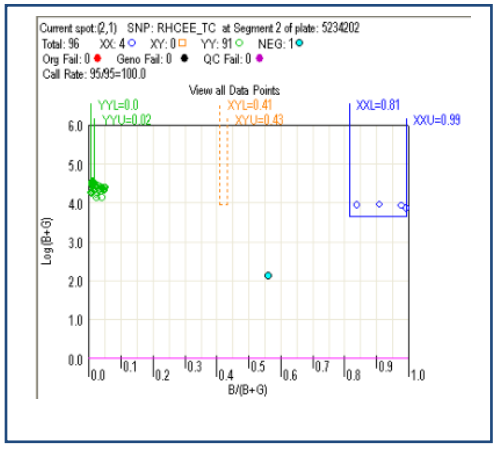

Current spot:(2,1) SNP: RHCEE_TC at Segment 2 of plate: 5234202
Total: 96 XX: 4 XY: 0 YY: 91 NEG: 1
Org Fail: 0 Geno Fail: 0 QC Fail: 0
Call Rate: 95/95=100.0
View all Data Points
YYL=0.0
YYU=0.02
XYL=0.41
XYU=0.43
XXL=0.81
XXU=0.99
Log (B+G)
B/(B+G)

Figure 18: Non-standard PCR and extension oligo designs for the antigens c/C.

A) Detailed design of the nonstandard polymorphism **c/C-rs1053344** (nucleotide in green). The sequences highlighted in pink represent the manual PCR oligo designs that are specific for the *RHCE_c* sequence allowing for the precise amplification of the specific amplicon containing the *RHCE_c* antigen. The sequence highlighted in red represents the extension oligos used in the extension pool G**A**. In the reaction only the dd**A**-TAMRA is incorporated, leading to a non-standard analysis of A/- detection. The same design logic is applied for the reverse complement assay, where in the extension pool CT the dd**T**-TAMRA is incorporated.

B) Detailed design of the nonstandard assay for the **c/C-Intron**. The sequence in yellow represents the 109 bp insertion that is only present in the *RHCE_C* allele. The sequences highlighted in pink represent the manual PCR oligo designs, and only individuals with the insertion have a positive amplification of 155 bp. The sequences in green, bold and underline represent the extension oligos used in the extension pool **GA** and **CT**. Since the interrogated nucleotides (highlighted in green) are not SNPs, only the dd**G**-BODIPY-Fluorescein and ddC-TAMRA are incorporated during the extension reaction, leading to a non-standard analysis of **G**/- and **C**/- respectively.

C) Generation of 155 bp amplicon bands using the PCR oligos design detailed in section (B). After completing the PCR cycling, 5 µL out of 10 total µL of PCR reaction were mixed with 10 µL PCR grade water and loaded on a 4% agarose gel (50 bp mass ladder). Only individuals with the 109 bp insertion have a positive amplification of 155 bp.

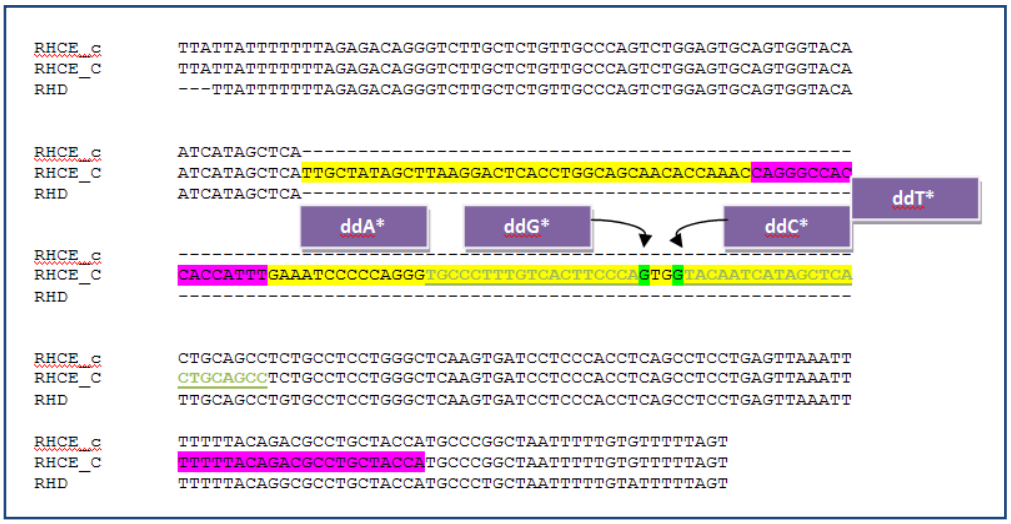
D) Polar genotyping cluster view that illustrates the ddA* detection (G**A** extension pool) of the non-standard assay design for the polymorphism **c/C-rs1053344.** The cluster in green represents the positive ddA* extension, therefore it reflects the presence of the A allele, and since the PCR amplification is only specific to *RHCE_c*, the individuals located in this cluster are either A/G (c/C) or A/A (c/c), whereas the individuals in the blue cluster are not RHCE_c carriers, leading to C/C. The same logic is applied for the design in the reverse direction using the C**T** extension pool (dd**T*** detection).

E) Polar and Cartesian genotyping cluster views that illustrate the ddG* detection (GA extension pool) of the non-standard designs of the polymorphism **c/C-Intron.** Since the PCR amplification is only possible in individuals with the 109 bp insertion, the ddG* detection reflects only the "presence" of the insertion and is translated in the genotyping clusters. The blue cluster represents the positive ddG* extension, therefore it reflects the presence of the G nucleotide and individuals located in this cluster are either c/C or C/C, whereas the individuals in the green cluster are non-extended with ddG* (no carrier of insertion) and have the presence of RHCE_c/c allele.

A


RHCE_c
RHCE_C
RHd
consensus
CCACCCTAAATCTCGTCTGCTTCCCCCTCCTCCTTCTCACCATCTCCCCACCGAGCAGTCGGCCAAGATCTGACCGTGAT
CCACCCTAAATCTCGTCTGCTTCCCCCTCGTCCTTCTCGCCATCTCCCCACCGAGCAGTTGGCCAAGATCTGACCGTGAT
CCACCCTAAATCTCGTCTGCTTCCCCCTCGTCCTTCTCGCCATCTCCCCACCGAGCAGTTGGCCAAGATCTGACCGTGAT
CCACCCTAAATCTCGTCTGCTTCCCCCTCcTCCTTCTCaCCATCTCCCCACCGAGCAGTcGGCCAAGATCTGACCGTGAT
17121....17130.....17140.....17150.....17160.....17170.....17180.....17190.....
ddG*
ddA*
RHCE_c
RHCE_C
RHd
consensus
GGCGGCCCTTGGCTTGGGCTTCCTCACCTCAATTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
GGCGGCCATTGGCTTGGGCTTCCTCACCTCGAGTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
GGCGGCCATTGGCTTGGGCTTCCTCACCTCGAGTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
GGCGGCCcTTGGCTTGGGCTTCCTCACCTCaAaTTTCCGGAGACACAGCTGGAGCAGTGTGGCCTTCAACCTCTTCATGC
17201....17210.....17220.....17230.....17240.....17250.....17260.....17270.....
RHCE_c
RHCE_C
RHd
consensus
TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTCCTGGGAAGGTGGTCATCACACTG
TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTCCTGGGAAGGTGGTCATCACACTG
TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTTCTGGGAAGGTGGTCATCACACTG
TGGCGCTTGGTGTGCAGTGGGCAATCCTGCTGGACGGCTTCCTGAGCCAGTTCCCTcCTGGGAAGGTGGTCATCACACTG
17281....17290.....17300.....17310.....17320.....17330.....17340.....17350.....

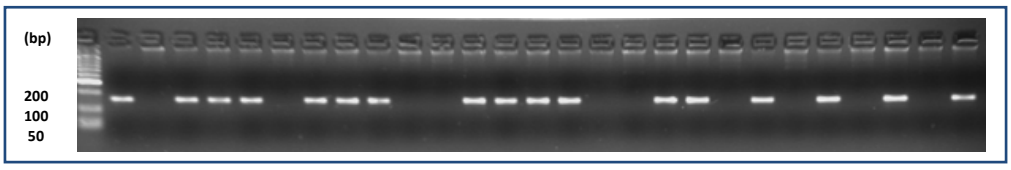
B


RHCE_c
RHCE_C
RHD
TTATTATTTTTTTAGAGACAGGGTCTTGCTCTGTTGCCCAGTCTGGAGTGCAGTGGTACA
TTATTATTTTTTTAGAGACAGGGTCTTGCTCTGTTGCCCAGTCTGGAGTGCAGTGGTACA
---TTATTTTTTTAGAGACAGGGTCTTGCTCTGTTGCCCAGTCTGGAGTGCAGTGGTACA
RHCE_c
RHCE_C
RHD
ATCATAGCTCA----------
ATCATAGCTCATTGCTATAGCTTAAGGACTCACCTGGCAGCAACACCAAACCAGGGCCAC
ATCATAGCTCA----------
ddT*
ddA*
ddG*
ddC*
RHCE_c
RHCE_C
RHD

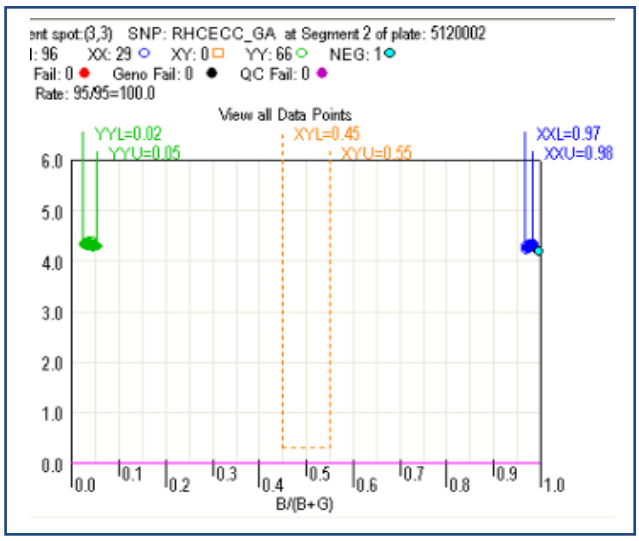
CACCATTTGAAATCCCCCAGGGTGCCCTTTGTCACTTCCCAGTGGTACAATCATAGCTCA

RHCE_c
RHCE_C
RHD
CTGCAGCCTCTGCCTCCTGGGCTCAAGTGATCCTCCCACCTCAGCCTCCTGAGTTAAATT
CTGCAGCCTCTGCCTCCTGGGCTCAAGTGATCCTCCCACCTCAGCCTCCTGAGTTAAATT
TTGCAGCCTGTGCCTCCTGGGCTCAAGTGATCCTCCCACCTCAGCCTCCTGAGTTAAATT
RHCE_c
RHCE_C
RHD
TTTTTACAGACGCCTGCTACCATGCCCGGCTAATTTTTGTGTTTTTAGT
TTTTTACAGACGCCTGCTACCATGCCCGGCTAATTTTTGTGTTTTTAGT
TTTTTACAGGCGCCTGCTACCATGCCCTGCTAATTTTTGTATTTTTAGT

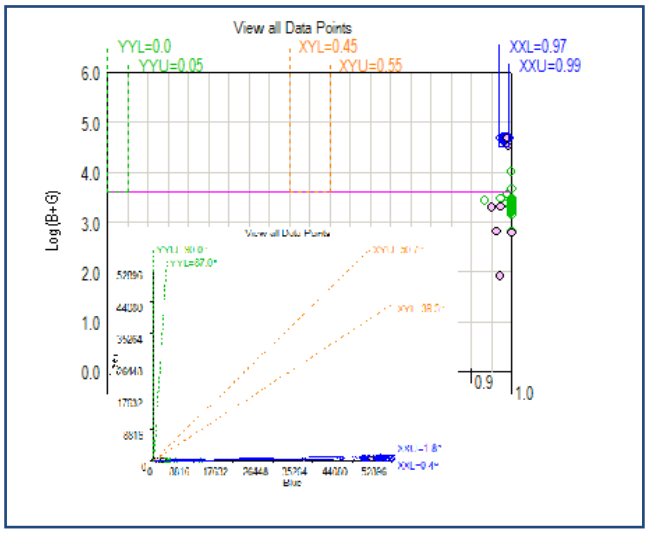
C


(bp)
200
100
50

D


ent spot:(3,3) SNP: RHCECC_GA at Segment 2 of plate: 5120002
I: 96 XX: 29 XY: 0 YY: 66 NEG: 1
Fail: 0 Geno Fail: 0 QC Fail: 0
Rate: 95/95=100.0
View all Data Points
YYL=0.02
YYU=0.05
XYL=0.45
XYU=0.55
XXL=0.97
XXU=0.98
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
B/(B+G)

E


View all Data Points
YYL=0.0
YYU=0.05
XYL=0.45
XYU=0.55
XXL=0.97
XXU=0.99
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Log (B+G)
0.9
1.0

1.3. SNPstream minor blood group antigen genotyping assay optimisation

To evaluate the quality of the SNPstream design (PCR and extension oligos version 1), a total of 44 Coriell Institution of Medical Research (Camden, NJ) DNAs (15 Caucasians, 15 African American, 14 Asians) were genotyped. In addition, since the ultimate source of DNAs for the clinical genotyping is expected to be from the FTA cards, DNAs from forty different FTA sample were also tested. The genotyping results were evaluated following five different parameters; 1- the clustering quality, which shows how well the three expected clusters are separated (homozygote AA, heterozygote AB and homozygote BB), 2- the call rate percentage, referring to the number of samples that were successful genotyped Vs. the number of samples that failed, 3- the genotyping concordances of results obtained in both DNA directions (Panels GA and CT), 4- the comparison of allele frequency (Minor Allele Frequencies; MAF) with the expected values published by dbSNP, and 5- the Hardy-Weindberg (HW), which is a statistical value that reflects the normal allele distribution in a non-related population. The three first parameters are key indicators of the assay robustness whereas the last two parameters are mainly an indicator of assay design deficiency in relation to the genomic background.

1.3.1. Phase I assay optimisation

To know which FTA punching techniques will be further applied in the genotyping experiments, a total of 200 FTA punches were performed following five different punching size techniques. All the punches were then diluted 5, 10 and 20 times in water and tested for multiplex PCR amplification using the SNPstream PCR oligos version 1. Analysis of PCR products on a 4% agarose gel indicated that the for Coriell DNA, the 1.2 mm punches (one and two spots) and the 2 mm (one spot) were the samples that were amplified with the highest yield, whereas the 2.0 mm (two spots) and the 3 mm (one and two spots) showed the weakest amplification, especially when the DNAs were diluted (Figure 19-A). The lack of PCR amplification when the DNAs were titrated down was a clear indication that the limit of detection had been reached for these punching techniques. The extracted DNAs from the different punching techniques were also tested for a complete genotyping

analysis and as per the results previously observed on a 4% agarose gel, the DNAs extracted from the 2 mm (two spots) and 3.0 mm (on and two spots) did not reach the required yield to generate good clustering and results were not interpretable (Figure 19-B). Based on these results, the DNAs from 2mm punches (one spot) were further tested for integrity using two PCR amplifications (APOE and CYP2D6) known in-house to be problematic on degraded or poor quality extracted DNAs. Figure 19-C illustrates clean and perfect PCR amplification for all ten samples for both PCR amplifications, indicating that DNAs were extracted in a high quality manner. The DNAs were also tested for Picogreen and Nanodrop quantification and reliable results were not obtained due to the presence of high background possibly due to DNA levels that are below detection (results not shown). No further investigations have been performed for quantification and we decided to use the DNAs directly from the extraction for the subsequent genotyping assay optimisation.

The genotyping results obtained are categorized into two groups which are "markers failed optimisation" and "markers passed optimisation". Within the group that needs further optimisation; two action plans are required which are "PCR condition modification" and "assay redesign". After the phase 1 genotyping optimisation on 84 total samples (44 Coriells and 40 FTA cards) for a total of 24 markers (12 markers in the two DNA directions), only five markers needed improvement (see Table 8). In summary, all markers that passed the phase 1 optimisation, for over 2,000 genotyping calls we found to have a 97% call rate in both extension pools, which represented failure of only one sample or less, and more than 99% of the genotyping results were concordant in both extension pools. Also, perfect matches with the expected allele frequencies and Hardy-Weindberg equilibrium was observed (Figure 20-A). In the group of markers that needed improvement, the marker k/K-rs8176058 showed good results for all different quality control steps, but clear lack of sensitivity was observed in the GA extension mix using the DNA extracted from FTA cards. This result suggested that the problem was not related to the PCR amplification since both CT and GA extensions share the same PCR amplification fragment but highlighted that the GA extension oligo was not sensitive enough using low DNA concentration as template. Cautionary Hardy-

Weindberg values were also obtained for both the GA and CT extension pools for the k/K-rs8176058 marker only in the genotyping of FTA samples. These results are likely because the FTA samples were compiled by Héma-Québec based on interesting phenotypes, therefore the samples selection may not be completely unbiased (Figure 20-B). For the marker Jka/b, a clear PCR amplification problem was observed since in both the GA and CT extension pools very low and poorly defined clusters were achieved. The same results were obtained using either the DNA from the Coriell samples or from the 2 mm FTA cards, confirming that the problems resided mainly at the PCR level. After reviewing the assay designs, the Autoprimer had generated a pair of oligos that amplified a 301 bp DNA sequence, which correlated perfectly with the PCR amplification problems, as the longest amplicon is always under-favoured in a multiplex PCR amplification (Figure 20-C).

The markers M/N in the GA and CT extension pools had failed genotyping results because of Hardy-Weindberg (HW) and the allelic frequency failures (Figure 20-D). For this marker the MAF is expected to be around 50% and only 27% was observed, also a HW value of 22 was observed, which represents a major allele distribution failure (normal = 0 to 5). Without the Hardy-Weindberg values, the problem would not have been identified, as good clustering separations and call rates were initially observed with the SNPstream software calculation (Figure 20-D, blue circles). After closer examination of the clustering there was evidence of two heterozygote clusters, one less abundant (all African American samples; cluster 1) and one more abundant (all Caucasians; cluster 2) and a manual re-clustering was performed (the actual blue, orange and green clusters). Following the manual re-clustering, the HW values and allele frequencies were normal, indicating that the double heterozygote clusters were the basis of the problem. Additionally, the problem is clearly associated with a specific population genetics issue since heterozygote calls from African American and Caucasian populations were split into two different clusters. Considering the same problem was observed in the GA and CT pools, it indicated that the issue was linked mostly with the PCR design and not with the extension probe designs. In order to pinpoint the problem and modify the designs accordingly, twelve samples were selected from the sub-clusters 1 and 2 and were

sequenced for the M/N marker DNA sequence region. In Figure 20-E, the sequencing results indicated that all samples were heterozygous, and as well as the M/N marker (position 233), two more SNPs were observed, one upstream of the M/N variant at position 254 and one downstream at position 221. The SNP located at 221 was equally present for all samples from both cluster 1 and 2 whereas the SNP located at position 254 was only present for samples located in cluster 1 (African American specific). Sequencing analysis was then performed for the DNA sequences that encompassed the forward and reverse PCR oligo sequences. No differences were observed for all tested samples for DNA sequences that encompassed the forward oligo design (position 320 to 340), but a G/A SNP (rs4867) specific to cluster 1 (African American) was observed at position 199 of the surrounding reverse PCR oligo DNA sequence (start position = 180, end position = 199) (Figure 20-F). This finding was in direct line with the observed genotyping issues since the underlying SNP is specifically associated with the outlier samples (African Americans), and it was located in the 3' end of the reverse oligo which is known to affect amplification of the non-matching allele. In this case, the A allele (position 199) seems to be tightly linked with the A allele of the M/N variation since the outlier samples were found with the A signals.

1.3.2. Phase II assay optimisation

Based on these observations, appropriate actions were taken to create improved PCR and extension oligo designs for version 2. For the phase II optimisation, a total of 192 samples (32 Caucasian, 32 African American, 32 Asians and 96 FTA) were genotyped. In the case of the marker k/K, to overcome the lack of sensitivity when low DNA template was used, the GA extension pool oligo was increased from 5 μM to 10 μM. Also, after reviewing the PCR and extension sequence annotation, appropriate modification of the first base located at the 5'end was performed on the forward PCR oligo. As pictured in Figure 21-A, the GA assay sensitivity is now comparable to the CT extension oligo. Both assays (GA and CT) now have a call rate of more than 99.5% and 100% genotyping concordance. Perfect

HW and MAF values are now observed. The PCR oligos of Jk a/b were manually redesigned to diminish the amplicon size. With respect to Tm, GC% and multiplex PCR compatibility, the amplicon size was then cut from 301 bp to 194 pb. Also, the GA extension oligo was modified to correct a mistake in the sequence. Following these modifications, both GA and CT genotyping clusters were perfectly separated with high intensities and the observed call rates were equal to 100%. A concordance of 100% was also achieved when genotyping results from both extension pools were compared. A warning regarding the HW value was obtained, but similar to the explanation for k/K during phase I, the HW values deviate from reality only in the FTA samples, which were pre-selected by Héma-Québec (Figure 21-B). Both forward and reverse PCR oligos of M/N-rs7682260 were modified to avoid any underlying SNPs that could ultimately contribute to a bad cluster distribution. The forward oligo was moved approximately 60 bp upstream in order to incorporate two mismatches with *GYPB* and *GYPE* and the reverse oligo was modified at the 3' to incorporate a degenerate bases (G and A) in order to accommodate the presence of the rs4867 SNP. Also, following the sequencing results, the GA and CT extension oligos were modified with degenerated bases to accommodate the rs7658293 and rs4449373 markers (Figure 21-C). In both the GA and CT assays, the double heterozygote clusters disappeared leading to perfect HW and allele frequency values. The African American samples were perfectly incorporated to a unique heterozygote cluster indicating that the new reverse PCR oligo amplifies equally the A and G alleles (Figure 21-D). Finally, both the reverse PCR oligo and forward oligo of HPA1 a/b and Fy a/b were corrected with degenerate nucleotides to accommodate the previously unknown underlying SNPs (Table 2 of the Materials and Methods section).

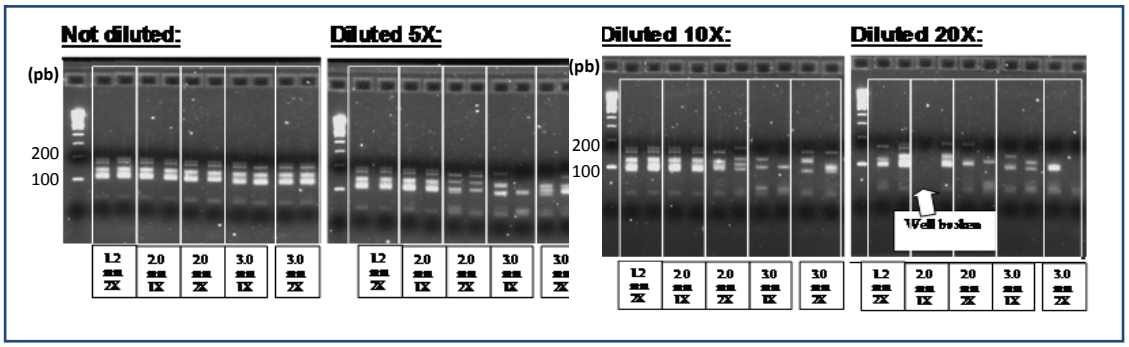
Figure 19: Extraction of DNA from FTA cards

A) 4% agarose gel after multiplex PCR using the SNPstream PCR oligo mixes version 1. Typical example of multiplex PCR results obtained using 2 µL of DNA extracted from 1.2 mm, 2.0 mm and 3.0 mm spots sizes, undiluted and diluted 5, 10 and 20X respectively. DNAs from each punching technique were amplified in duplicate and are represented within the white boxes.

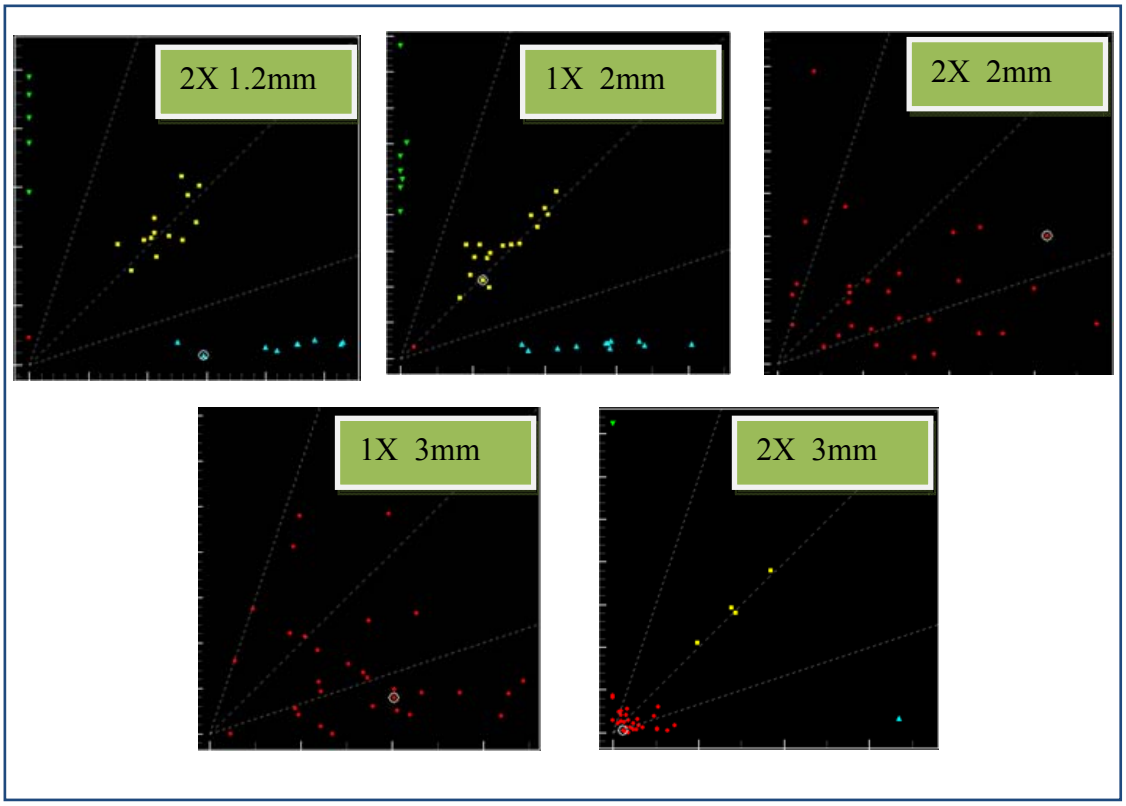
B) Typical genotyping clusters of marker Fy a/b using the multiplex PCR and CT extension pool from version 1 and five different FTA punching techniques (two spots 1.2 mm, one and two spots 2.0 mm and one and two spots 3 mm). The actual examples pictured are the results obtained using 1/10 dilution. The blue cluster represents the C allele, the green represents the T allele and the yellow represents the CT alleles. The six samples points shown in the genotyping cluster of 3.0mm (two spots) represent the Coriell DNA samples.

C) 1% agarose gel analysis of the CYP2D6 and APOE fragments. PCR bands are expected at 1.7 kb and 900 bp respectively, when positive amplifications of CYP2D6 and APOE are obtained. The wells 1 to 10 represent the CYP2D6 amplification and wells 36 to 45 represent the APOE amplification.

A


Not diluted:
Diluted 5X:
Diluted 10X:
Diluted 20X:
(pb)
200
100
Well broken

B


2X 1.2mm
1X 2mm
2X 2mm
1X 3mm
2X 3mm

C

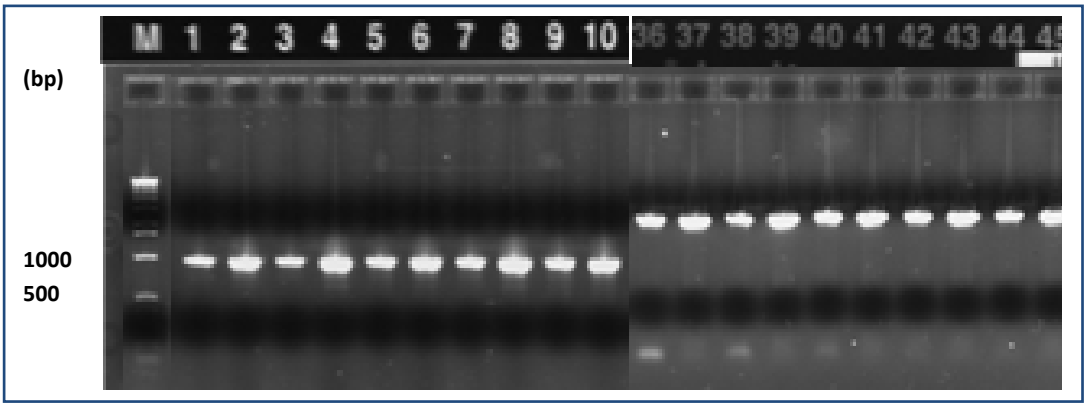


Table 8: Summary results after the phase 1 genotyping optimisation

Antigens	Marker ID	Extension Pools	Cluster Separation	Call rate %	F/R Genotyping concordance	Reported MAF (Caucasian)	Experimental MAF	HW	Improvement needed
Kap a/b	rs8176059	GA	Perfect Cluster Separation	100%	>99%	2%	2%	Passed	
		CT	Perfect Cluster Separation	100%	>99%			Passed	
HPA1 a/b	rs5918	GA	Perfect Cluster Separation	100%	>99%	15%	16%	Passed	
		CT	Perfect Cluster Separation	100%	>99%			Passed	
M/N	rs7682260	GA	Good Separation, Double Heterozygote	100%	>99%	50%	27%	Failed	Assay Redesign
		CT	Good Separation, Double Heterozygote	>97%	>99%			Failed	Assay Redesign
s/S	rs7683365	GA	Good Separation	100	>99%	35%	32%	Passed	
		CT	Good Separation	100	>99%			Passed	
c/C	rs1053344	GA	Not Standard Clustering	100%	>99%	10%	N/D	N/A	
		CT	Not Standard Clustering	100%	>99%			N/A	
c/C-Intron	109 bp insertion	GA	Not Standard Clustering	100%	>99%	N/D	N/A	N/A	
		CT	Not Standard Clustering	100%	>99%			N/A	
HPA2 a/b	rs6065	GA	Perfect Cluster Separation	100%	>99%	10%	10%	Passed	
		CT	Perfect Cluster Separation	100%	>99%			Passed	
Fy a/b	rs12075	GA	Perfect Cluster Separation	100%	>99%	50%	48%	Passed	
		CT	Perfect Cluster Separation	100%	>99%			Passed	
Jk a/b	rs1058396	GA	Bad Cluster separation and Low Intensity	95%	>99%	55%	50%	Passed	Assay Redesign
		CT	Bad Cluster separation and Low Intensity	95%	>99%			Passed	Assay Redesign
k/K	rs8176058	GA	Perfect Cluster Separation, Low intensity	100%	>99%	4%	7%	Passed	Assay conditions
		CT	Perfect Cluster Separation	100%	>99%			Passed	
c/E	rs609320	GA	Not Standard Clustering	100%	N/A	10%	N/A	N/A	
		CT	Not Standard Clustering	100%	N/A			N/A	
HPA5 a/b	rs10471371	GA	Perfect Cluster Separation	100%	>99%	8%	7%	Passed	
		CT	Perfect Cluster Separation	100%	>99%			Passed	

The above table represents a summary of results obtained after the phase I genotyping optimisation. The following results are evaluated on 84 samples, where 44 are from the Coriell DNA panels (15 Caucasian, 15 African American and 14 Asians) and 40 are from the FTA cards (2 mm punches), for a total of 24 genotyping assays (GA and CT). The phase 1 genotyping optimisation evaluated five main criteria: cluster separation, which evaluates the quality of the heterozygote cluster separated from both homozygote clusters, the marker call rate, representing the ratio of successful genotype calls over the failed calls, the genotyping call concordance, representing the percentage of samples that have identical genotyping results in both extension pools (F/R; Forward and reverse DNA strands), the comparison of the theoretical and experimental minor allele frequencies (MAF), and finally the Hardy-Weindberg evaluation (HW), which represents a statistical interpretation of normal (Passed) and abnormal (Failed) allele distributions. Markers highlighted in red represent the markers that required phase II optimisation for either PCR condition optimisation or complete assay redesign.

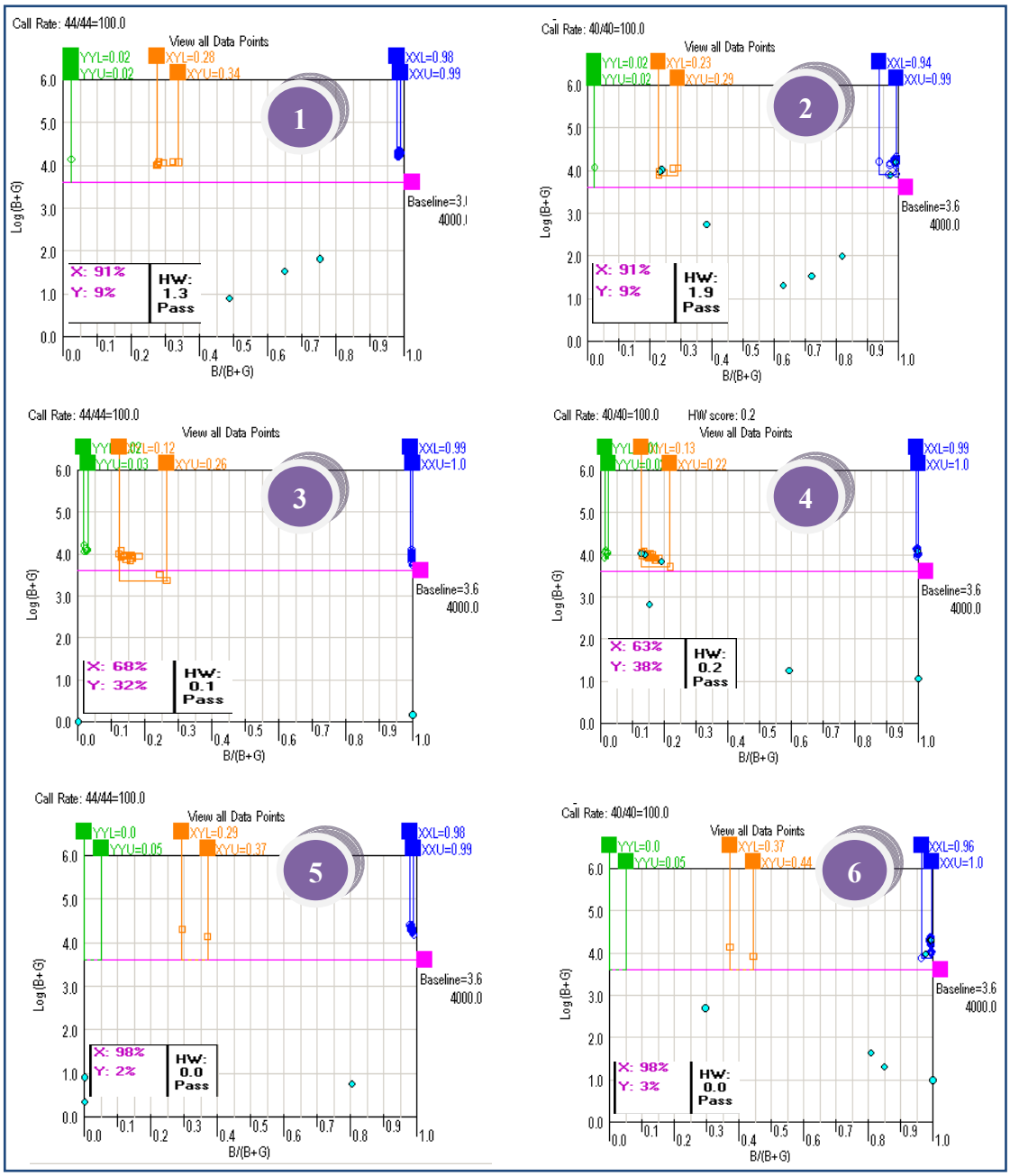
Figure 20: Phase I genotyping results optimisation

A) Typical polar genotyping representation of three successful markers from the phase 1 optimisation. The first two genotyping results (1 and 2) represent the HPA2 a/b (GA pool) from the Coriell DNAs and the FTA-2 mm DNAs respectively, the second two results (3 and 4) represent the s/S (GA pool) and the last two results (5 and 6) represent the genotyping clusters from Kap a/b (GA pool) from the Coriell and FTA-2 mm DNAs respectively. Computational marker call rates, allele frequencies and Hardy-Weinberg calculations are represented in the upper left corner and in the lower left side of each genotyping cluster respectively. The blue clusters represent the presence of the G allele whereas the green clusters represent the presence of the A allele and the orange clusters represent the presence of both G and A alleles. Samples in gray represent the negative controls.

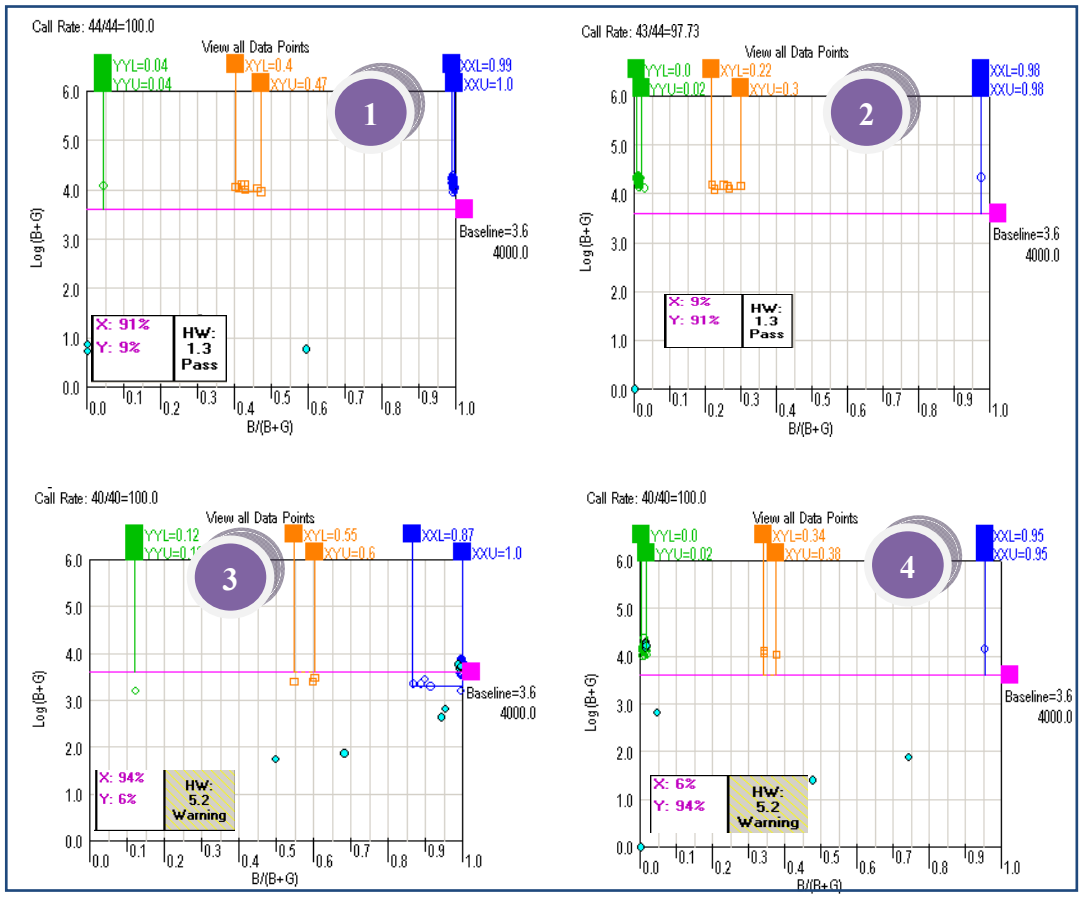
B) Polar representation of the genotyping results obtained for the marker k/K with the pools GA (1) and CT (2) generated with Coriell DNAs and FTA DNAs (3 and 4). These results depict an assay that needed optimisation after using low DNA concentration as the PCR template.

C) Polar representation of the genotyping results obtained for the marker Jka/b with the pools GA (1) and CT (2) generated with Coriell DNAs and FTA DNAs (3 and 4). The following results depict a typical assay that needed redesign due to clustering separation failure.

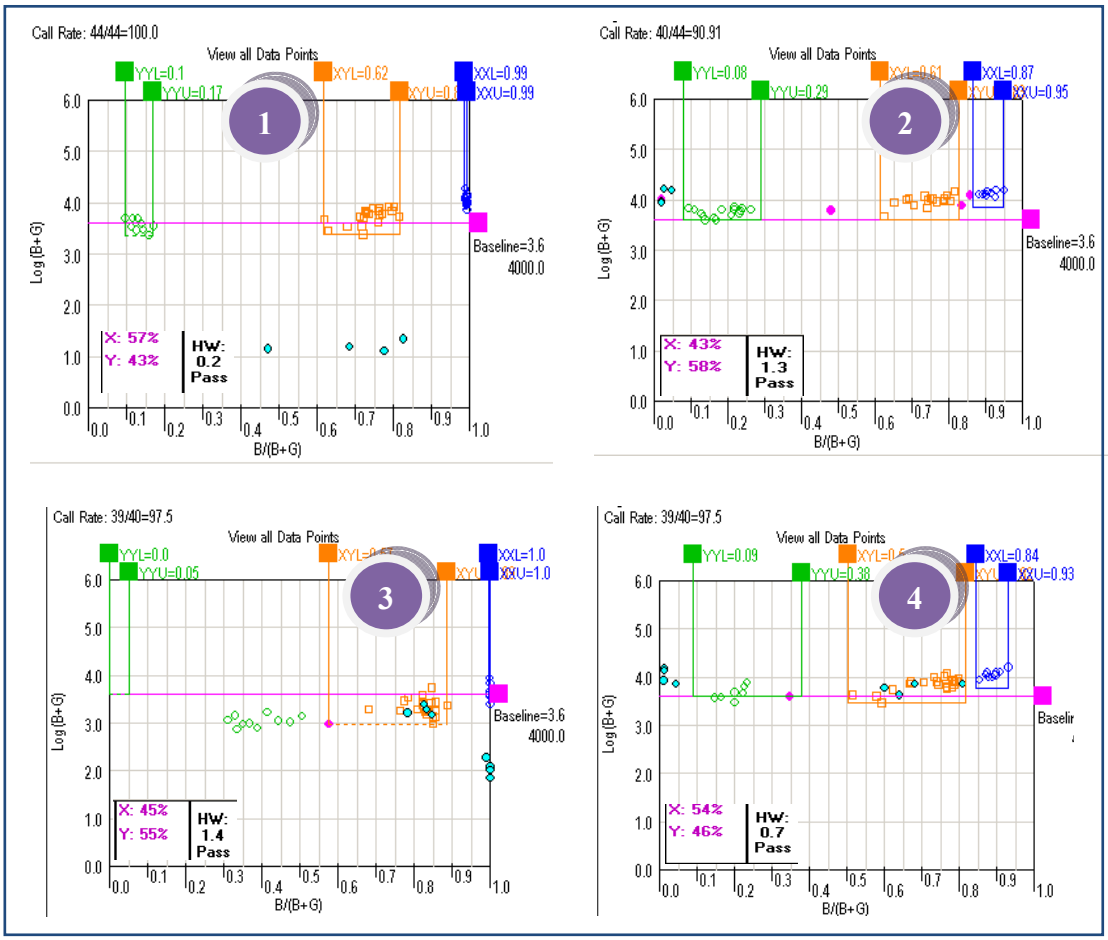
A


Call Rate: 44/44=100.0
View all Data Points
YYL=0.02
YYU=0.02
XYL=0.28
XYU=0.34
XXL=0.98
XXU=0.99
1
Log(B+G)
B/(B+G)
Baseline=3.
4000.
X: 91%
Y: 9%
HW:
1.3
Pass
Call Rate: 40/40=100.0
View all Data Points
YYL=0.02
YYU=0.02
XYL=0.23
XYU=0.29
XXL=0.94
XXU=0.99
2
Log(B+G)
B/(B+G)
Baseline=3.6
4000.0
X: 91%
Y: 9%
HW:
1.9
Pass
Call Rate: 44/44=100.0
View all Data Points
YYL=0.02
YYU=0.03
XYL=0.12
XYU=0.26
XXL=0.99
XXU=1.0
3
Log(B+G)
B/(B+G)
Baseline=3.6
4000.0
X: 68%
Y: 32%
HW:
0.1
Pass
Call Rate: 40/40=100.0
HW score: 0.2
View all Data Points
YYL=0.0
YYU=0.0
XYL=0.13
XYU=0.22
XXL=0.99
XXU=1.0
4
Log(B+G)
B/(B+G)
Baseline=3.6
4000.0
X: 63%
Y: 38%
HW:
0.2
Pass
Call Rate: 44/44=100.0
View all Data Points
YYL=0.0
YYU=0.05
XYL=0.29
XYU=0.37
XXL=0.98
XXU=0.99
5
Log(B+G)
B/(B+G)
Baseline=3.6
4000.0
X: 98%
Y: 2%
HW:
0.0
Pass
Call Rate: 40/40=100.0
View all Data Points
YYL=0.0
YYU=0.05
XYL=0.37
XYU=0.44
XXL=0.96
XXU=1.0
6
Log(B+G)
B/(B+G)
Baseline=3.6
4000.0
X: 98%
Y: 3%
HW:
0.0
Pass

B


Call Rate: 44/44=100.0
View all Data Points
YYL=0.04
YYU=0.04
XYL=0.4
XYU=0.47
XXL=0.99
XXU=1.0
1
Baseline=3.6
4000.0
X: 91%
Y: 9%
HW:
1.3
Pass
Log (B+G)
B/(B+G)
Call Rate: 43/44=97.73
View all Data Points
YYL=0.0
YYU=0.02
XYL=0.22
XYU=0.3
XXL=0.98
XXU=0.98
2
Baseline=3.6
4000.0
X: 9%
Y: 91%
HW:
1.3
Pass
Log (B+G)
B/(B+G)
Call Rate: 40/40=100.0
View all Data Points
YYL=0.12
YYU=0.1
XYL=0.55
XYU=0.6
XXL=0.87
XXU=1.0
3
Baseline=3.6
4000.0
X: 94%
Y: 6%
HW:
5.2
Warning
Log (B+G)
B/(B+G)
Call Rate: 40/40=100.0
View all Data Points
YYL=0.0
YYU=0.02
XYL=0.34
XYU=0.38
XXL=0.95
XXU=0.95
4
Baseline=3.6
4000.0
X: 6%
Y: 94%
HW:
5.2
Warning
Log (B+G)
B/(B+G)

C

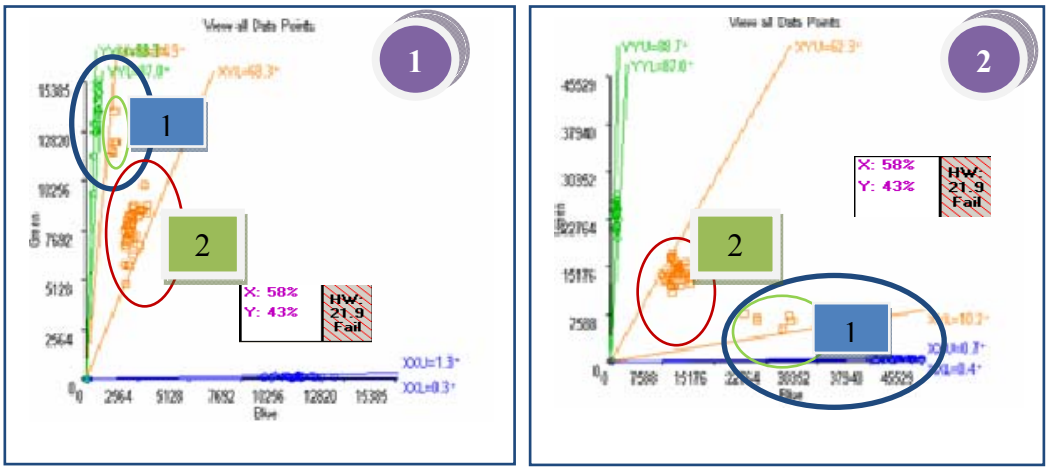

Call Rate: 44/44=100.0
View all Data Points
YYL=0.1
YYU=0.17
XYL=0.62
XXL=0.99
XXU=0.99
1
Baseline=3.6
4000.0
X: 57%
Y: 43%
HW:
0.2
Pass
Log (B+G)
B/(B+G)
Call Rate: 40/44=90.91
View all Data Points
YYL=0.08
YYU=0.29
XXL=0.87
XXU=0.95
2
Baseline=3.6
4000.0
X: 43%
Y: 58%
HW:
1.3
Pass
Log (B+G)
B/(B+G)
Call Rate: 39/40=97.5
View all Data Points
YYL=0.0
YYU=0.05
XXL=1.0
XXU=1.0
3
Baseline=3.6
4000.0
X: 45%
Y: 55%
HW:
1.4
Pass
Log (B+G)
B/(B+G)
Call Rate: 39/40=97.5
View all Data Points
YYL=0.09
YYU=0.38
XXL=0.84
XXU=0.93
4
X: 54%
Y: 46%
HW:
0.7
Pass
Log (B+G)
B/(B+G)

D) Cartesian representation of the genotyping results obtained for the marker M/N with the GA (1) and CT pools (2) generated with FTA DNAs. The blue circles represent the original clustering results performed by the software before the manual modification. The red and green circles represent the heterozygote subclusters 1 (less frequent) and 2 (more frequent) respectively. In the upper left figure, the blue, orange and green clusters represent the G, GA and A alleles respectively. The following results depict a typical assay that needs assay redesign because of the Hardy-Weindberg and allelic frequency failures.

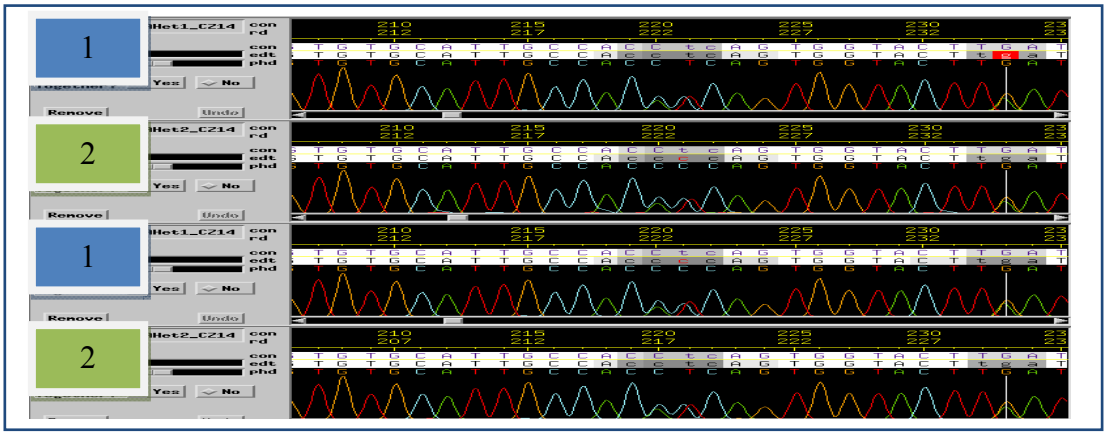
E) Sequencing chromatogram representation of four tested samples, two from the less abundant cluster (#1) and two from the most abundant cluster (#2). The sequences pictured in the upper side are covering around 35 bp of DNA sequences downstream of the M/N variation (position 233) encompassing the extension oligo sequence used for the SNP detection in the GA pool. The sequences pictured in the lower side are covering ~30 bp of DNA sequences upstream of the M/N SNP (position 233) including the extension oligo sequence used in the CT pool. The new SNPs are located at positions 221 and 254 of downstream and upstream sequences respectively.

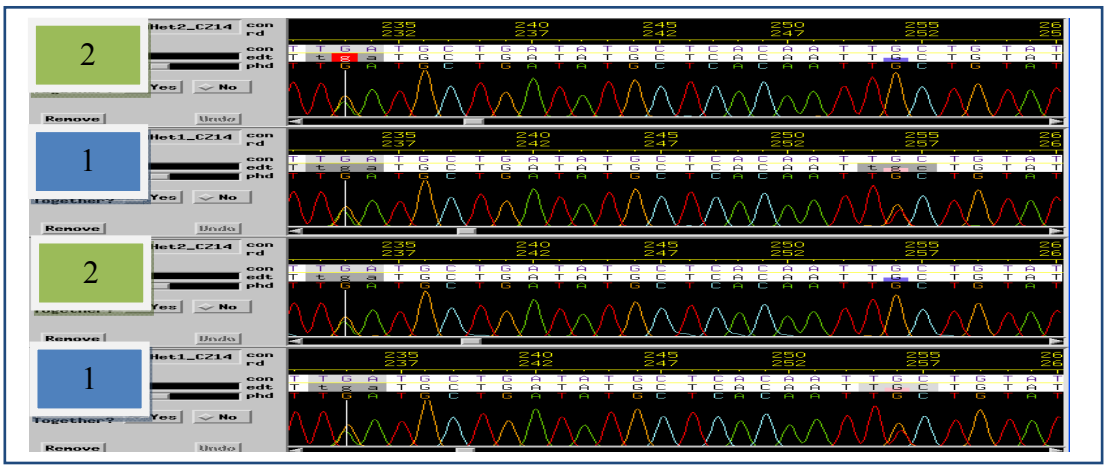
F) Sequencing chromatogram representation of four tested samples, two from the less abundant cluster (#1) and two from the most abundant cluster (#2). The sequences pictured in the upper side are covering around 30 bp of DNA sequences downstream of the starting base of the forward oligos. The oligos start at position 320 and finish at position 340. The sequences pictured in the lower side represented 30 pb of the surrounding DNA sequences of the reverse oligos. The reverse oligos is starting at 180 and finishing at 199 and the new SNP is located at position 199.

D

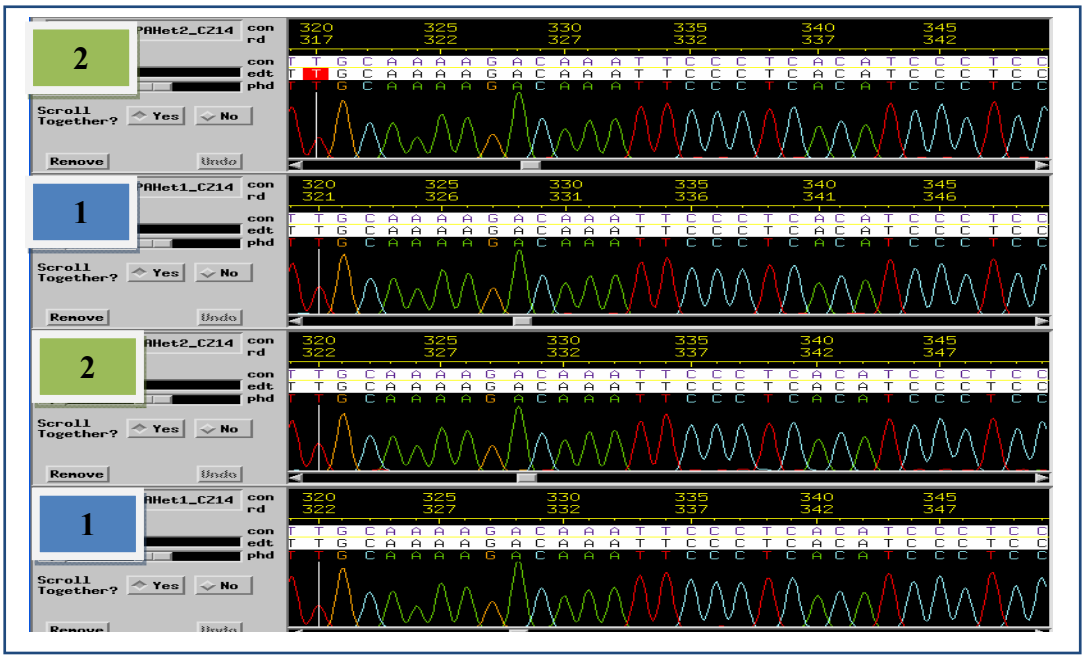

View all Data Points
1
2
X: 58%
Y: 43%
HW: 21.9 Fail
Green
Blue

E


1
2
1
2


2
1
2
1

F


2
1
2
1
Scroll Together?
Yes
No
Remove
Undo

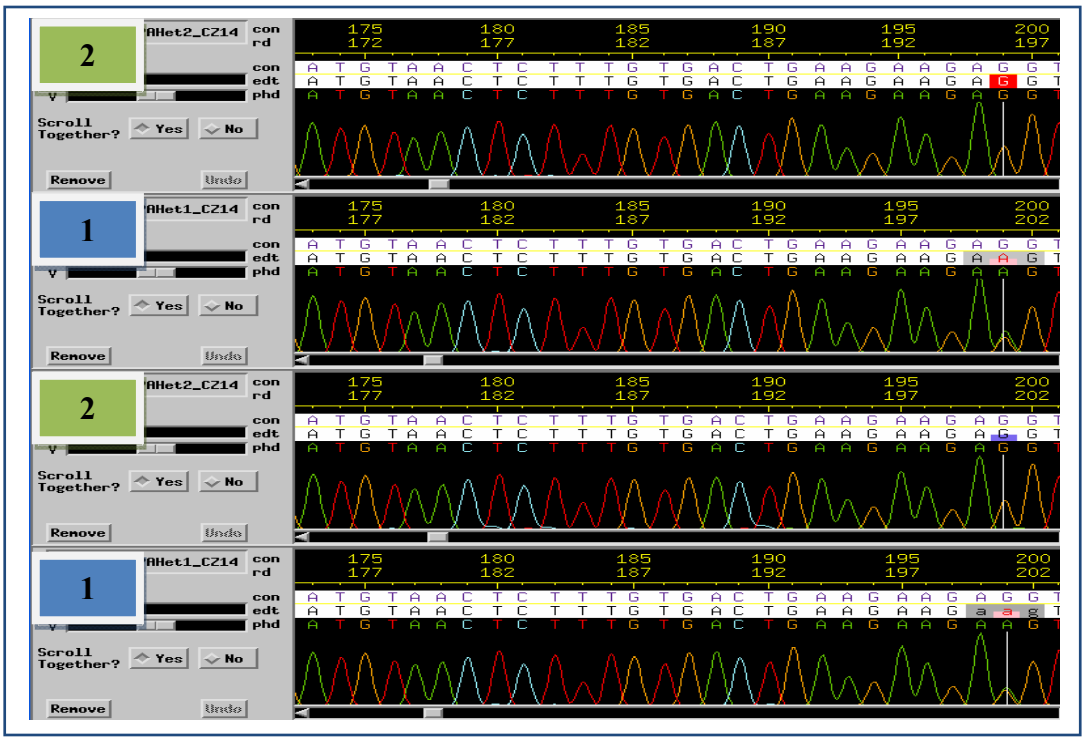

2
1
2
1
Scroll Together?
Yes
No
Remove
Undo

Figure 21: Phase II genotyping results optimisation

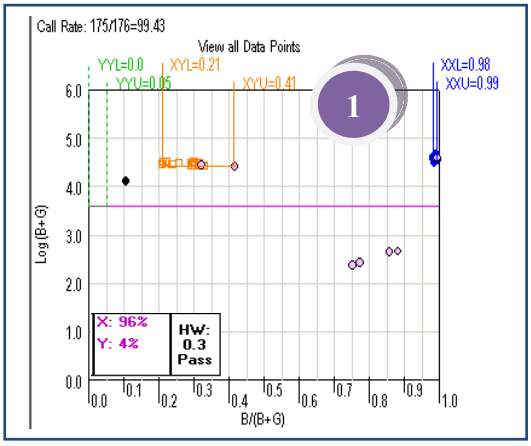
A) Polar graph representation of the k/K genotyping after the optimisation of the GA extension oligo concentration (1) and the non-modified CT assay (2). Coriell DNAs and FTA DNAs are analyzed all together for a total of 192 samples.

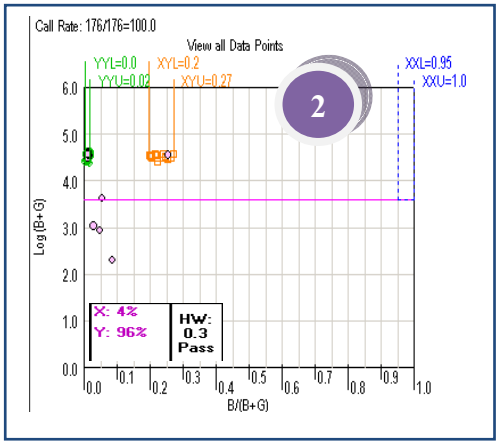
B) Genotyping representation of the optimized Jk a/b assays after the complete PCR oligos redesign and minor modification of the GA extension oligo. Coriell DNAs and FTA DNAs are analyzed all together for a total of 192 samples.

C) M/N PCR and extension oligos re-design based on the sequencing results. The new PCR oligos are represented in red whereas the forward oligo is located between position 28490 and 28510 and the reverse oligo is located between position 28630 and 28650. The M/N marker is highlighted in green.

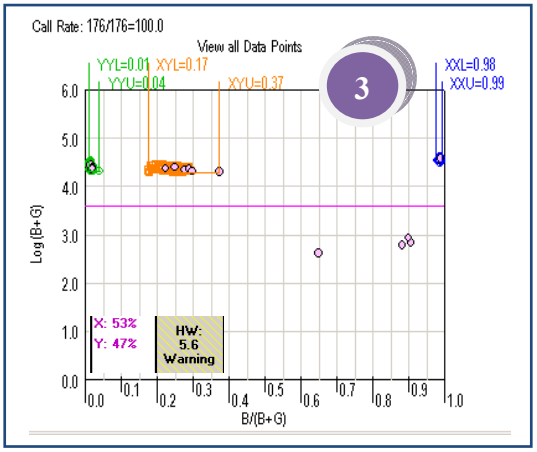
D) Clustering representation of M/N GA (1) and CT (2) assays after PCR and extensions oligos optimisation. Coriell DNAs and FTA DNAs are analyzed all together for a total of 192 samples.

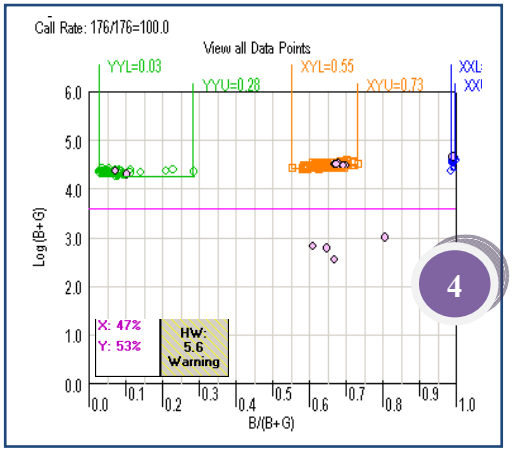
A





B





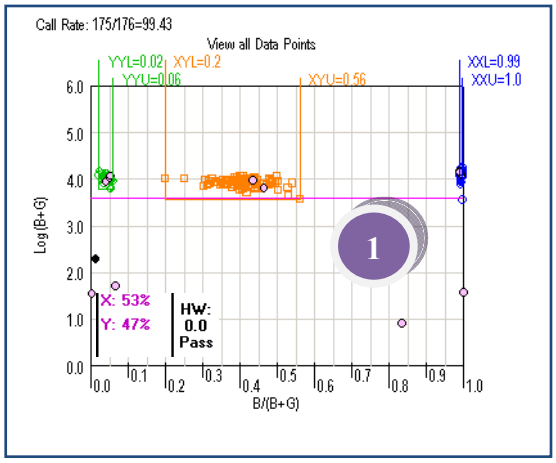
C

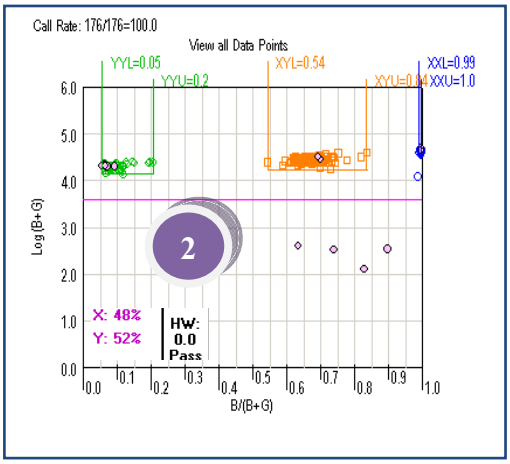
```
GYPA        TGGAGGGATGTGAG-GGAATTTGTCTTTTGCAATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
GYPB        TGGAGGGATGTGAGAGGAATTTGTCTTTTGTGATATGCTTTATGGTCTGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
GYPE        TGGAGGGATGTGAG-GGAATTTGTCTTTTGTGATATGCTTTATGGTCCGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
consensus   TGGAGGGATGTGAG-GGAATTTGTCTTTTGtgATATGCTTTATGGTCcGCTCAGTCACCTCGTTCTTAATCCCTTTCTCA
            28481....28490.....28500....|.28510.....28520.....28530.....28540.....28550.....

                                                    TTCATGGTGACYMCACCGTTACGT
GYPA        ACTTCTATTTTATACAGYAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACYTCTTCTT
GYPB        ACTTCTATTTTATACAGAAATTGTGAGCATATCAGCATTAAGTACCACTGAGGTGGCAATGCACACTTCAACCTCTTCTT
GYPE        ACTTCTATTTTATACAGGAATTGTGAGCATATCAGCATCAAGTACCACTGGTGTGGCAATGCACACTTCAACCTCTTCTT
consensus   ACTTCTATTTTATACAGaAATTGTGAGCATATCAGCATtAAGTACCACTGagGTGGCAATGCACACTTCAACcTCTTCTT
            28561....28570.....28580.....28590.....28600.....28610.....28620.....28630.....

GYPA        CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCATT
GYPB        CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATGATT
GYPE        CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATCGTT
consensus   CAGTCACAAAGAGTTACATCTCATCACAGACAAATGGTTTGTTTTCATTTTTATTTTTAAATTGTGGCTCCGAAATcaTT
            28641....28650.....28660.....28670.....28680.....28690.....28700.....28710.....
```

D


Call Rate: 175/176=99.43
View all Data Points
YYL=0.02
XYL=0.2
YYU=0.06
XYU=0.56
XXL=0.99
XXU=1.0
Log (B+G)
B/(B+G)
1
X: 53%
Y: 47%
HW:
0.0
Pass


Call Rate: 176/176=100.0
View all Data Points
YYL=0.05
XYL=0.54
YYU=0.2
XYU=0.94
XXL=0.99
XXU=1.0
Log (B+G)
B/(B+G)
2
X: 48%
Y: 52%
HW:
0.0
Pass

2. SNPSTREAM MINOR BLOOD GROUP ANTIGEN GENOTYPING ASSAY VALIDATION

To validate the SNPstream genotyping panel version 2, two distinct groups of experiments were conducted. The first set of experiments was designed to evaluate the accuracy of the assay and the second set of experiments was designed to evaluate the robustness of each marker. Within the accuracy group, we evaluated the genotype concordance of each marker from the SNPstream panel against the Sequenom panel. This evaluation is called "cross-technology" concordance, where two different assays and chemistries are compared side-by-side on the same set of samples. Since Sequenom and SNPstream technologies both use single base extension techniques for detection, selected samples from the previous development phases were sequenced and then genotyped to evaluate the concordance of genotypes using two completely different kinds of detection technologies (Table 9). The accuracy of the assay was further tested by investigating the concordance of the minor blood antigens phenotypes inferred from the genotyping results with known minor blood antigen phenotypes previously characterized by the standard and accredited serological methodologies used at Héma-Québec. For this part of the study, 618 extensively phenotyped donors were selected from Héma-Québec's regular blood donors (Table 10). The robustness of the assay was evaluated for reproducibility, where the final genotyping calls from different days were compared, and also tested for inter-technician concordance (Table 11).

2.1. Accuracy evaluation

The cross-technology experiment was conducted on 96 Coriell samples (1/3 Caucasian, 1/3 African-American and 1/3 Asian) and on the same 40 FTA samples used during the development phases. The comparison was performed analysing both DNA directions on both genotyping platforms. A total of 6528 genotyping calls and 13056 inferred phenotypes were compared for this purpose. A total of 7 (0.2%) and 16 (0.5%) genotyping calls failed the SNPstream and Sequenom experiments

respectively, leading to an average call rate for markers of 99.9% and 99.2%. The markers e/E and c/C-intron had the lowest call rates which are perfectly reasonable because of the "non-Standard" nature of the assays. Also, the Sequenom panel had the lowest overall marker call rate since no phase II optimisation had been performed for this panel. Excluding the failed genotypes, an average of 99.8% of genotyping call concordance was obtained from both DNA strands for the SNPstream and Sequenom platforms, and from the both DNA sources (Coriell DNAs and FTA samples). Finally, a concordance of 100% was observed between the results generated by the SNPstream and Sequenom panels for the Coriell DNAs and 97.5% (one discordant) for the FTA samples. These results were more than acceptable considering the variability of DNAs extracted from the FTA cards, and considering that the discordance was from the c/C-intron (Table 9).

Out of 96 Coriell samples, six samples (NA17175, NA17275, NA17223, NA17294, NA17284 and NA19007) were selected for their high frequency of minor alleles (heterozygous and homozygous mutant) and sent for sequencing. This would cross-validate the genotyping results using a different detection technology, and definitively validate these samples for use as internal controls for the genotyping production. 100% concordance was achieved between the results generated by the genotyping and the sequencing (Table 9 and Figure 22). Only the marker c/C-intron has not been sequenced since not enough specific PCR amplification with acceptable amplicon sizes was obtained to generate clean sequencing results (results not shown). Since the reverse PCR oligo was specific only in the presence of the insertion, the resulting concordance measure was obtained by comparing the presence or the absence of the band on a 4% agarose gel. The presence of a band on the gel signified that the sample was a carrier of the insertion and should generate a positive genotyping call (Figure 23).

The ultimate validation of the sensitivity of the assay was to measure how well the assay can predict actual donor phenotypes previously determined by Héma-Québec. Phenotypes were determined using serological based assays, and the inferred phenotypes from the genotyping results were obtained using version 2 of the

SNPstream panel. To perform the comparison, 618 unknown blood samples were collected by Héma-Québec on FTA cards and DNAs were extracted from 2 mm punches and genotyped on both DNA strands. In addition, 32 duplicates samples were present in the sample collection. An average of 99.9% of call rates were obtained when both DNA strands assays were combined. The lowest call rates were seen for the c/C-Intron and e/E markers, where 99.8 % and 99.2 % was obtained, indicating once again that the non-standard assays performed well, especially considering the variability in the quality and quantity of the DNA extracted from FTA cards. The genotyping accuracies between both strands were also very high, varying from 99.8% to 100% with an average of 99.9%. The assay c/C-intron gave the lowest accuracy with 99.8%, corresponding to 1 discordant sample. The sample was not associated with general sample failure and was not border line between the two clusters which could explain the discordances.

As 32 samples were in duplicate in the sample set, a reproducibility analysis was then performed and only two discordant calls were found, one for the e/E assay and one for the c/C-intron assay. For a total of 1963 different phenotypes compared, the concordance rate between the serological phenotypes and the inferred phenotype by genotyping calls was also very high, varying from 97 to 100% for 18 of the 19 antigens. The serological data for Kp b, HPA-2a and HPA-2b were not available.

The only exception to the otherwise excellent concordance was the antigen HPA-1a where only 50% concordance was achieved (3 correct samples out of 6 total tested samples). The genotyping results were confirmed afterwards by sequencing, confirming at the same time that the serological data was wrong. All other discrepancies (on 12 samples) were verified by an alternative technique and the genotyping results were confirmed for 10 of the 12 samples. With these confirmations, the accuracy level was in reality 99.9% and not 99.3% as reported by the first round of analysis. The two discordant samples were, as before, the e/E and c/C-intron assays (Table 10).

Table 9: Summary results of the accuracy validation phase

			Accuracy (%)		
Markers name	Antigens	Genotype alleles	Concordance between DNA strands	SNPstream Vs. Sequenom concordance	SNPstream Vs. Sequencing concordance
c/C-rs1053344	C	G	100	100	N/A
	c	A			
c/C-Intron	C	G	98	100 (97.5)	N/A
	c	A			
e/E-rs609320	E	G	N/A	100	100
	e	C			
M/N- rs7682260	M	G	100	100	100
	N	A			
s/S-rs7683365	S	A	100	100	100
	s	G			
k/K-rs8176058	K	A	100	100	100
	k	G			
Kp a/b-rs8176059	Kp a	A	100	100	100
	Kp b	G			
Fy a/b-rs12075	Fy a	G	100	100	100
	Fy b	A			
Jk a/b-rs1058396	Jk a	G	100	100	100
	Jk b	A			
HPA1 a/b-rs5918	HPA-1a	A	100	100	100
	HPA-1b	G			
HPA2 a/b-rs6065	HPA-2a	G	100	100	100
	HPA-2b	A			
HPA5 a/b-rs10471371	HPA-5a	G	100	100	100
	HPA-5b	A			
Average			**99.8**	**100 (99.8)**	**100**

Summary of accuracy results obtained after the genotyping of 96 Coriell DNAs and 40 DNAs extracted from 2 mm FTA cards. The concordance between DNA strands represents the concordance between genotyping results obtained from the GA and CT extension pools. The SNPstream Vs. Sequenom concordance corresponds to the average genotyping concordance from both DNA strands between the SNPstream panel version 2 and the Sequenom panel. The percentage in brackets for the SNPstream Vs. Sequenom call concordance referred only to the DNAs from FTA cards. The concordance between the genotyping and the sequencing was based on six selected samples (NA17175, NA17275, NA17223, NA17284, NA17294 and NA19007).

Table 10: Inferred phenotypes from the genotyping results and correlation with the serological phenotypes

			Accuracy (%)				
Markers name	Antigens	Genotype alleles	Call rate	Concordance between DNA strands	Reproducibility	Concordance between serological and Genotyping phenotypes	Number of samples*
c/C-rs1053344	C	G	100	100	100		131
	c	A				100	323
c/C-Intron	C	G	99.8	99.8	97	99.2	131
	c	A					323
e/E-rs609320	E	G	99.2	N/A	97	99.5	196
	e	C			100	99.5	181
M/N- rs7682260	M	G	100	100	100	98.7	76
	N	A			100	100	26
s/S-rs7683365	S	A	100	100	100	100	70
	s	G			100	98.5	66
k/K-rs8176058	K	A	100	100	100	100	22
	k	G			100	100	344
Kp a/b-rs8176059	Kp a	A	100	100	100	100	26
	Kp b	G			100	N/D	0
Fy a/b-rs12075	Fy a	G	100	100	100	99.2	130
	Fy b	A			100	97.3	75
Jk a/b-rs1058396	Jk a	G	100	100	100	98.6	146
	Jk b	A			100	98.9	97
HPA1 a/b-rs5918	HPA-1a	A	100	100	100	100	30
	HPA-1b	G			100	50	6
HPA2 a/b-rs6065	HPA-2a	G	100	100	100	N/D	0
	HPA-2b	A			100	N/D	0
HPA5 a/b-rs10471371	HPA-5a	G	100	100	100	100	9
	HPA-5b	A			100	100	9
Average			99.9	99.9	99.7	99.3	1963

* **Number of samples that serological data was available**

Summary of the accuracy results obtained after genotyping of 618 DNAs extracted from regular blood donors. The call rate is an average of both DNA strands call rates and the concordance between DNA strands represents the concordance between genotyping results obtained from the GA and CT extension pools. The reproducibility was evaluated based on the 32 duplicate samples present within the 618 test DNAs. The concordance for the phenotypes was evaluated based upon the phenotypes provided by Héma-Québec (serological) and the inferred phenotypes from the genotyping results.

Figure 22: Genotyping Vs. Sequencing results for the antigen M/N

A) Typical example of genotyping and sequencing results concordance from sample NA17294 for the marker M/N. The left genotyping (1) cluster shows the typical SNPstream genotyping results obtained during the validation phase with Coriell and FTA samples. The blue circle shows the location of the sample NA17294. Similarly, the right genotyping cluster (2) shows typical Sequenom results with the NA17294 sample in the heterozygote cluster. The sequencing trace represents 30 bases of DNA sequence downstream of the M/N variants, and the cursor indicates the SNP position.

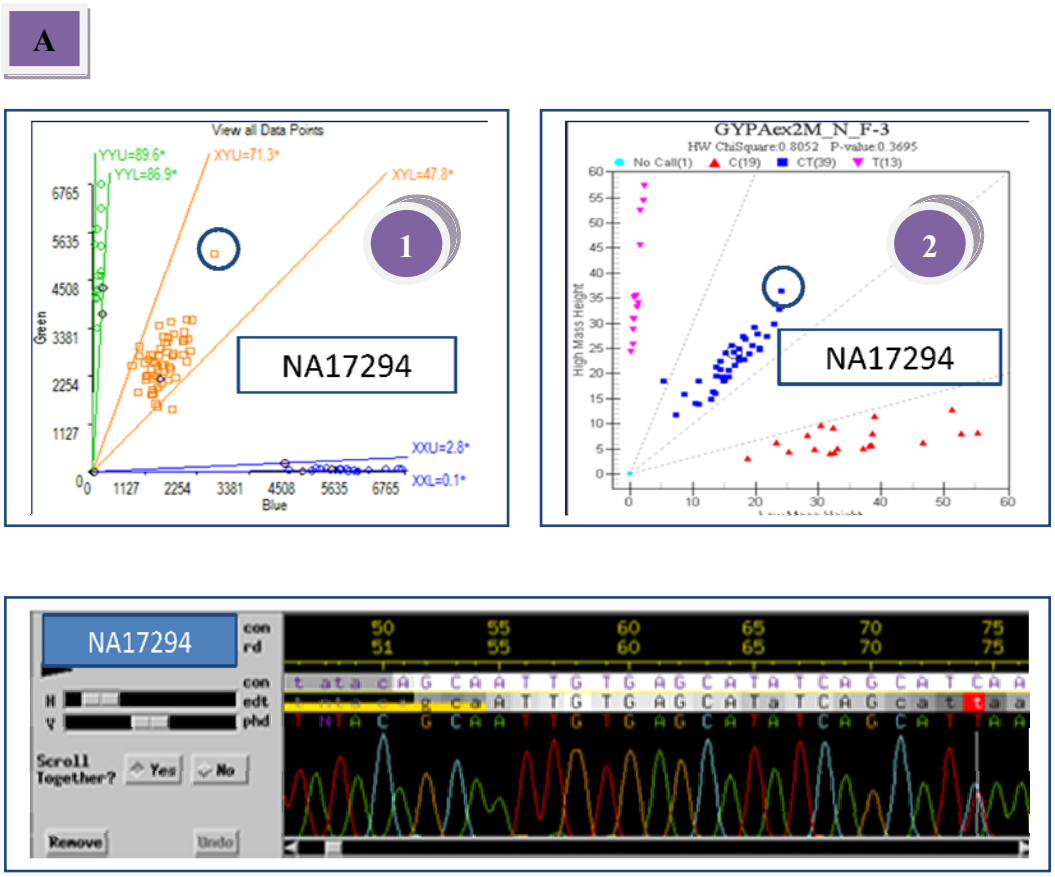
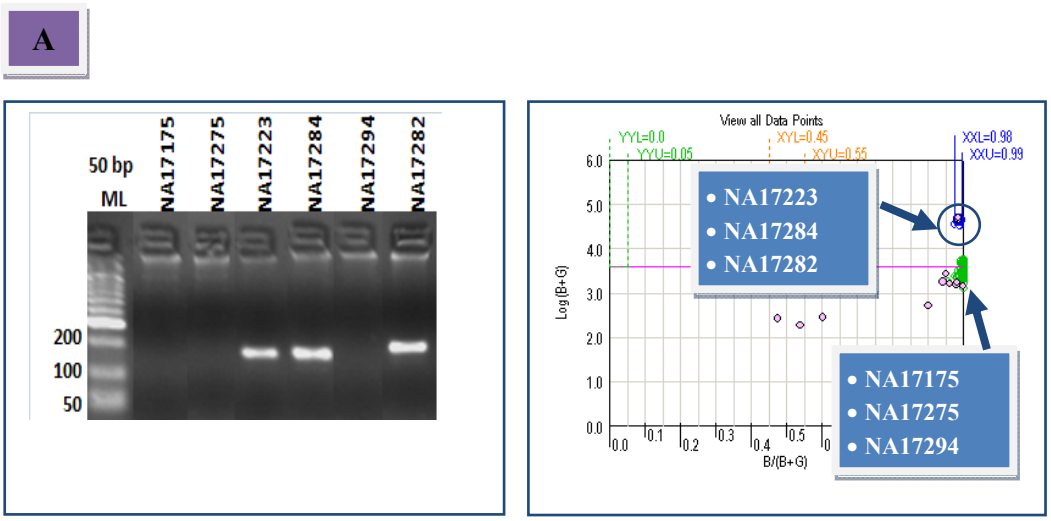


Figure 23: Genotyping Vs. agarose gel results of the marker c/C-intron

A) Typical agarose 4% gel results after the amplification of the c/C-intron. The gel results demonstrate the evaluation of six selected samples for the presence or absence of a band that correlates with the 109 bp insertion (amplicon size = 155bp). The presence of the band indicates the presence of the *RHCE_C* allele (c/C or C/C) and the absence signifies the presence of *RHCE_c* allele (c/c only). The blue cluster in the genotyping graph represents the presence of the insertion whereas the green cluster represents the absence of *RHCE_C*.



2.2. Robustness evaluation

The assays were evaluated for robustness based on the reproducibility of results at different time points, and between technicians. The day-to-day data reproducibility was conducted on 176 FTA and 12 Coriell DNAs samples (all having known genotypes) using the GA and CT extensions pools reactions. In total the same samples were genotyped three times over three consecutive days leading to a total of 13536 different genotyping calls for comparison. The average call rate obtained per experiment was 99.6% and only three discordant genotyping calls were observed (0.07%). 100% of the discordant calls were from the marker e/E-rs609320 and exclusive to FTA samples (Table 11). These results are in direct line with the results shown earlier where the accuracy and the intra-assay reproducibility results of e/E were the worst of the twelve markers tested.

The reproducibility of results between technicians was also very good with only one discordant result from 1,152 total genotyping calls, leading to a discrepancy rate of only 0.09%. For 96 FTA samples (1,152 genotyping calls), only 0.9% of genotype calls were failed and therefore rejected from the comparison study. The identified discrepancy was for the marker c/C-intron (Table 11). Both reproducibility measures revealed that the SNPstream panel version 2 was very robust for all markers, but the results also demonstrated that the two non-standard makers were the most susceptible to generating unreliable results. However, the high percentage of concordance obtained following the accuracy and robustness experiments was sufficient to convince Héma-Québec to move forward on a proof of concept study to integrate the SNPstream minor blood group antigen genotyping panel into their clinical operations.

Table 11: Summary results of the assay robustness validation phase

			Robustness (%)		
Markers name	Antigens	Genotype alleles	Mean genotyping Call rate of both DNA directions results	Genotyping calls reproducibility over three experiments	Genotyping concordance from two technicians results
c/C-rs1053344	C	G	100	100	100
	c	A			
c/C-Intron	C	G	99.8	100	98.9
	c	A			
e/E-rs609320	E	G	99.2	98.5	100
	e	C			
M/N- rs7682260	M	G	100	100	100
	N	A			
s/S-rs7683365	S	A	100	100	100
	s	G			
k/K-rs8176058	K	A	100	100	100
	k	G			
Kp a/b-rs8176059	Kp a	A	100	100	100
	Kp b	G			
Fy a/b-rs12075	Fy a	G	100	100	100
	Fy b	A			
Jk a/b-rs1058396	Jk a	G	100	100	100
	Jk b	A			
HPA1 a/b-rs5918	HPA-1a	A	100	100	100
	HPA-1b	G			
HPA2 a/b-rs6065	HPA-2a	G	100	100	100
	HPA-2b	A			
HPA5 a/b-rs10471371	HPA-5a	G	100	100	100
	HPA-5b	A			
Average			**99.9**	**99.9**	**99.9**

The day-to-day reproducibility results were obtained on three consecutive days by the same technician using the SNPstream version 2 panel and 188 total samples (176 FTA samples and 12 Coriells samples). The inter-technician reproducibility results were obtained on two different days using 96 different FTA samples.

3. CLINICAL APPLICATION: PROOF OF CONCEPT STUDY

In order to accurately genotype 22,000 samples over a period of one and a half years, the genotyping blood antigen panel had to be incorporated into standard and controlled process. Based upon the genotyping panel development, optimisation and validation processes, detailed SOPs (Standard Operational Procedure) have been developed and currently applied to help prepare the genotyping panel for clinical application (see Annexes A and B). Also, the SOPs have been adapted into our LIMS (Laboratory Information Management System) to ensure complete traceability from the sample reception to the results reporting to Héma-Québec. Essentially, as described in Figure 11 of the Materials and Methods, the complete turnaround process for a test sample encompasses five major steps: blood collection on FTA cards by the Héma-Québec nurses, the management of sample reception, DNA plate management and extraction, the genotyping of extracted DNAs, and finally results quality control (QC) and reporting. For each of these steps, a controlled process has been established to minimize the human error and make the procedure more automated and as accurate as possible (Figure 24).

3.1. Clinical genotyping process and key results

As illustrated in Figure 24-1, to ensure complete sample traceability, barcodes were generated from our LIMS and shipped to Héma-Québec in order to uniquely identify each FTA card sample with our unique identifiers. In addition to our barcodes, Héma-Québec also identified each sample with their own unique identifiers.

Samples are received in batches of 500 at the Pharmacogenomics Centre. The sample list provided by Héma-Québec is downloaded from our secure share website folder and pasted into the “sample checker” tool (Figure 12-A; Materials and Methods). To match the FTA samples against the sample list sent by Héma-Québec,

the Héma-Québec barcode is scanned first and immediately a notification of "Passed" or "Failed" is seen based upon whether the sample matches the original submitted list. The LIMS barcode is then scanned to make the final link between the Héma-Québec and LIMS identifiers (Figure 12-B). The final file containing all sample related information is created only if the previous step provides 100% "Passed" comments and is then uploaded into the LIMS (Figure 12-C). Afterwards, the LIMS associates both identifiers in their specific fields and all sample related information is saved into the LIMS database (Figure 25; samples screen).

As presented in detail in the sections 9.3 and 9.4 of the annexe A (Figure 24-2), the FTA cards are submitted to the auto puncher (BSD, Brisbane, Australia, Duet 600) to create the 2 mm spots that are required for the automatic extraction on the BiomekFX (Beckman Coulter, Fullerton, CA). The extraction is mediated by a LIMS procedure transforming the FTA sample type to a DNA sample type (Figures 26-A LIMS procedure and B DNA plate layout). A visual quality control check is performed after the automatic punching and after the extraction steps, respectively, to ensure that all wells contain only one 2 mm spot and also that all wells contain the expected volume of DNA at the end of the extraction process. At the time of writing, 16,598 samples have been punched and no problems have been documented regarding this step, however approximately 1% of the FTA cards (5 samples out of every 500 samples) are randomly removed from the well by the BiomekFx tips during the extraction procedure. The lost of the 2 mm FTA disk during the extraction procedure steps does not automatically lead to a failed genotype unless the loss happens before the second wash, suggesting that DNA is slowly released from the disk during the final stages of the extraction procedure.

To accurately control the genotyping processes, all steps described in the Annexe B are also supported by the LIMS (Figure 24-3). The DNA from two 96 well plates is transferred in duplicate (to support forward and reverse assays) to one 384-well assay plate for PCR amplification. This is then followed by the LIMS genotyping protocol (Figure 27-A). After the robotic transfer of DNA to the assay plate, a visual QC check is performed to make sure that the volume is uniformly

transferred across the plate, and all mis-transferred samples are noted in the genotyping LIMS run. Mis-transfer problems happen rarely and to date less than 0.1% of samples (~160 samples) have experienced transferred issues. When these do occur, they can be easily corrected by a manual transfer. Furthermore, to evaluate the multiplex PCR amplification efficacy, a 1 µL aliquot of twenty-four samples located in key positions of the assay plate (384 total PCR reactions) are loaded onto a 4% agarose gel to act as part of our quality control process. Different actions are taken for failed samples ranging from the re-extraction of test samples to re-starting the multiplex amplification with another master mix (Annex B, section 8.5). With more than 95 experiments involving 176 samples having been performed so far, only one experiment has been rejected because more than 4 samples failed the PCR amplification (Figure 27-B).

To accurately interpret the genotyping results obtained from the SNPstream scanner, the raw results are transferred to our in-house quality software tool called "SHERPA" to perform the genotype clustering (Figure 28-A) and evaluate the overall experiment quality (Figure 24-4). The SHERPA software assists in reporting quality issues in two ways, the first one reports the issues at the assay plate level, where marker conversion and failure rates are graphically highlighted for better visual interpretation of the general experiment performance, and the second one reports the issues at the genotyping level, giving a good comprehension of the accuracy of the results. As demonstrated in Figure 28-B, a clear issue with the entire row K is captured by the software where only 10 markers have generated satisfactory genotyping results from a possible 12 markers. The highlighted failed samples were subsequently linked to an array spotting problem on the SNPstream hybridization plate that encompassed the markers Fy a/b and Jk a/b. Also, the heat map gives a quick representation of possible plate inversion or sample layout issues since for the blank samples O9 (P9), O10 (P10), O11 (P11) and O12 (P12) a null genotyping result is expected.

Multiple other issues have been flagged by the SHERPA assay plate QC tool, for example, an inadequate plate sealing technique leading to a high percentage of failed

samples located at the plate extremities was identified and has since been resolved. Furthermore, the SHERPA software also assists in the evaluation of the accuracy of the results by interpreting three levels of concordance: the concordance of results from both DNA strands (Figure 28-C), the concordance of results from two independent technicians (Figure 28-D), and the concordance between the actual and expected results of known control DNAs (Figure 28-E). The possibility to rapidly check for discordances in our results has allowed the technicians to quickly identify and to reanalyze problematic samples and markers and correct the genotypes accordingly. For instance, sometimes the discordances from both DNA strands are generated by a contaminant fixed on the glass of the SNPstream array plate, leading to a bias in the fluorescence intensities and ultimately creating a call mismatch. In addition to the negative heat map results given by blank samples, the concordance between the genotypes of known controls samples obtained during an experiment with the expected genotyping calls is used to flag plate inversion problems. Such problems have not been seen in this study to date. The known control genotypes are also used to evaluate the stability of the multiplex PCR and extension reagents batches over time. So far, no observable variation within the same reagent batches has been observed (more than 99.9% of concordance was observed) and the mismatches are not related to the age of the reagent aliquots. SHERPA also allows for the double reviewing of the raw genotyping data by two independent technicians. This greatly increases our ability to accurately call genotypes. Results of genotyping mismatches between technicians are rare and are easily identified and resolved by the more experienced technician. On averages only four samples per assay plate (~1%) have cross-technician genotyping mismatches. Mismatches are most often observed for the markers c/C-intron and e/E. Since, the clustering of these two non-standard markers requires technical interpretation, a manual interpretation is needed and therefore these generated the highest discordant rates. After reviewing and correcting all cross-technician discordances, the genotyping results are ultimately saved as final call batches.

Based upon the predefined criteria described in Tables 1 to 4 (Annex B, appendix 8), the software called “Area51” analyses the final genotypes saved into the SHERPA

database and translated the interpreted results into three comprehensive reports: technical, inferred phenotypes for tested samples, and concordance with known controls. If for any test sample, the rules described in Table 1 (Annex B, appendix 8) failed, the sample was automatically removed from the sample list, added to the failed sample section of the phenotype report, and flagged for repeat (Figure 29-A). Based on these criteria, only 382 samples out of 16,598 samples (2.3%) have been failed and subjected to a second genotyping experiment. More than 90% of all failed samples are associated with "> 1 discordant genotype from DNA strand assays" criteria, and 95% of them are associated with the c/C-intron marker. The other 5% are associated with the criteria E "Low intensity signals or samples outside of normal clustering thresholds". These spurious samples were re-genotyped and 100% of the repeated samples successfully passed the Table 1 criteria.

The criteria of Tables 3 and 4 (Annex B, appendix 8) are simultaneously analysed. If one of these criteria fails an automatic warning message is highlighted in the technical report (Figure 29-B). After analysing more than 95 assay plates using the above rejection criteria, only two have been rejected to date. The experimental failures were in both cases attributable to mismatching errors in known controls. The problem was tracked to an evaporation issue where samples in row P led a to a high failure rate of control samples.

Ultimately, all samples that passed all of the quality control criteria were translated into phenotypes and reported according to HQ clinical requirements as "+ and –" reports (Figure 29-C). The translation from genotypes to + and – signs is performed electronically following the conversion Table 12, where basically each genotype allele is associated with a specific antigen. For instance, if the genotype is heterozygous for Fy a/b markers, the inferred phenotypes are (+) and (+) for the antigens Fy a and b respectively, and if the genotype is homozygous for the major allele, the inferred phenotypes are Fy a (+) and b (-) (therefore the sample is only carrying the a antigen and not the b).

3.2. Clinical results: accuracy and robustness

To date, 16,598 samples have been genotyped for a total 398,352 genotype calls and 365,165 inferred phenotypes. Of that number, 9.7% (35,275) of inferred phenotypes were already known by Héma-Québec, but the Centre was blinded to the results. These known phenotypes were used as a comparison to check the accuracy of the genotyping experiments in production. So far, only 157 discordances were observed (0.45%), and most of them are related to the antigens c/C (9), e/E (69) and Fy b (57) (Table 13). All discordant results have been submitted to a confirmation test to assess if either the serological or the genotyping results generated the discrepancies. The validation tests confirmed the initial genotyping results for eighty-six phenotypes (55%), whereas the test confirmed the initial serological results for seventy-one phenotypes (45%). As previously described in the genotyping validation phases, the antigen C, inferred from the non-standard c/C-intron assay, and e/E gave the worst initial genotyping confirmation results, where 9 out of 78 total c/C and e/E related discordant results were observed. The discordances generated from the non-standard assays c/C-intron and e/E explained 97.2% (69 out of 71) of all initial discrepancies with the serological results. Furthermore, the antigen Fy b also showed discordances, but in this case all of the discordances were validated for the initial genotyping results. These discordances were fully explainable by two silencing markers that are not present in either the initial serological results or the genotyping data but are essential for the accurate interpretation of the final Fy b phenotype. In addition, discordant results have been observed for antigens M, s/S, K, Fy a and Jk a/b. All the discrepancies have been resolved upon re-sequencing to show that the original genotyping results were correct, except for one Jk b result and one Fy a result. Moreover for accuracy evaluation, Héma-Québec wanted to evaluate the stability of the blood at 4°C before the transfer to the FTA cards, and the stability of FTA cards once they were spotted with bloods. To do so, they kept ten samples from blood donors in the refrigerator and transferred the blood in triplicate onto FTA cards after 6, 27 and 55 days. Subsequently the FTA cards were stored at room temperature for 4, 12 and 26 weeks before shipping. All samples were unknown to the Centre and

distributed in different shipping boxes and consequently dispersed in different genotyping runs. All inferred phenotypes from all replicated samples were compared for inter-condition robustness with 100% concordance and reproducibility (Table 13).

Figure 24: Clinical process flow summary and QC check points

A) A detailed process flow has been established to accurately genotype 21,000 Héma-Québec blood donors that have agreed to participate in the creation of a database for minor blood group antigens. The whole process involves four major steps: 1- sample reception, 2- FTA sample spotting and DNA extraction, 3- multiplex PCR amplification and 4- genotyping, analysis and results reporting. Each of these steps has detailed quality control steps in order to reduce errors and ensure the quality of results.

A

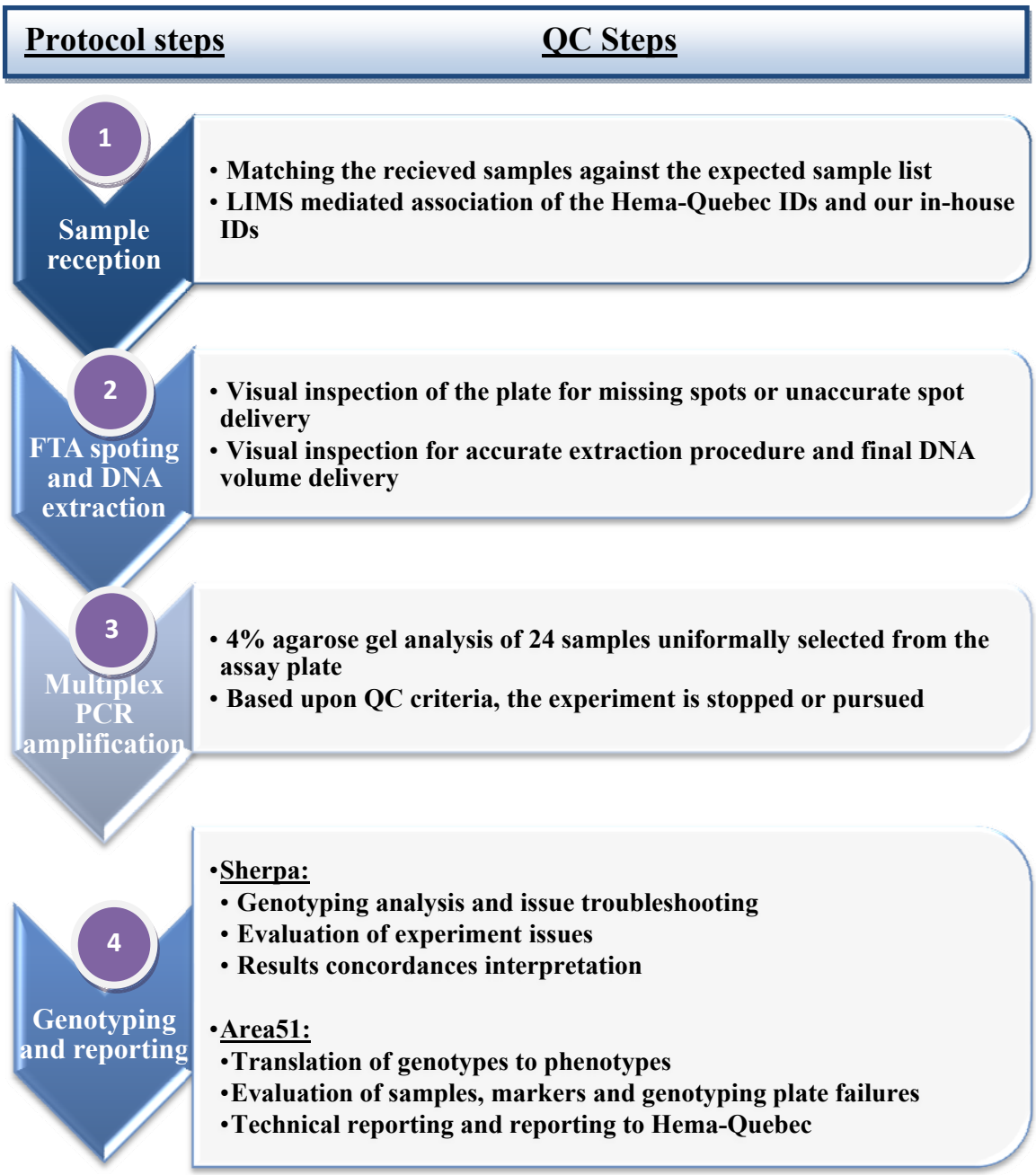

Protocol steps
QC Steps
1
Sample reception
• Matching the recieved samples against the expected sample list
• LIMS mediated association of the Hema-Quebec IDs and our in-house IDs
2
FTA spoting and DNA extraction
• Visual inspection of the plate for missing spots or unaccurate spot delivery
• Visual inspection for accurate extraction procedure and final DNA volume delivery
3
Multiplex PCR amplification
• 4% agarose gel analysis of 24 samples uniformally selected from the assay plate
• Based upon QC criteria, the experiment is stopped or pursued
4
Genotyping and reporting
•Sherpa:
• Genotyping analysis and issue troubleshooting
• Evaluation of experiment issues
• Results concordances interpretation
•Area51:
•Translation of genotypes to phenotypes
•Evaluation of samples, markers and genotyping plate failures
•Technical reporting and reporting to Hema-Quebec

Figure 25: Sample reception and LIMS management

A) Typical view of the information from the received FTA samples after they were uploaded into our Laboratory Information Management System (LIMS). 1- Shows the side-by-side FTA IDs provided by the Héma-Québec (HQ, Client Sample ID) and the FTA IDs generated by our LIMS (PGx name). 2- Represents the most important related information for each sample, such as the date of sampling, reception date and its storage box identifier. 3- Describes specific information related to the samples, such as the storage location, the sample quantity and concentration. Also, handling and transfer information is annotated once the samples have been moved to different labware (i.e. new plate association) or how the samples have been manipulated in the lab (i.e. information about DNA extraction, genotyping or sequencing run (Usage)).

A

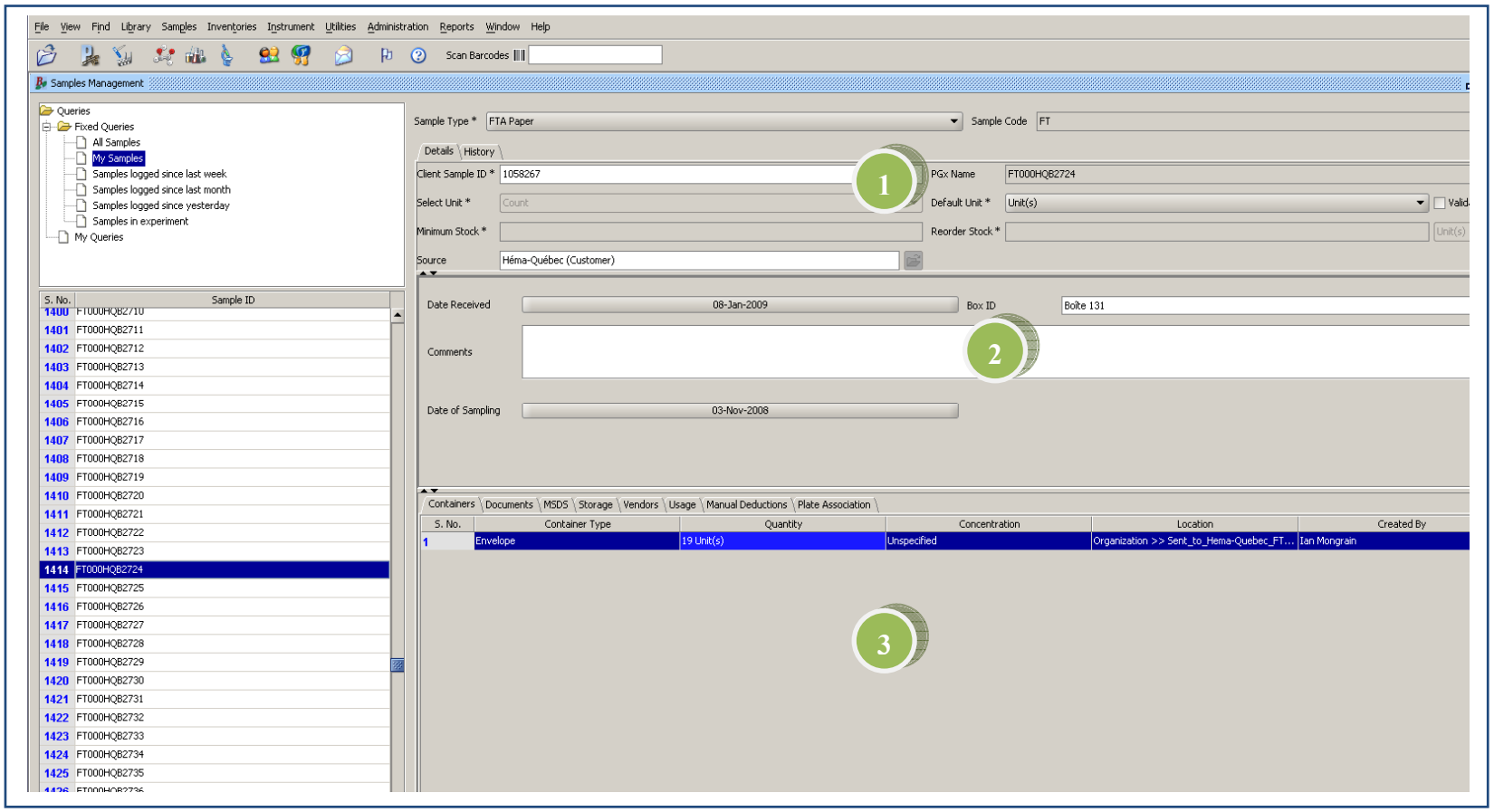
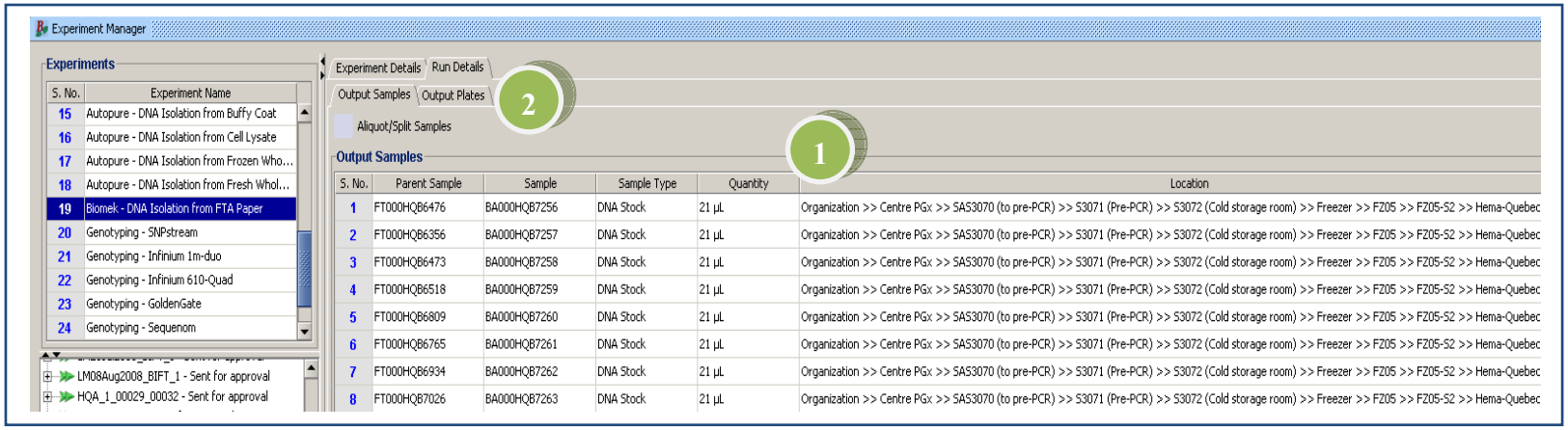

File View Find Library Samples Inventories Instrument Utilities Administration Reports Window Help
Scan Barcodes
Samples Management
Queries
Fixed Queries
All Samples
My Samples
Samples logged since last week
Samples logged since last month
Samples logged since yesterday
Samples in experiment
My Queries
S. No.
Sample ID
1401 FT000HQB2711
1402 FT000HQB2712
1403 FT000HQB2713
1404 FT000HQB2714
1405 FT000HQB2715
1406 FT000HQB2716
1407 FT000HQB2717
1408 FT000HQB2718
1409 FT000HQB2719
1410 FT000HQB2720
1411 FT000HQB2721
1412 FT000HQB2722
1413 FT000HQB2723
1414 FT000HQB2724
1415 FT000HQB2725
1416 FT000HQB2726
1417 FT000HQB2727
1418 FT000HQB2728
1419 FT000HQB2729
1420 FT000HQB2730
1421 FT000HQB2731
1422 FT000HQB2732
1423 FT000HQB2733
1424 FT000HQB2734
1425 FT000HQB2735
Sample Type * FTA Paper
Sample Code FT
Details History
Client Sample ID * 1058267
PGx Name FT000HQB2724
Select Unit * Count
Default Unit * Unit(s)
Minimum Stock *
Reorder Stock *
Source Héma-Québec (Customer)
Date Received 08-Jan-2009
Box ID Boîte 131
Comments
Date of Sampling 03-Nov-2008
Containers Documents MSDS Storage Vendors Usage Manual Deductions Plate Association
S. No. Container Type Quantity Concentration Location Created By
1 Envelope 19 Unit(s) Unspecified Organization >> Sent_to_Hema-Quebec_FT... Ian Mongrain
1
2
3

Figure 26: LIMS mediated DNA extraction protocol

A) View of the typical output sample file obtained after performing the DNA extraction LIMS protocol. 1- Shows the input samples (Parent sample), its related output samples (Sample) and the related output sample type, quantity and location. 2- Shows the output plate layout

B) Typical view of the output plate obtained at the end of the DNA extraction protocol. The final plate name and related information, such as the future storage location and storage container type are indicated in point 1. For the plate name, the nomenclature (HQB_1_00095) represents the project (HQB=Héma-Québec), the nature of samples contained in the plate (1 = DNA) and the actual plate number (DNA plate 95). Point 2 represents the plate layout at the wells level (wells details) and the full list at the sample level (sample). All future usage of the plate for either genotyping or sequencing will be documented in the tab (usage).

A


Experiment Manager
Experiments
S. No.
Experiment Name
15 Autopure - DNA Isolation from Buffy Coat
16 Autopure - DNA Isolation from Cell Lysate
17 Autopure - DNA Isolation from Frozen Who...
18 Autopure - DNA Isolation from Fresh Whol...
19 Biomek - DNA Isolation from FTA Paper
20 Genotyping - SNPstream
21 Genotyping - Infinium 1m-duo
22 Genotyping - Infinium 610-Quad
23 Genotyping - GoldenGate
24 Genotyping - Sequenom
LM08Aug2008_BIFT_1 - Sent for approval
HQA_1_00029_00032 - Sent for approval
Experiment Details
Run Details
Output Samples
Output Plates
Aliquot/Split Samples
Output Samples
S. No. | Parent Sample | Sample | Sample Type | Quantity | Location
1 FT000HQB6476 BA000HQB7256 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
2 FT000HQB6356 BA000HQB7257 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
3 FT000HQB6473 BA000HQB7258 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
4 FT000HQB6518 BA000HQB7259 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
5 FT000HQB6809 BA000HQB7260 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
6 FT000HQB6765 BA000HQB7261 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
7 FT000HQB6934 BA000HQB7262 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
8 FT000HQB7026 BA000HQB7263 DNA Stock 21 µL Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
1
2

B

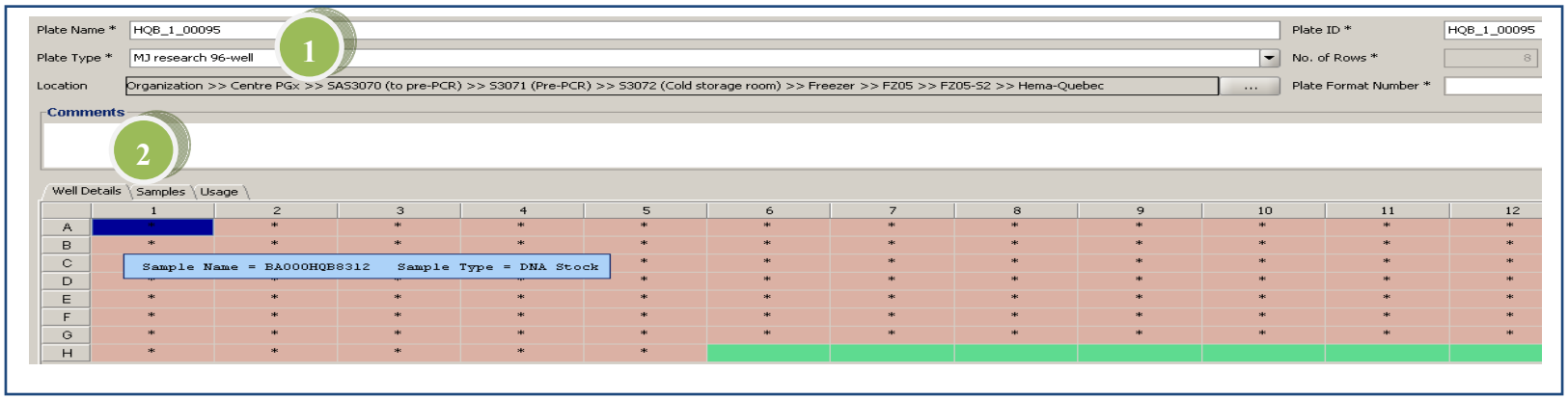


Plate Name * HQB_1_00095
Plate Type * MJ research 96-well
Location Organization >> Centre PGx >> SAS3070 (to pre-PCR) >> S3071 (Pre-PCR) >> S3072 (Cold storage room) >> Freezer >> FZ05 >> FZ05-S2 >> Hema-Quebec
Plate ID * HQB_1_00095
No. of Rows *
Plate Format Number *
Comments
Well Details
Samples
Usage
1 2 3 4 5 6 7 8 9 10 11 12
A B C D E F G H
Sample Name = BA000HQB8312 Sample Type = DNA Stock
1
2

Figure 27: LIMS mediated genotyping protocol

A) Typical view of the output plate obtained at the end of the DNA transfer step. The final plate name and related information, such as the future storage location and storage container type are indicated in point 1. The HQB_2_00095_00096 located in the plate name field represents the project (HQB=Héma-Québec), the PCR plate (_2 = PCR plate) and DNA plates IDs encompassed into the assay plate (00095_00096). Point 2 represents the plate layout at the well level (wells details) and the full list at the sample level (sample). The plate usage is ultimately described in the tab "usage".

B) Represents the typical 4% agarose gel QC results obtained after multiplex PCR. Typically only wells 10 and 12 are expected to have null PCR amplification. Following the criteria described in Table 1 of the annex B, the experiment shown here was rejected since more than four samples had failed the PCR amplification (wells; 3, 6, 15, 17 and 19).

A

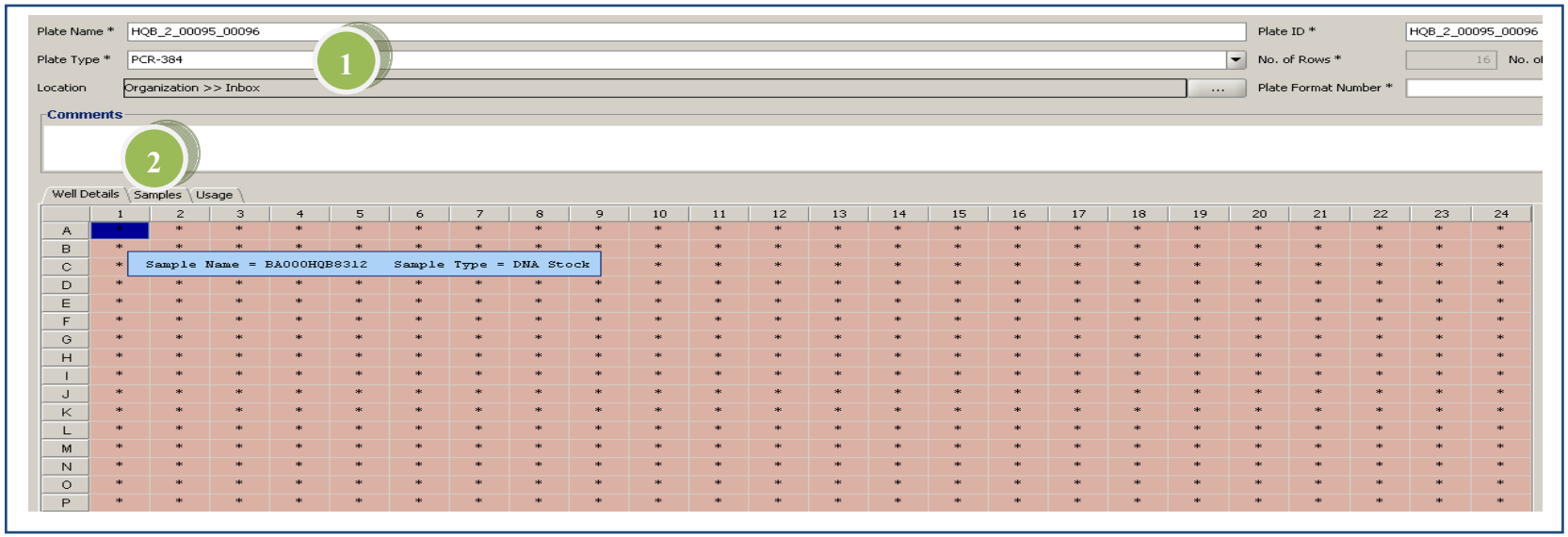


Plate Name * HQB_2_00095_00096
Plate ID * HQB_2_00095_00096
Plate Type * PCR-384
No. of Rows * 16
Location Organization >> Inbox
Plate Format Number *
Comments
1
2
Well Details
Samples
Usage
Sample Name = BA000HQB8312 Sample Type = DNA Stock

B

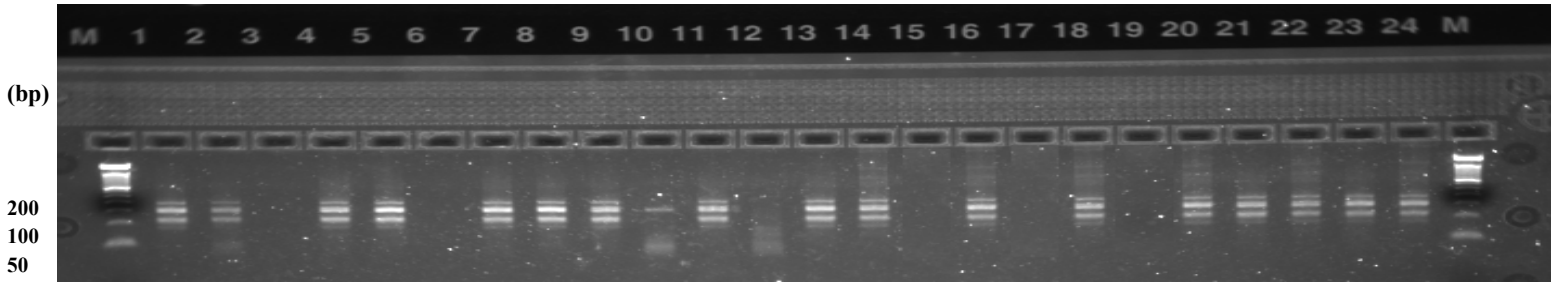
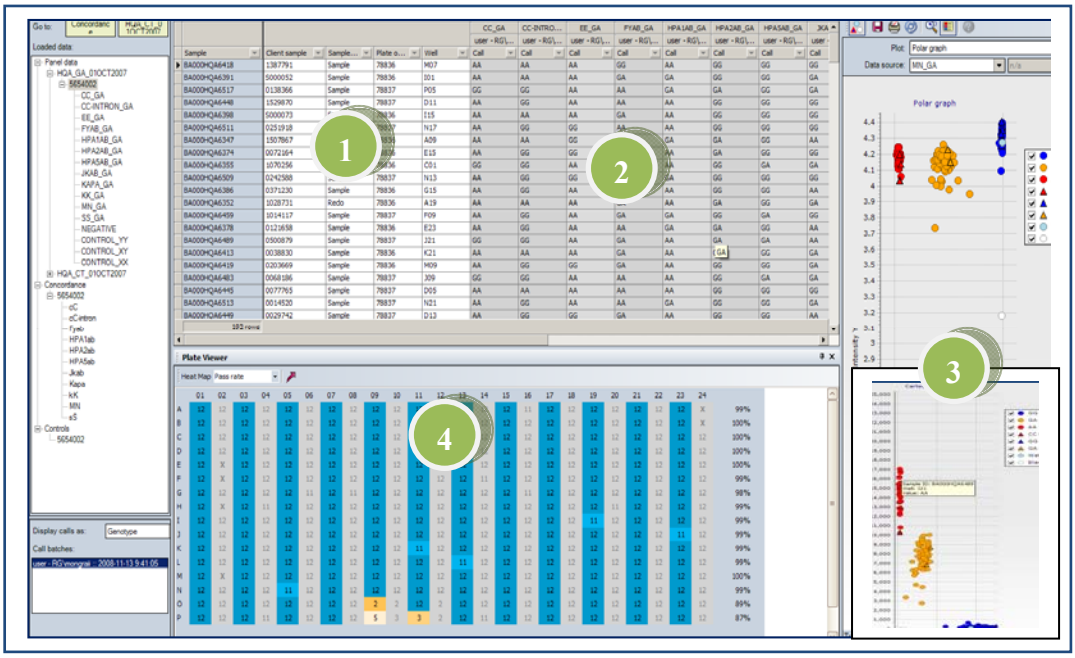

(bp)
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 M
200
100
50

Figure 28: Genotyping quality control assessment using the in-house SHERPA software

A) Overall illustration of our in-house developed quality assessment software tool, "SHERPA". 1- Represents the sample and the assay plate IDs generated by our LIMS, along with the client ID, sample types and the well positions. 2- Demonstrates the tabular representation of all genotype calls per tested antigen marker. 3- Shows the visual representation of genotype calls using Polar and Cartesian graphic plots. 4- Represents the overall performance of genotyping for the entire detection plate. Basic statistics are performed to evaluate the percentage of calls per row, per column and per well. These statistics are used to identify failure patterns that occur during the genotyping procedure.

B) Close-up of the SHERPA heat map view after the raw data upload from the SNPstream database. The highlighted columns 01, 03, 05... represent the GA extension SNPstream pool whereas the columns 02, 04, 06... represent the CT extension pool. The number represented in each well shows the actual number of markers that are within the passed genotyping thresholds, for instance, 11 means that one out of twelve markers is considered failed. The percentage located on each row and column demonstrates the overall ratio of failed makers over the successful markers and the arrows indicate that an investigation is required to elucidate possible failure patterns. The wells O9 to O12 and P09 to P12 represent the blank samples. 1- Shows clear failure pattern that has occurred in row K and the highlighted failed samples are represented in the genotyping clusters 2- Fy a/b and 3- Jk a/b.

A


1
2
3
4
Plate Viewer
Heat Map Pass rate
Polar graph

B

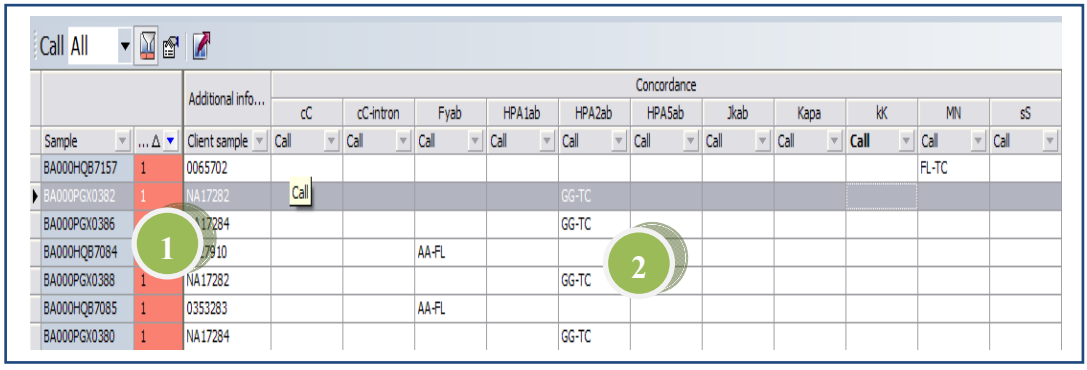

Plate Viewer
Heat Map Pass rate
1
2
3
Intensity Y

C) Illustration of the module that evaluates the concordance between both DNA strands. 1- Shows the samples with DNA strand results discordances. 2- Shows the actual call discordances from the GA and CT extension pools.

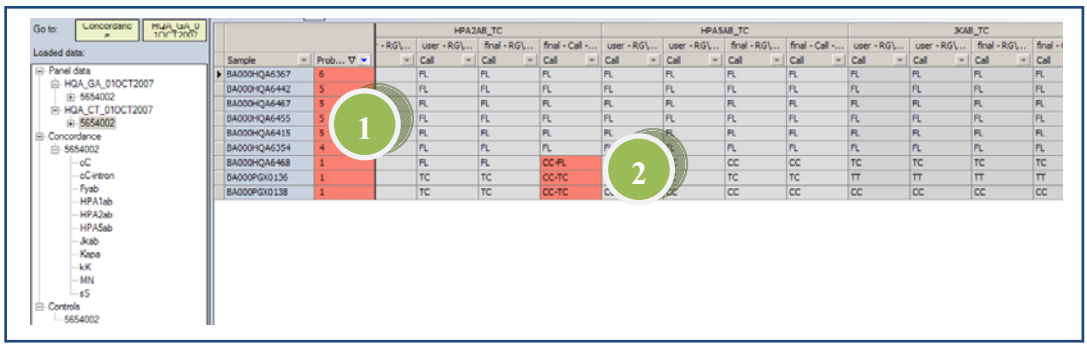
D) Illustration of the module that evaluates the cross-technician concordance results. 1- Highlights the samples that experience cross-technician genotyping call discordances. 2- The red highlighted samples show a typical example of final genotyping call discordances.

E) Representation of the module that evaluates the known control concordance between actual genotyping calls and the expected results. 1- Highlights the samples and number of markers with miss-matched genotyping results. 2- The red highlighted samples show a typical example of discordances of actual genotypes (***XX***-*XY*) compared to the expected results (*XX*-***XY***).

C


Call All
Additional info...
Concordance
cC
cC-intron
Fyab
HPA1ab
HPA2ab
HPA5ab
Jkab
Kapa
kK
MN
sS
Sample
Client sample
Call
BA000HQB7157 1 0065702 FL-TC
BA000PGX0382 1 NA17282 GG-TC
BA000PGX0386 1 17284 GG-TC
BA000HQB7084 1 7910 AA-FL
BA000PGX0388 1 NA17282 GG-TC
BA000HQB7085 1 0353283 AA-FL
BA000PGX0380 1 NA17284 GG-TC
1
2

D


Loaded data:
Panel data
HQA_GA_01OCT2007
5654002
HQA_CT_01OCT2007
5654002
Concordance
5654002
cC
cC-intron
Fyab
HPA1ab
HPA2ab
HPA5ab
Jkab
Kapa
kK
MN
sS
Controls
5654002
HPA2AB_TC
HPA5AB_TC
JKAB_TC
Sample
BA000HQA6367 6
BA000HQA6442 5
BA000HQA6467 5
BA000HQA6455 5
BA000HQA6415 5
BA000HQA6354 4
BA000HQA6468 1 CC-FL
BA000PGX0136 1 TC-TC
BA000PGX0138 1 TC-TC
1
2

E

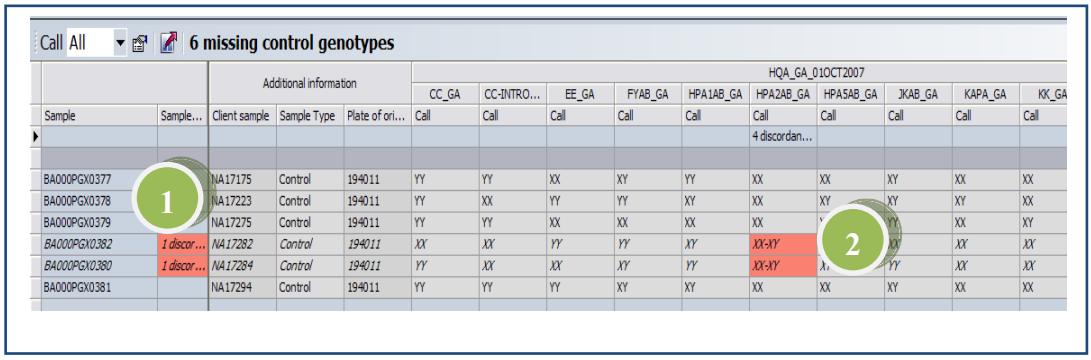

Call All
6 missing control genotypes
Additional information
HQA_GA_01OCT2007
CC_GA
CC-INTRO...
EE_GA
FYAB_GA
HPA1AB_GA
HPA2AB_GA
HPA5AB_GA
JKAB_GA
KAPA_GA
KK_GA
Sample
Client sample
Sample Type
Plate of ori...
4 discordan...
BA000PGX0377 NA17175 Control 194011 YY YY XX XY YY XX XX XY XX XX
BA000PGX0378 NA17223 Control 194011 YY XX YY YY XY XX XY XY XY XX
BA000PGX0379 NA17275 Control 194011 YY YY XX XX XX XX XY XX XY
BA000PGX0382 1 discor... NA17282 Control 194011 XX XX YY YY XY XX-XY XX XX XX
BA000PGX0380 1 discor... NA17284 Control 194011 YY XX XX XY YY XX-XY XX YY XX XX
BA000PGX0381 NA17294 Control 194011 YY YY YY XY XY XX XX XY XX XX
1
2

Figure 29: Results reporting using the in-house Area51 software

A) Illustration of failed samples after a typical genotyping experiment. 1- Represents the complete list of samples that failed the criteria described in Table 1 of the Data Analysis and Data reporting section of the annex B. 2- Shows the reason for the sample failure.

B) Illustration of the technical report generated after each genotyping experiment. Three main components are included in this report, 1- the experiment status, 2- genotype discordances between the technicians 1st and 2nd analysis, and 3- the samples that failed and required further genotyping experiments. The actual report shows a genotyping failure because more than 10% of the control DNAs failed.

C) Representation of the + and - inferred phenotypes from successful sample genotypes that are reported to Héma-Québec as a final output file of results. 1- List of Héma-Québec and PGx sample IDs reported for which genotypes passed all samples, markers, and plate failure criteria. 2- Inferred phenotypes for all test markers. RH2 (C), RH3 (E), RH4 (c), RH5 (e), MNS1 (M), MNS2 (N), MNS3 (S), MNS4 (s), KEL1 (k), KEL2 (K), KEL3 (Kp a), KEL4 (Kp b), FY1 (Fy a), FY2 (Fy b), JK1 (Jk a), JK2 (Jk b), HPA-1a (a), HPA-1b, (b), HPA-2a (a), HPA-2b (b), HPA-5a (a), HPA-5b (b). The – and + signs indicate the absence and presence of the antigen respectively.

A

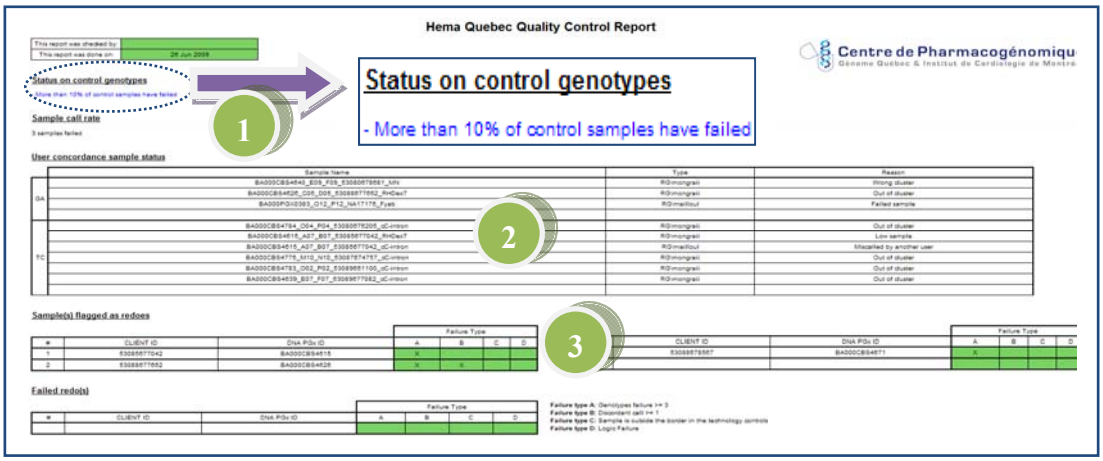
REPORTS CGR Controls Tech Report

-5835602-(RG\mongraii :: 2009-05-26 16:57:25) [HQ]

Échantillons échoué

Échantillons	PGX Id	Date du test	Raisons
0213767	BA000HQB7624	27/05/09	Plus que deux resultats FL
0420703	BA000HQB7708	27/05/09	Fyab GA et CT ne sont pas égales;kK GA et CT ne sont pas égales;Plus que deux resultats FL
1634236	BA000HQB7750	27/05/09	Plus que deux resultats FL
BLANK	BA000PGX0429	27/05/09	eE GA or CT FL;Plus que deux resultats FL
BLANK	BA000PGX0430	27/05/09	eE GA or CT FL;Plus que deux resultats FL
WATER	WT000PGX0001	27/05/09	eE GA or CT FL;Plus que deux resultats FL
WATER	WT000PGX0002	27/05/09	eE GA or CT FL;Plus que deux resultats FL

B



C

-5835502-(RG\mongraii :: 2009-05-26 16:57:25) [HQ]

			B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
				RH					MNS			Kell				Duffy		Kidd		Plaquettes					
Echantillons	PGX Id	Date du test	RH2	RH3	RH4	RH5	MNS1	MNS2	MNS3	MNS4	KEL1	KEL2	KEL3	KEL4	FY1	FY2	JK1	JK2	HPA-1a	HPA-1b	HPA-2a	HPA-2b	HPA-5a	HPA-5
0011123	BA000HQB7453	27/05/09	+	-	-	+	+	+	+	+	-	+	-	+	-	+	+	-	+	+	+	+	+	-
0017956	BA000HQB7570	27/05/09	+	-	-	+	+	+	+	-	-	+	-	+	+	-	+	+	+	-	+	-	+	-
0024306	BA000HQB7573	27/05/09	-	-	+	+	+	+	-	-	-	+	-	+	-	+	+	+	+	-	+	+	+	-
0024640	BA000HQB7509	27/05/09	+	+	+	+	+	+	-	-	-	+	-	+	+	+	-	+	+	-	+	-	+	-
0027045	BA000HQB7575	27/05/09	+	+	+	+	+	+	+	+	-	+	-	+	-	+	+	+	+	-	+	-	+	+
0029329	BA000HQB7449	27/05/09	+	-	-	+	+	-	+	+	-	+	-	+	-	+	-	+	+	+	+	+	+	-
0032882	BA000HQB7572	27/05/09	+	-	+	+	+	-	+	+	-	+	-	+	-	+	-	+	+	-	+	-	+	-

Table 12: Conversion table of the genotyping results to phenotype

						Inferred Phenotype from Genotype		
Gene	Markers name	Antigens	Hema-Quebec Phenotype Code	Genotype alleles	Allele Frequency	Homozygote Major Allele	Homozygote Minor Allele	Heterozygote
RHCE	c/C-rs1053344	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		c	RH4	A	Major	+	-	+
	c/C-Intron	C	RH2	G	Minor	-	+	+
		N/A	N/A	N/A	N/A	N/A	N/A	N/A
	e/E-rs609320	E	RH3	G	Major	+	-	+
		e	RH5	C	Minor	-	+	+
GYPA	M/N- rs7682260	M	MNS1	G	Minor	-	+	+
		N	MNS2	A	Major	+	-	+
GYPB	s/S-rs7683365	S	MNS3	A	Minor	-	+	+
		s	MSN4	G	Major	+	-	+
Kel	k/K-rs8176058	K	KEL1	A	Minor	-	+	+
		k	KEL2	G	Major	+	-	+
	Kp a/b-rs8176059	Kp a	KEL3	A	Minor	-	+	+
		Kp b	KEL4	G	Major	+	-	+
DARC	Fy a/b-rs12075	Fy a	FY1	G	Major	+	-	+
		Fy b	FY2	A	Minor	-	+	+
SLC14A1	Jk a/b-rs1058396	Jk a	JK1	G	Minor	-	+	+
		Jk b	JK2	A	Major	+	-	+
ITGB3	HPA1 a/b-rs5918	HPA-1a	HPA-1a	A	Major	+	-	+
		HPA-1b	HPA-1b	G	Minor	-	+	+
GP1BA	HPA2 a/b-rs6065	HPA-2a	HPA-2a	G	Major	+	-	+
		HPA-2b	HPA-2b	A	Minor	-	+	+
ITGA2	HPA5 a/b-rs10471371	HPA-5a	HPA-5a	G	Major	+	-	+
		HPA-5b	HPA-5b	A	Minor	-	+	+

Detailed strategy used by Area51 to translate the genotypes to phenotypes for each marker. The presence and the absence of the inferred antigens are represented with the positive sign (+) and the negative sign (-) respectively and are reported in a comprehensive table as presented in Figure 29-C.

Table 13: Inferred phenotypes accuracies and reproducibility results obtained during the clinical genotyping of blood donors.

				Accuracy (%)				Robustness (%)
Markers name	Antigens	Hema-Quebec Phenotype Code	Genotype alleles	# of available phenotypes data	# of known phenotypes discordant with the inferred phenotypes	Concordance between serological and Genotyping phenotypes	Confirmation of discordances*	Reproducibility
c/C-rs1053344	C	N/A	G	N/A	N/A	N/A	N/A	N/A
	c	RH4	A	3905	5	99.9	5 =Genotyping	100
c/C-Intron	C	RH2	G	2909	4	99.9	3 = Serological 1= Genotyping	100
	c	N/A	A	N/A	N/A	N/A	N/A	N/A
e/E-rs609320	E	RH3	G	4219	63	98.5	60 = Serological 3 = Genotyping	100
	e	RH5	C	2236	6	99.7	6= Serological	100
M/N- rs7682260	M	MNS1	G	1465	2	99.9	2= Genotyping	100
	N	MNS2	A	458	0	100.0		100
s/S-rs7683365	S	MNS3	A	2134	3	99.9	3= Genotyping	100
	s	MSN4	G	1275	1	99.9	1= Genotyping	100
k/K-rs8176058	K	KEL1	A	5939	1	100.0	1= Genotyping	100
	k	KEL2	G	199	0	100.0		100
Kp a/b-rs8176059	Kp a	KEL3	A	1117	0	100.0		100
	Kp b	KEL4	G	31	0	100.0		100
Fy a/b-rs12075	Fy a	FY1	G	2689	6	99.8	1 = Serological 5 = Genotyping	100
	Fy b	FY2	A	1694	57	96.6	57= Genotyping	100
Jk a/b-rs1058396	Jk a	JK1	G	2670	7	99.7	7= Genotyping	100
	Jk b	JK2	A	2109	2	99.9	1= Serological 1 = Genotyping	100
HPA1 a/b-rs5918	HPA-1a	HPA-1a	A	130	0	100.0		100
	HPA-1b	HPA-1b	G	36	0	100.0		100
HPA2 a/b-rs6065	HPA-2a	HPA-2a	G	2	0	100.0		100
	HPA-2b	HPA-2b	A	6	0	100.0		100
HPA5 a/b-rs10471371	HPA-5a	HPA-5a	G	24	0	100.0		100
	HPA-5b	HPA-5b	A	28	0	100.0		100
Average/Total				35275	157	99.7	71 = Serological (45%) 86 = Genotyping (55%)	100

* Confirmation test were achieved by Héma-Québec using alternative molecular based experiments.

16,598 DNAs from donors were genotyped and 365,156 phenotypes were inferred from the genotyping alleles. Out of the total number of phenotyped samples, 9.7% (35,275) were also characterized using serological techniques (Héma-Québec). A 99.7% accuracy was obtained. The discordances were tested at Héma-Québec using alternative molecular based techniques and 55% validated the genotyping results and 45% the initial serological data. The reproducibility was determined after the genotypes of ten unknown different samples were run in replicates, tested for different blood (6, 27, 55 days) and FTA card (4, 12 and 26 weeks) stabilities.

DISCUSSION

Worldwide blood banking facilities have developed rapid and efficient serological tests to evaluate the presence of major blood antigens, such as the A, B, AB, O and the RhD antigens. These tests are routinely performed to avoid any major immune responses that can happen during blood transfusion. However, in total there are 30 known blood groups leading to more than 300 different blood antigens. For a patient receiving one or two transfusions during their lifetime, mismatches for the minor blood antigens pose no problem, but a patient requiring frequent blood transfusions (patients with leukemia, hemophilia, sickle cell disease) can develop strong and fatal immune responses. For many reasons, including the continual increase in reagent costs, unreliable serological results, and lack of tests for some antigens, the detection of minor blood antigens is not routinely performed on blood donors (Quill, 2008). In addition, blood banking facilities have to screen dozens of different blood samples for specific antigens before they arbitrarily find the right donor. In order to efficiently match donors with patient who undergo frequent blood transfusions, we have developed a minor blood group antigen genotyping assay that can simultaneously detect 22 different minor blood and platelet antigens (c/C, e/E, S/s, M/N, K/k, Kpa/b, Fya/b, Jka/b, HPA-1a/b, HPA-2a/b and HPA-5a/b) in a single multiplex PCR reaction only using a few nanograms of DNA. A total of 21,000 Héma-Québec donors have been recruited for genotyping and have agreed to take part in the creation of a centralized database consisting of their minor blood antigen profiles. The creation of this database provides Héma-Québec with the instant ability to match patients who undergo frequent transfusions to the appropriately matched donors for a fraction of the cost of standard serological testing. The final, optimized

and validated, minor blood group antigen genotyping assay has been developed for the Beckman Coulter GenomeLab SNPstream platform (Fullerton, CA) and uses a DNA extraction method that can support FTA cards. The use of FTA cards has many advantages for Héma-Québec's blood banking facility, where only one drop of blood is needed, and the DNA on the FTA cards is highly stable for at least a few years at room temperature (Belgrader et *al.*1995).

Since the blood genotyping panel was developed for a clinical setting, the highest quality of genotyping was essential. To do so, we devoted a lot of effort into assay design, and optimisation. We also developed a comprehensive phased development and validation plan to fully convert all markers with the highest genotyping call rates, robustness, and accuracy. Firstly, we bioinformatically analyzed the DNA sequences that encompassed the selected minor antigens to fully characterize any genomic anomalies that could compromise our assay designs, such as gene duplications, underlying genetic variation and gene homology. The results obtained from this bioinformatics analysis permitted the rapid identification of markers that were suitable for routine automated assay design using the commercial assays design tools provided by the genotyping platform companies. For those markers that were assessed to contained irregular and challenging genomic regions, we performed manual design strategies. As shown in Table 7, high DNA sequence homologies around the markers c/C and e/E were found for the genes *RHCE* and *RHD,* and high DNA sequence homologies around the markers M/N and S/s were identified for the *GYPA, GYPB, GYPE* genes. Laboratories that only use the standard automatic design tools would have automatically rejected these four markers that are absolute requirements for our panel design, however, this would ultimately reduce the sensitivity of our genotyping panel and reduce its viability for clinical application.

Therefore, a significant amount of effort was made so that the markers M/N (*GPYA*) and S/s (*GPYB*) could be manually designed to optimise PCR oligos with the highest specificities in 3' region of their specific locus, whilst having favorable profiles with regards to Tm and other compatibilities within the PCR oligo mix. Likewise, the genes *RHCE* and *RHD* were aligned and PCR oligos were manually optimised for the markers c/C and e/E. The manual assay designs were initially based off the M/N and S/s standard single base extension designs generated by the automatic software programs, where one pair of PCR oligos and two extension probes ending one base 5' of the SNPs were created. The PCR oligo designs then needed to be manually optimised to provide greater specificity for the PCR amplicons to overcome the presence of high homology flanking the SNPs. Following phase I optimisation, the S/s marker was completed as the assay achieved all the required acceptance criteria: good cluster separation, 100% call rate, >99% concordance between both DNA strands, accurate allele frequency calls, and accurate Hardy-Weindberg values (Table 8). On the other hand, the marker M/N failed with respect to the expected allele frequency and Hardy-Weindberg values. After thorough analysis of the genotyping clusters, a clear double heterozygote cluster was observed for the assays on both DNA strands (Figure 20-D). By manually re-clustering the two heterozygous sub-clusters, the expected allele frequencies and the HW values were returned to normal, which indicated that the problem resided into the PCR oligo designs (Figures 20-E and F). Additionally, the minor sub-cluster was composed of African-American samples only which suggested the possibility of a unknown underlying variation that was impacting the African population genotyping patterns. Upon the sequencing of samples from both heterozygous populations, a previously unreported G>A variation was discovered at the extreme 3' of the reverse PCR oligo specific only for the

African-American heterozygous population cluster. This finding explained the loss of heterogeneity calls for the African-American samples that were carrying the variation at the 3' of the reverse PCR oligo. This issue highlights the need for using control samples from diverse populations during the panel development and validation phases. This issue would never have been discovered with exclusive use of Caucasian samples.

More complex genotyping designs were developed for the markers e /E and c/C. As described in the results sections (Section 1.2), the marker e/E was quite challenging to develop because it is a G>C variation and the SNPstream GenomLab chemistry allowed only the same extension types (GA and CT) in multiplex extension reactions. To bypass this issue, we detected the presence of each allele (G and C) using two different extension mixes, thus the allele **G** was detected using the **C**T extension mix and the allele **C** was detected using the **G**A extension mix (Figure 17). The results, as explained in Table 8 and Figure 17 showed the MAF fitted perfectly with the expected value and ultimately no critical problems were observed in the phase I experiment. However, during the phase I optimisation, the assay performed with the GA extension mix showed very high non-specific extension (Figure 17-B; green cluster). This non-specific extension was expected in the GA mix as there was no C nucleotide available for the normal extension on the template containing the G allele. This led to a high potential of non-specific G nucleotide incorporation (Yang *et al.*, 2002). Nevertheless, the non-specific G incorporation genotyping profile was constant over time, permitting the establishment of reasonable clustering thresholds between samples with specific and non-specific nucleotide incorporation (Figure 17-B). For the CT extension, some non-specific extension was also observed, but in this case, instead of the same C nucleotide being specifically

and non-specifically extended, the T nucleotide was incorporated for samples carrying the G allele, leading to two opposite clusters (Figure 17-B). Normally, the absence of a T allele template should lead to a failed cluster, but the increased amount of ddT in the extension mix provided an excellent source of nucleotide and the DNA polymerase incorporated it into a mismatched template, ultimately leading to non-specific extension. Since the final genotyping results have to be inferred from both DNA strand assays (GA and CT) we do not have the same quality control concordance values that we have for most of the other SNPs, however, the assay results for the marker e/E was consistant and accurate enough to be successfully incorporated to the minor blood group antigen genotyping panel.

The marker c/C was the most challenging marker to be incorporated into the genotyping panel. This marker design was also conducted manually using the aligned sequences of the genes *RHCE* and *RHD*. The complexity of this design came from the very high homology of the *RHCE-C* allele with the *RHD* gene. This homology prevented the conventional genotyping designs for the simultaneous detection of the *RHCE-C* and *RHCE-c* alleles. To avoid any genotyping signal background from the *RHD* gene, which would lead to misinterpretation of the c/C antigen phenotypes, we optimized the PCR amplification to be specific only to the *RHCE-c* allele using sequences mismatches between the *RHCE-c* allele and *RHD* (Figure 18-A). The detection of the small c allele was once again not trivial. The *RHCE-c* allele was amplified, and the only possible detectable alleles were A (for the extension GA) and T (for the extension CT). However, in both assays (CT and GA) non-specific extensions were observed, similar to the issue described for e/E (CT). The detection of the *RHCE-C* allele was obtained by designing an additional assay, which detected the presence of the 155 bp insertion in intron 2 of the gene *RHCE*. Reported by Tax

*et al.,*2002, the presence of this insertion in intron 2 is highly associated with the presence of *RHCE-C* antigen. Even if some non-specific ddG and ddT incorporation occurred for both DNA strand assays, the correct genotyping clustering metrics and expected allele frequency values obtained during the optimisation phases permitted the establishment of a clear clustering threshold, which allowed for the successful incorporation of the c/C antigens into the genotyping panel.

For validation, we evaluated the minor blood group antigen genotyping panel for its accuracy and robustness. According to Bell and coworkers (2002), the expected call rate for the SNPstream GenomeLab system is 95%. With the idea of moving the assay into clinical application, a 95% call rate combined with external factors including possible variability in blood quality, genotyping errors, possible variability in array hybridization plate quality was not ideal. The overall genotyping call rate obtained during our validation phase was 99.9% on both DNA strands. Thus, we concluded that the 95% call rate estimated by Bell *et al.* (2002) reflected only the average call rate that is achieved when standard automatic designs are used. However, our current result of 99.9% assay call rate is a clear indication that our design and optimisation strategy that utilizes an extensive bioinformatics evaluation followed by a phased development process significantly enhances the quality of genotyping call rates, as well as the marker conversion rate. The lowest call rates obtained were 99.8% and 99.2% for the markers c/C-intron and e/E respectively (Table 10). Our lowest call rates were expected for our non-standard marker designs, but regardless, the very high call rate % obtained for these assays during the reliability testing demonstrated a high degree of confidence for both markers. In

addition, the minor blood group antigen panel has been tested on different days, and by two different technicians with a 99.9% correlation (Table 11).

To determine accuracy, we cross-compared the genotyping results obtained from the panel developed on the SNPstream GenomeLab technology and the panel developed on the Sequenom platform. The data showed more than 99.8% genotyping call concordance from the both DNA strands on both platforms, and 100% cross platform concordance if we considered only the Coriell samples, and 99.8% if we included the DNA extracted from the FTA samples (Table 9). The discordances observed with the FTA samples were associated with the c/C-intron marker, which was not surprising since this assay demonstrated the worst sensitivity to low DNA concentrations. Of particular relevance from the cross-technology comparison is that, except for c/C-intron, e/E and M/N, all the PCR and extension oligos were different between the SNPstream and the Sequenom panels, further re-enforcing the validity and accuracy of the genotyping results obtained during this study. The high accuracy levels were ultimately confirmed by sequencing where 100% of the identified sequences were concordant with the genotyping results.

The final accuracy test evaluated 618 well characterized FTA blood donor samples, where the antigen phenotypes inferred from the genotyping panel were cross-compared with the antigen phenotypes determined through standard serologic techniques (FDA and Health Canada approved). Within the 618 samples, 32 samples were duplicated and 99.7% concordance was seen for the replicates, indicating that the reproducibility of the panel is constant over a diverse source of DNA. The data obtained from this accuracy experiment also showed the reliability of the genotyping

technique, as a 97% correlation was obtained between the genotype and serological techniques.

The only exception to the otherwise high accuracy rate was seen for HPA-1b where only a 50% correlation was found (out of 6 samples). However, all samples that showed discrepancies were re-tested using an alternative molecular based methodology and all but two of the HPA-1b results were confirmed to be accurate from genotyping. The two exceptions, which confirmed the serologic results, are associated with the two non-standard markers, c/C-intron and e/E. Despite these two discordant results the concordance percentage is still very high, as 99.2% and 99.5% concordance was obtained for the c/C-intron (C allele) and e/E (e and E alleles), and an overall concordance of 99.3% was obtained during this validation step. This confirmed the utility of the genotyping blood antigen panel for the creation of the Héma-Québec minor blood group antigen database.

As a comparison, Denomme *et al.* 2005 performed a similar study and was unable to achieve an acceptable conversion rate for these markers. They obtained a significantly lower concordance rate, especially for the two non-standard markers c/C (C antigen; 87.7%) and e/E (e antigen; 73.9%), which would preclude the use of their panel in a clinical context. They explained the problem by the co-amplification of the genes *RHCE* and *RHD* during the multiplex PCR. This example clearly illustrates the effectiveness of our assay design efforts and highlights the need to perform a deeper bioinformatics analysis to accurately designing oligos that are specific to the desired templates. In addition, to overcome the problem of inferring the C allele, we designed an alternative assay with greater specificity for the presence of *RHCE-C* allele.

There are many factors that could lead to a difference between the inferred phenotypes from the genotyping results and the phenotype determined by serological

techniques. First of all, it's well known that more than one polymorphism can contribute to the protein conformation and ultimately affect the antigen expression. For example, the SNP located in the promoter region of the *FY* gene occurred because of the presence of the malarial parasite *Plasmodium vivax* in endemic African regions and therefore is mostly associated with African descent (Storry *et al.,* 2007). The mutation located in the GATA-1 box region (called Fy0) is erythroid specific, meaning that individuals with this variation lack the Fy glycoprotein. For most carriers this silencing mutation is located on the genomic background of the Fyb allele, but it was also found on the Fya allele background (Storry *et al.,* 2007). In our validation phases, the alternative molecular based assays confirmed that two of our three samples that differed with the serological results had this Fy0 mutation. Also another SNP, FyX, more prevalent in individuals of European descent (~2.0%), is known to produce antigens almost undetectable by most anti-Fyb reagents. These examples highlight the limitations of serologic testing. Storry et al. 2007 showed in their study that 7 out of 74 samples had both the Fy0 and Fyx variants and thus the RBC carried only one dose of the antigen but they appeared, by serologic testing, to express both antigen doses.

Because of the high homology between the *RHCE* and *RHD* genes, particularly around exon 2, we had to develop a non-standard second assay. Instead of detecting directly any polymorphism resulting in C antigens, we detected the155 bp insertion in intron 2 which is *cis* with respect to the RHCE-C allele. The insertion is in high linkage for individuals of Caucasian descent, but 25% of individuals of African descent carry a hybrid gene (D-CE-D) containing exons 1 through 3 of *RHD* gene. These individuals express a weak *RHC* antigen and therefore give false negative results (Daniels *et al.,* 1998 and Storry *et al., 2007*). As discussed above, the biggest

discordance in our results was observed for the HPA-1b antigen. It is well documented that for erythrocyte antigens, except for HPA-1a where a monoclonal antibody is available, the serologic detection techniques are highly variable as the tests require human antisera and immune fluorescence based assays (Beiboer *et al.,* 2005). The possible variability in the HPA-1b antigen is most likely due to poor reagents and not the sensitivity of our genotyping assay and could explain the initial 50% discordance observed between our genotyping data and the phenotype data provided by Héma-Québec.

In addition to the limitations of the serological assays described above, there is a general problem in the blood banking world regarding the accessibility to well characterized and reliable reagents. Many important minor antigens have to be assessed using polyclonal antibodies, which are often only weakly reactive. In addition, the Knops antigens system has been well documented to generate inaccurate phenotypes when serologic test are used (Storry *et al.,* 2007). The same study reported an issue when using hemagglutination for the detection of the Jsa, Kpa, Cw, Goa, Dia and Mia minor antigens. Inconsistent serological results for these antigens can be very problematic as these antigens are highly clinically relevant, especially when used to screen specific populations where these antigens are frequent. Storry and colleagues (2007), have also reported an issue with using serologic assays for the detection of the antigens S/s. They explained that evolutionary pressure by the malarial parasite, *Plasmodium falciparum* has led in some individuals of African descent, to a deleted and mutated *GYPB* gene, which transmits a selective advantage against infection. This polymorphism ultimately leads to unreliable serological results as current testing methods cannot accurately determine the correct antigen. As

described in the introduction, there is a great need for correctly matched blood donors for individuals who have been exposed to multiple transfusions and who are now refractory to PLTs or RBCs because of immunological responses that have resulting in antibodies to a variety of antigens. This is a major concern for blood banks worldwide. This is particularly true for patients with diseases like sickle cell disease or thalassemia, which necessitate regular transfusions over long periods of time. Screening blood bank donations for compatible blood when the patient has a mixture of relatively uncommon antibodies can be difficult, sometimes impossible, because the reagents, rare antisera, are difficult to obtain (Anstee, 2005).

Taken altogether, the challenges and limitations surrounding serological screening of rare antigens are driving the world-wide effort to develop new and improved methodologies. Specifically, DNA-based assays are now gaining acceptance for the detection of the minor blood group antigens. This new initiative is currently gaining acceptance in the form of either establishing minor blood group antigen databases or performing greater quality control on established serological tests.

Following our very encouraging results obtained with our genotyping panel during the validation phase, we have established a validated process starting from sample reception to the final generation of reports (Figure 11). We have developed successful standard operation procedures (SOPs) that have firstly allowed us to efficiently handle 21,000 test samples from their reception to the reporting stage; and secondly, we have developed robust quality control procedures which we have integrated into all aspects of the genotyping process to assess the quality of the final inferred phenotypes. These SOPs support all aspects of the process from sample reception, DNA extraction, genotyping, through to the predicted phenotype (Figures

24 to 29). The complete process is mediated by our custom LIMS (Laboratory Information Management System) and unique barcode identifiers which ensure full traceability of samples, reagents and experiment time. To eliminate the possibility of mixing up samples provided by Héma-Québec, we have developed a system that electronically links the client barcodes with our own unique barcodes in our LIMS. Even though our laboratory and this process has not yet become officially accredited (eg. CLIA), the process that we have develop meets GLP standards as our laboratory practice during the advancement of this project followed all the documentation required for a controlled GLP process as closely as possible. It is important to mention that the system was audited by two independent pharmaceutical companies, who were in agreement that our processes were very close to being GLP compliant.

A critical part of our ability to perform accurate clinical genotyping on a large number of samples is the establishment of quality control steps. We have efficiently introduced a series of quality control steps from the DNA extraction through to the reporting of phenotypes to Héma-Québec. The main quality control steps that we successfully incorporated into our process were the introduction of the SHERPA software and the reporting tool Area51. As presented in the Results section, the SHERPA software has the potential to flag, at an early stage, any plate problem patterns that may ultimately impact on the high-throughput genotyping process. Also, SHERPA contributed to increasing the accuracy of the results by automatically interpreting the genotyping calls from two different technicians, and by interpreting the concordance from both DNA strands. The comparison of two independent genotyping results obtained from two technicians was a particularly useful quality control step for the non-standard marker e/E, where no comparison of both DNA strands was available.

Finally, the reporting interface Area51 has raised the quality of our results by automatically inferring the phenotypes from the genotype data, thus limiting the impact of human error during the reporting steps. In addition, Area51 is directly linked to the SHERPA database; consequently it is able to interpret the data, based on predefined rules, allowing it to report on sample failures, control failures, and ultimately on experimental failures. This rapid and efficient reporting has elevated the quality of final results by excluding samples that failed more than two markers (out of 24 total markers from GA and CT extension pools) and samples that had more than one discordant call from both DNA strands. Area51 has so far flagged two major problems related to control DNAs which were later tracked to a specific, fixable cause, but this example illustrates the robustness of Area51 to report any kind of issues that could contribute to decreased quality of results. By having the quality controls and the reporting tools in place during the project, we were able to evaluate to date 16,598 different frequent donors for a total of 365,156 inferred phenotypes. A total of 382 samples were re-genotyped because they failed the Area51 reporting filters and 100% of them were recovered in a second round of genotyping. This demonstrated that the samples failing the first round of genotyping were related to technical issues and not the reliability of the genotyping designs.

The most impressive results that came out from this project were the accuracy of our assay over thousands of samples. Approximately 10% of the total results generated so far were compared with the gold standard serological assays and we have achieved, similar results to those obtained during the validation phase, 99.7% accuracy and 100% reproducibility. A total of 157 phenotypes were discordant and 55% (86 phenotypes) were confirmed for the genotyping results using an alternate molecular method, whereas 45% (71 phenotypes) were confirmed to be accurate for

the serological assays. Interestingly, as observed during the validation phase, the most problematic antigens were e/E (69 discordant), Fya/b (63 discordant) and c/C (9 discordant), which account for more than 90% of the discrepant results. In the Fya/b assay, the Fyb antigen generated 57 discordant calls and all were related to the Fy0 or FyX variants, which clearly demonstrate the limitation of the gold standard serological assay when silencing markers are present. In the case of the antigen c/C, only 3 phenotypes (C antigen) were confirmed for the serological data, which led to a final concordance for that antigen of 99.9%, confirming that the strategy developed to interpret the C allele was valuable and applicable in a high throughput genotyping process. The most problematic assay remains the e/E antigen, where 60 out 69 discrepancies were positively replicated by the serological results. The discordance between both assays is 100% related to the non-standard design, and there is no double DNA strand data to test for concordance during the quality control phases since both strands are needed for the final phenotype interpretation.

As expected from the validation phase, the two non-standard markers (c/C-intron and e/E) generated more than 97% of all serological confirmed discrepancies. Fortunately, 100% of the e/E discordance encountered during this project was fully resolvable using the Sequenom platform, where no limitations exist for the extension reactions of various SNPs.

The process that we have developed to integrate minor blood group antigen genotyping into our laboratory, to support Héma-Québec, was also recently used to support the genotyping of 10,000 samples from the Canadian Blood Services. This group has the same aim, which was the vision of building an informatics database of 20 minor blood antigens. By using the same panel, with the exception of replacing the platelets antigens by two *RHD* markers, and including the Fy0 marker, we have

been able to complete the genotyping of 10,000 samples (from Toronto) over three months. We currently have no data comparing genotyping with serological phenotypes, but our internal results showed that only 0.4% of total genotyped samples failed our filters, and an overall call rate of 99.8% was achieved, which is significantly higher than the 95% expected by Bell *et al.,*2002.

In the near future, both the Canadian Blood Services and Héma-Québec are interested in the expansion of the current panel. We aim to incorporate 60-70 more antigens into the panel using the Sequenom platform. A preliminary list has been created, which includes the Knops, the Dombrock, the Lutheran, and the Colton systems. The addition of these extra antigen systems will significantly increase the amount of clinical relevant information and will ultimately provide better characterization of the minor blood group antigens fro each blood donor, improving the sensitivity of the assay the ultimate selection of suitable blood for specific patient needs. Moreover, we expect to incorporate some silencing mutations, such as the Fy0 and FyX in the next generation of the panel, in order to eliminate the recurrent discrepancies observed for the Fyb antigen.

With the use of the Sequenom platform, we will eliminate the extension issues observed for the e/E (G>C) assay obtained with the SNPstream platform, as the Sequenom chemistry can handle all different extension types in a single multiplex reaction. Even though Sequenom will eliminate most of the concerns related to the e/E antigens, we expect difficulties for all markers located in the *RHCE*, *GYPA* and *GYPB* genes. To overcome this homology problem for future markers, the same design strategy used in this work will be applied to successfully incorporate all crucial makers into the extended panel.

In conclusion, we have successfully developed a DNA-based assay suitable for the creation of a minor blood group antigen database of 21,000 frequent blood donors for Héma-Québec. Also we have successfully integrated a GLP-like process dedicated to high-throughput multiplex genotyping. To date, 16,598 donors have been genotyped for a total of 365,156 total inferred phenotypes with 99.7% accuracy. The high accuracy and reproducibility achieved during this project has proven that the development strategy that we have put in place in the early stage of the project has enhanced the quality of the genotyping results. Similar development approaches used in this project, have been adapted for use in the laboratory to support other major pharmacogenomics projects. For instance, we have used similar development and clinical application approaches for a project involving lipid related genes and 400 well characterized patients that have experienced myotoxic adverse effects as a result of their Statin medication.

Recently, the same slightly modified minor antigen blood panel has been used by the Canadian Blood Services to generate a similar minor blood group antigen database, where 10,000 frequent donors have been phenotyped using our panel. Finally, in March 2008, Elizabeth Quill cited in Science Magazine that our laboratory and Héma-Québec were the first blood units in Canada and one of the leaders in North America that were actively moving ahead in the large scale genotyping arena. This article demonstrated that our work has initiated a new era for blood antigens genotyping. The DNA-based assay we developed here will provide significantly more information about each donor than previously used methods. This will reduce future complications in patients receiving multiple transfusions.

The minor blood group antigen informatics database created from this work is currently being used for all hospitals in Quebec and has already demonstrated great improvements in the selection of the right blood for the right patient. It is anticipated that the same usage will become common practice in additional hospitals across Canada. This work was the first personalized medicine application developed and implemented at the Génome Québec and Montreal Heart Institute Pharmacogenomics Centre and is paving the way for the next set of pharmacogenomic assays that are moving towards the clinic.

BIBLIOGRAPHY

1. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *Science* **322**, 881-888(2008).
2. Anstee, D.J. Goodbye to agglutination and all that? *Transfusion* **45**, 652-653(2005).
3. Avent, N., Martinez, A. et al. The bloodGen project : toward mass-scale comprehensive genotyping of blood donors in the European Union and beyond. *Transfusion* **47**, 40S-46S(2007).
4. Beiboer, SH., Wieringa-Jelsma, T., Maakant-Can-Wijk, PA., et al. Rapid genotyping of blood group antigens by multiplex polymerase chain reaction and DNA microarray hybridization. *Transfusion* **45**, 667-679(2005).
5. Belgrader, P., Del Rios, M., Reid, M.E. Automated DNA purification and amplification from blood-stained cards using a robotic workstation. *Biotechniques* **19**, 426-432(1995).
6. Bell., PA., Chaturvedi, S., Gelfand, CA, et al. SNPstream UHT:ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* **32**, S70-S77(2002).
7. Brennan, M.D. High throughput genotyping technologies for pharmacogenomics. *Am J Pharmacogenomics* **1**, 295-302(2001).
8. Conner, B.J. et al. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci U S A* **80**, 278-82(1983).
9. Daniels, G.L. et al. The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion* **38**, 951-958(1998).
10. de Denus, S. et al. Key articles and guidelines in the management of hypertension. *Pharmacotherapy* **24**, 1385-99(2004).
11. Debrauwere, H. et al. Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. *Biochimie* **79**, 577-86(1997).
12. Denomme, G.A. & Van Oene, M. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* **45**, 660-666(2005).
13. Doxiadis, I.I.N. & Claas, F.H.J. The short story of HLA and its methods. *Dev Ophthalmol* **36**, 5-11(2003).
14. Ekman, G.C., Billingsly, R. & Hessner, M.J. Rh genotyping: avoiding false-negative and false-positive results among individuals of African ancestry. *Am. J. Hematol* **69**, 34-40(2002).
15. Fu, Y.H. et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047-58(1991).

16. Goidts, V. et al. Complex patterns of copy number variation at sites of segmental duplications: an important category of structural variation in the human genome. *Hum Genet* **120**, 270-84(2006).
17. Gonzalez, E. et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434-40(2005).
18. Hanai, J. et al. The muscle-specific ubiquitin ligase atrogin-1/MAFbx mediates statin-induced muscle toxicity. *J Clin Invest* **117**, 3940-51(2007).
19. Iwamoto, S. et al. Identification of a novel exon and spliced form of Duffy mRNA that is the predominant transcript in both erythroid and postcapillary venule endothelium. *Blood* **87**, 378-85(1996).
20. Jannetto, P.J., Laleli-Sahin, E. & Wong, S.H. Pharmacogenomic genotyping methodologies. *Clin Chem Lab Med* **42**, 1256-64(2004).
21. Jones, T.S. et al. Using HapMap tools in pharmacogenomic discovery: the thiopurine methyltransferase polymorphism. *Clin Pharmacol Ther* **81**, 729-34(2007).
22. King, C.R. et al. Performance of commercial platforms for rapid genotyping of polymorphisms affecting warfarin dose. *Am J Clin Pathol* **129**, 876-83(2008).
23. Iakoubova, O. Association of the Trp719Arg Polymorphism in kinesin-Like Protein 6 with Myocardial Infarction and Coronary Heart Disease in 2 Prospective Trials. *Journal of the American College of Cardiology* **51**, 435-443(2008).
24. Liao, J.K. Isoprenoids as mediators of the biological effects of statins. *J.Clin. Invest.* **110,** 285-288(2002).
25. Lopez-Miranda, J. et al. Influence of mutation in human apolipoprotein A-1 gene promoter on plasma LDL cholesterol response to dietary fat. *Lancet* **343**, 1246-9(1994).
26. Mackay, T.F. The genetic architecture of quantitative traits. *Annu Rev Genet* **35**, 303-39(2001).
27. Mahley, R.W. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622-30(1988).
28. Manolopoulos, V.G. Pharmacogenomics and adverse drug reactions in diagnostic and clinical practice. *Clin Chem Lab Med* **45**, 801-14(2007).
29. McGuire, W. et al. Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* **371**, 508-10(1994).
30. McLeod, H.L. & Siva, C. The thiopurine S-methyltransferase gene locus -- implications for clinical pharmacogenomics. *Pharmacogenomics* **3**, 89-98(2002).
31. McLeod, H.L. et al. High-througput Genotyping Methods for Pharmacogenomic Studies. *Current Pharmacogenomics* **2,** 21-33(2004).
32. Moskowitz, S.M. et al. Clinical practice and genetic counseling for cystic

fibrosis and CFTR-related disorders. *Genet. Med* **10**, 851-868(2008).

33. Nachman, M.W. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet* **17**, 481-5(2001).
34. Polin, H., et al. Introduction of real-time based blood group genotyping approach. *Vox Sanguinis* 1-6(2008).
35. Poulter, M., Kemp, T.J. & Carritt, B. DNA-based rhesus typing: simultaneous determination of RHC and RHD status using the polymerase chain reaction. *Vox Sang* **70**, 164-168(1996).
36. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69**, 124-37(2001).
37. Elizabeth, Q., Blood-Matching Goes Genetic. *Science* **319,** 1418-1419(2008).
38. Reid, M.E. Applications of DNA-based assays in blood group antigen and antibody identification. *Transfusion* **43**, 1748-1757(2003).
39. Reid, M.E. Overview of molecular methods in immunohematology. *Transfusion* **47**, 10S-16S(2007).
40. Risch, N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* **10**, 733-41(2001).
41. Roden, D.M. et al. Pharmacogenomics: challenges and opportunities. *Ann Intern Med* **145**, 749-57(2006).
42. Roses, A.D. Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nat Rev Genet* **5**, 645-56(2004).
43. Sachidanandam, R. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-33(2001).
44. Sheffield, L.J. & Phillimore, H.E. Clinical use of pharmacogenomic tests in 2009. *Clin Biochem Rev* **30**, 55-65(2009).
45. Shi, M.M. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin Chem* **47**, 164-72(2001).
46. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881-5(2007).

    Storry, JR., Olsson, ML. Genetic baisis of blood group diversity. *Br J Heamatol* **126**, 759-771(2004)
47. Sutherland, G.R. & Richards, R.I. Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci U S A* **92**, 3636-41(1995).
48. Swen, J.J. et al. Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med* **4**, e209(2007).
49. Syvanen, A. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* **2**, 930-942(2001).
50. Tax, M.G.H.M. et al. RHC and RHc genotyping in different ethnic groups.

Transfusion **42**, 634-644(2002).

51. Taylor, J.G. et al. Using genetic variation to study human disease. *Trends Mol Med* **7**, 507-12(2001).
52. Toscano, C. et al. Impaired expression of CYP2D6 in intermediate metabolizers carrying the *41 allele caused by the intronic SNP 2988G>A: evidence for modulation of splicing events. *Pharmacogenet Genomics* **16**, 755-66(2006).
53. Van Rooij T. Behind the PGx challenge: Biofx. *Genome Technology*. July/August 2007.
54. Van Schie et al. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. (2009)
55. Wallace, R.B. et al. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* **6**, 3543-57(1979).
56. Weiss, K.M. In search of human variation. *Genome Res* **8**, 691-7(1998).
57. Yang, L. et al. Local deformations revealed by dynamics simulations of DNA polymerase Beta with DNA mismatches at the primer terminus. *J. Mol. Biol* **321**, 459-478(2002).
58. Zhang, W., Ratain, M.J. & Dolan, M.E. The HapMap Resource is Providing New Insights into Ourselves and its Application to Pharmacogenomics. *Bioinform Biol Insights* **2**, 15-23(2008).
59. Zhu, M. & Zhao, S. Candidate gene identification approach: progress and challenges. *Int J Biol Sci* **3**, 420-7(2007).
60. Zuckerkandl, E. & Cavalli, G. Combinatorial epigenetics, "junk DNA", and the evolution of complex organisms. *Gene* **390**, 232-42(2007).
61. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA* **285**, 2486-97(2001).
62. The International HapMap Project. *Nature* **426**, 789-96(2003).

ANNEXES

Annex A

This Standard Operating Procedure is currently used at the Genome Quebec and Montreal Heart Institute Pharmacogenomics Centre as reference document for the sample management and DNA extraction from FTA cards. This SOP is referred at the Material and Methods under the Clinical Application section.

Automated DNA Extraction of Blood Spots from FTA Cards using the Generation DNA Purification Technology by Qiagen for Héma-Québec's Project

Author Signature:

_____ Signature	_____ Date (dd-mmm-yyyy)

Lizette Mailloux, Clinical Technician

Reviewer and Approver Signature:

_____ Signature	_____ Date (dd-mmm-yyyy)

Ian Mongrain, Senior Technician

1. Purpose

1.1. This SOP defines the procedure for automated DNA extraction of blood spots from FTA cards with the Generation DNA Purification Technology by Qiagen for Héma-Québec's Project**.**

2. Scope

2.1. This SOP applies to DNA extraction from a 2 mm punch of FTA card impregnated with whole blood.

2.2. This SOP applies to either automated or manual procedures of FTA card punching and DNA isolation.

3. Abbreviations

3.1	DCN	:	Document Control Number
3.2	DNA	:	Deoxyribonucleic acid
3.3	LIMS	:	Laboratory Information Management System
3.4	PCR	:	Polymerase chain reaction
3.5	PGx	:	Pharmacogenomics
3.6	QA	:	Quality Assurance
3.7	SOP	:	Standard Operating Procedure

4. Contents

5. Definitions: N/A

6. Responsibilities

6.1. Directors, Clinical Lab Leaders and QA representatives

6.1.1. Ensure that all PGx employees required to isolate DNA from blood spots on FTA cards using the Generation Technology are trained and follow this SOP.

6.1.2. Maintain and, when required, review and sign this SOP.

6.1.3. Any of these responsibilities can be delegated to the appropriate person provided he/she was trained on this SOP.

6.2. Technicians

6.2.1. Follow this SOP.

6.3. Reviewer

6.3.1. All reviewers must have the required background in order to understand the content of the document.

7. Supplies

7.1. Equipment

- Biomek FX Laboratory Automation workstation
- Biomek Software
- BSD600-Duet Automatic puncher
- 2 mm puncher
- Cutting mat
- Centrifuge
- Multichannel pipettes (20 µL and 200 µL)
- Heating plate sealer
- Thermocycler (96-well module)
- Barcode reader
- Biotracker (LIMS)

7.2 Consumables

- FTA Classic Card (Whatman) or FTA Mini Card (Whatman)
- 55 mL Reservoir Basin
- 1.2 mL 96-well plate
- Thermo-Seal Heat Sealing Foil
- Adhesive plastic film
- 20 uL and 200 uL sterile filter tips

- Robotic Workstation 250 uL sterile filter tips
- Robotic Workstation 200 uL conductive sterile filter tips
- Robotic Workstation 1000 uL conductive sterile filter tips
- Quarter Reservoir, Divided by Length, 19 mL/Section
- Reservoir Holders
- Waste containers
- FTA Cardboard box

7.3 Chemicals

- Generation DNA Purification Solution #1 (Qiagen)
- Generation DNA Elution Solution #2 (Qiagen)
- Distilled water (RNase/DNase–free grade)

8. Generalities

8.1. Delimited zones for DNA extraction, pre-PCR and post-PCR are to be used accordingly to the step being performed. This will avoid any carry-over contamination (refer to SOP PGx-IG-016 *Material Sample and Personnel Flow and Hygiene*).

8.2. Every technical step performed has to be documented and recorded in the LIMS.

8.3. Every consumable, chemical and equipment has to be documented and recorded in the LIMS.

8.4. Any observation, abnormality or incident has to be documented and recorded in the LIMS. Furthermore, incidents have to be reported to QA in order to be documented as per SOP PGx-IG-006 *Deviation, Incident and Investigation reporting*.

8.5. Lab coat and gloves should be worn at all times.

Exception: in order to avoid static problems and loss of samples, the hand that handles the 1.2 mL 96-well plastic plate during the FTA card punching step should always be glove-free.

9. Procedure

9.1. Pre-Generation of FTA card PGx barcodes

Note: The amount of barcodes to be generated relies on Héma-Québec's request.

9.1.1. Open the folder *R:\Commun\Biotracker\Generic sample Batch Import\HQA-HQB* to know what the last numbers used were for the previous generation of FTA barcodes (e.g: HQ_Temporary_0001-1000).

9.1.2. Open the template file: *HQ_PreGeneration_FTA Barcode_3Sept08.xlsx* located under: *R:\Commun\Héma-Québec\Temporaire-Sample-reception-checker*.

9.1.3. Replace the "*ClientSampleID"* list with a new sample list that corresponds to the new incremental numbers: (e.g. SampleUpdate1001 to SampleUpdate2000) and insert the location number.

9.1.4. Save the previous file according to the corresponding new numbers: in the folder *R:\Commun\Biotracker\Generic sample batch import* (e.g. HQ_Temporary_1001 to 2000).

9.1.5. Open the LIMS, and go to the *"Samples"* tab and again select *"samples"* in the roll down list.

9.1.6. Select a *"sample type"* (FTA paper) and click on the *"Import"* button at the bottom of the screen.

9.1.7. Retrieve the previously created file from the *"Select a file"* row and select the mapping *"HQ_Import".*

9.1.8. Click on *"Import".*

9.1.9. When all temporary samples are imported into the LIMS, save the new entries and select all the samples and print the corresponding barcodes.

9.1.10. Send the printed barcodes to the following address:

Care of: **Josée Perreault**
Héma-Québec
1070 avenue des Sciences-de-la-vie
Quebec, Qc
G1V5C3

9.2. Sample Reception and Update

9.2.1. Upon arrival of Héma-Québec's FTA card boxes, create a new location and a barcode for all boxes with the Location Explorer of the LIMS (refer to the LIMS User's manual).

9.2.2. Print the barcodes and label the corresponding boxes.

9.2.3. In Knowledge Tree, open the file: */Shared/Héma-Québec/Tools/Sample_Reception_Checker_SampleBatchUpdate_template_3Sep08*

9.2.4. In Knowledge Tree, open the HQ file: *PGC/Clients/Héma-Québec/Input*/.

9.2.5. Copy and paste all data from the previous file to the "*File from HQA"* tab in the *Sample Reception Checker*.

9.2.6. In the "*Date Received"* column, document the sample reception date in the following format: YYYY-MM-DD.

9.2.7. Save the previous file under: R:\Commun\Héma-Québec*Temporaire-Sample-reception-checker* following this nomenclature: HQA or HQB_DATE_Start PGx Barcode – Last PGx Barcode (e.g. *HQB_07Aug2008_FT000HQB8001-FT000HQB9000.xlsx).*

9.2.8. Open the "*Step 1"* tab and scan the Client Sample ID and the PGx barcodes on each FTA Card starting with the box that has the lowest barcode number.

Note: Throughout the scanning process, verify that the status is *Passed* or *Failed*. Address any kind of problem (e.g. wrong FTA papers, wrong order, missing sample etc.) and notify the client right away.

9.2.9. After scanning all samples, fill out Héma-Québec's form (comes with the shipment) and fax it to the appropriate fax number shown on the document.

9.2.10. Save again the file and make a copy available on Knowledge Tree: */Shared/Héma-Québec/Final_Sample Reception Checker*.

9.2.11. Open the "*Upload to LIMS"* tab from the previous file. Select the cells containing the info and paste (special value) them on a new Excel worksheet.

9.2.12. In the *"Location"* column, assign the new locations that were previously created (see section 9.2.1) and that correspond to the FTA card box barcodes.

9.2.13. Save the new worksheet in .xls format under *R:\Commun\Biotracker\Generic sample batch update*. Follow the same nomenclature as for the *Sample Reception Checker.*

9.2.14. Open the LIMS, and go to the *"Samples"* tab and again select *"samples"* in the roll down list.

9.2.15. Select a *"sample type"* (FTA paper) and click on the *"Import"* button.

9.2.16. Select the *"Update"* option instead of the default *"Import"* on the upper portion of the window

9.2.17. Retrieve the previously created file from the *"Select a file"* row and select the mapping *"HQ_Update"*.

9.2.18. Click on *"Import"* at the bottom of the window.

9.2.19. Once the new updated samples are imported into the LIMS, save them and close the screen.

9.3. Creating a Plate Layout

9.3.1. Open the template file in Knowledge Tree: */Shared/Héma-Québec/Tools/Plate_Maker.xlsx*.

9.3.2. Save the file in the following folder *R:\Commun\Héma-Québec\Plate_Maker*. Follow the plate maker naming scheme to name the file (refer to Appendix 3 *Naming Scheme for Héma-Québec's Project*).

9.3.3. In Knowledge Tree, open the file */Shared/Héma-Québec/REDO/redo.xlsx*

9.3.4. Copy the PGx Names with their corresponding Client Sample ID for all redoes (if there are any).

9.3.5. In the "*Plates_list"* tab of the Plate Maker file, paste the redoes at the beginning of the *FTA PGx_Barcode* and *ClientSampleID* columns of *Plate #1*.

9.3.6. Open in Knowledge Tree the folder */Shared/Héma-Québec/Final_Sample Reception Checker/* and select the *Sample Reception Checker* file previously created (refer to section 9.2).

9.3.7. In the "*Upload to LIMS"* tab of the *Sample Reception Checker* file, copy 88 PGx Names with their corresponding Client Sample ID.

Note: The number of samples to be copied should correspond to 88 minus the number of redoes (if there are any).

9.3.8. In the Plates_list tab of the Plate Maker file, paste the 88 previously copied samples to the *FTA PGx_Barcode* and *ClientSampleID* columns of *Plate #1*.

9.3.9. Repeat steps 9.3.7 and 9.3.8 with 88 new samples for *Plate #2.*

9.3.10. Go to the "*FTA_Plate_Layout"* tab and identify the two plates (refer to Appendix 3 *Naming Scheme for Héma-Québec's Project*).

9.3.11. Also add the ID of one PGx blank FTA card under each H5 well of the two plates.

9.3.12. Print the Plate layout from the "*FTA Plate_Layout"* tab of the Plate Maker.

9.3.13. Following the plate-layout, retrieve the 88 FTA Card samples, chosen for *Plate #1,* in an empty processing box.

9.3.14. If needed, scan the PGx barcodes in the LIMS for location of samples and redoes.

9.3.15. Add the corresponding PGx blank FTA Card at the end of every 88 samples.

9.3.16. Manually identify the processing box. (Refer to Appendix 3 *Naming Scheme for Héma-Québec's Project*).

9.3.17. Repeat the step 9.3.13 to 9.3.15 for *Plate #2.*

9.4. **Automated FTA Card punching**

Note: If necessary, refer to Appendix 1 *Manual Punching of FTA cards*. (not include in this Annex).

9.4.1. Identify a 1.2 mL 96-well plate according to the plate layout.

9.4.2. Add 1 µL of distilled water at the bottom of each well.

9.4.3. Open the BSD Automated puncher.

9.4.4. Enter the User name and the Password.

9.4.5. Open the software BSD duet.

9.4.6. Click on *"Distribute Spots".*

9.4.7. Select *"OK"* on the upper right.

9.4.8. Click *"Continue"* twice.

9.4.9. Select the method according to the number of plates to be punched (e.g. HQA_One Plate 88 Samples for one punching plate).

9.4.10. Select *"Samples Only"* on the upper right.

9.4.11. Click *"Continue".*

9.4.12. Scan the first 1.2 mL 96-well plate's barcode or enter it manually.

9.4.13. Insert the previous plate in the BSD Automated puncher.

9.4.14. Click *"Continue"* twice.

9.4.15. Scan the PGx barcode of the first FTA Card sample**.**

Note: If the card is not yet punched, it is still possible to go back and scan the card again by clicking on the option *Scan Barcode again*.

9.4.16. Bring the FTA paper to the punching holder.

9.4.16.1. If you are using the BSD automatic punch trigger, wait until the punch is completed.

9.4.16.2. If you are using the punching pedal, press on it.

9.4.17. If there is a *Punching Error*, the BSD Automated puncher will give you options.

9.4.17.1. *Card not Punched:* Click on this option if the FTA card is not punched. The BSD Automated puncher will then prompt you to punch the card again.

9.4.17.2. *Inspect Trays:* Click on this option to move the plate forward to get a closer look at it. Verify if the spot is in its well or not.

A) If the spot is in the well, click "*Continue"* and click on "*Spot placed in cell"* or "*Spot in cell".* The BSD automated puncher will resume the punching.

B) If the spot is not in the well, click *"Continue"* and click on "*Spot discarded"* or "*Spot not found".* The BSD automated puncher will then prompt you to punch the card again.

9.4.18. As long as the run is not completed, it is possible to punch again a sample.

9.4.18.1. Double-click on the desired well.

9.4.18.2. Rescan the barcode (the barcode has to be the same otherwise the BSD puncher will not authorize the punch).

9.4.18.3. The BSD Automated puncher will then prompt you to punch the card again.

9.4.19. Repeat steps 9.4.15. to 9.4.18 for every sample.

9.4.20. At the end of the run, the "*Completed Plate"* screen will appear. Verify that all spots are present by taking out the plate and doing a visual check.

Note: At this point, it is impossible to go back and punch samples again. If spots are missing, refer to Appendix 1 *Manual Punching of FTA Card*.

9.4.21. Click on "All Spots Present".

9.4.22. Refer to the BSD Duet automated puncher's manual for daily and weekly maintenance.

9.5. Creating a run in the LIMS

9.5.1. In the LIMS, go to the "*File"* tab and select "*Experiment Manager"*.

9.5.2. Select the experiment name: *Biomek - DNA Isolation from FTA Paper.*

9.5.3. Name the run following the naming scheme of the *DNA plates* (refer to Appendix 3 *Naming Scheme for Héma-Québec's Project*).

9.5.4. Add samples for all the plates to be extracted and add (at the end) the ID of the PGx blank FTA card that was used for each plate (Refer to the plate-layout created in section 9.3.12.).

9.5.5. Click on "*Samples"* in the "*Output"* section of the run.

9.5.6. Generate PGx DNA ID for all samples and FTA blanks by clicking on the "*Auto Accession"* function.

9.5.7. With a right-click, export the table data in csv format to the folder *R:/Commun/Héma-Québec/FTA_BA_Export.*

9.5.8. Open the previously saved file and copy BA and FTA sample ID corresponding to the plate maker created in section 9.3.2.

9.5.9. Paste the FTA and the BA sample ID in the "*FTA&BA_sampleID"* tab of the *Plate Maker* file.

9.5.10. From that same tab, copy the first 88 samples and click on the "*FTA_BA-SamplesOrder"* tab and paste them under the columns *FTA PGx_Barcode* and *BA PGx_Barcode* of Plate #1.

9.5.11. Repeat the previous step with the next 88 samples for Plate #2.

9.5.12. Go back to the "*FTA&BA_sampleID"* tab. Copy the FTA blank that corresponds to Plate #1 and paste it under the 88th sample.

9.5.13. Repeat the previous step for Plate #2.

9.5.14. Click on the "*BA_Plate_Layout"* tab and identify the plates (refer to Appendix 3 *Naming Scheme for Héma-Québec's Project*).

9.5.15. Print the BA plate-layout and staple it to the corresponding FTA plate-layout (see section 9.3.12.).

9.5.16. Click on the "*BA-SampleImport_Plate 1"* tab and save the sheet as a csv file under *R:/Commun/Héma-Québec/DNA Plate Management*.

9.5.17. Repeat the previous step for "*BA-SampleImport_Plate 2"*.

9.5.18. In the LIMS, click on *"Samples"* and select *"Plate Management"* in the roll down list.

9.5.19. Create a DNA plate for each plate previously saved and print the corresponding barcodes.

9.5.20. Record the *DNA Plate's* name and barcode in the *"Comment"* section of the extraction run of the LIMS.

9.5.21. Attach the Plate Maker file to the extraction run.

9.6. Automated DNA Extraction

Note: If necessary, refer to appendix 2 *Manual DNA extraction of FTA Cards*.

9.6.1. Empty the waste container for pod2 (the 8 pipette channels arm) and fill up the water tank (if necessary).

9.6.2. Turn on the computer and the robot.

9.6.3. Start the Biomek software.

9.6.4. Make sure that the light on the upper left side and on the upper right side of the instrument is green.

9.6.5. If the light is red, this could mean a connection error. Restart the computer and the instrument.

9.6.6. If the problem persists, refer to the Biomek's user guide to resolve the issue.

9.6.7. Click on "*Instrument"* tab and select "*Home All axes"* in the roll down menu.

9.6.8. At the first dialog box, read all instructions. Click "*OK"* if the Biomek meets the specifications.

9.6.9. At the second dialog box, make sure that the span 8 does not have any tips. Then click on "*OK*".

9.6.10. Watch for air bubbles in the tubes and syringes. When the tubes and syringes are empty of bubbles, press "*OK".*

Note: If a bubble remains in a syringe after 5 min of purging, stop the robot and refer to the Biomek's user guide to resolve the issue.

9.6.11. Open the method *FTA cards extraction pgx 6 plates_span8_1000µL*, in the PGx folder of the Biomek software.

Note: If robotic workstation 1000 uL conductive sterile filter tips are not available, the method *FTA cards extraction pgx 6 plates_span8_200µL* can be used.

9.6.12. In a section of a Quarter Reservoir, add 22.5 mL of *Generation DNA Purification Solution #1.* The number of sections filled should be the same as the number of plates to be extracted.

9.6.13. In a section of a Quarter Reservoir, add 15 mL of *Generation DNA Elution Solution #2.* The number of sections filled should be the same as the number of plates to be extracted.

9.6.14. For the disposition of all materials and solutions, refer to figure 1 "*Deck Layout of the Biomek FTA cards extraction pgx 6 plates_span8_1000µL"* (or figure 2 *"FTA cards extraction pgx 6 plates_span8_200µL")*.

9.6.15. The deck layout will be slightly different if there are less than 6 plates to be extracted.

9.6.16. Use tip boxes lids for the 2 waste containers.

9.6.17. Press "*Run"* (green arrow) to start the method.

9.6.18. Enter the number of plates to be extracted.

9.6.19. Make sure that the deck layout is the same as the layout shown on the screen. If it is the same, click *"OK"*. The method will start.

Note: To immediately stop the robot, cross the UV light curtain with your hand. Make corrective actions if necessary. Document and record any observation, abnormality or incident in the LIMS.

9.6.20. At the end of the run, do a visual check of the plates. Any missing spot should be considered as a failed sample. Document and record the event in the LIMS.

9.6.21. Apply a Thermo-Seal Heat Sealing Foil on each plate and seal them using a heating plate sealer.

9.6.22. Centrifuge the plate ~30 seconds at 2000 RPM.

9.6.23. Put the plate into a 96-well thermocycler and run the following program:

99ºC – 15 min

4 ºC for ever

9.6.24. Once the cycling program is completed, centrifuge the plate ~30 seconds at 2000 RPM.

9.6.25. Do a visual check of the plate. Document and record any evaporation or sample abnormalities in the LIMS

9.6.26. Label a new 1.2 mL 96-well plate as a *DNA Plate* with the barcode previously printed (see section 9.5.19).

9.6.27. Remove very carefully the Thermo-Seal Heat Sealing Foil of the first extraction plate.

9.6.28. Using a multi-channel pipette, mix the first row of samples 10 times and transfer ~30 uL to the corresponding wells of the new *DNA Plate*.

9.6.29. Repeat the previous step for the remaining rows of the extraction plate.

Note: Make sure that there is no spot transferring on to the DNA plate.

9.6.30. Seal the *DNA Plates* with a plastic adhesive film and get rid of the now empty extraction plate.

9.6.31. The plate can be stored up to three days at +4ºC. For longer storage, store the plate at -20 ºC.

9.7. Return of FTA Cards boxes.

Note: The FTA card boxes should be returned to Héma-Québec within a month following their reception.

9.7.1. Using a cardboard box, package all FTA card boxes to be returned.

9.7.2. Go to the Fedex web site and schedule a pickup.

9.7.3. Print the shipment receipt 3 times and keep them with the box.

9.7.4. Forward the package to the shipping deck of the Montreal Heart Institut.

Reference Documents

Biomek FX Laboratory Automation Workstation Quick Start Guide

Biomek FX Laboratory Automation Workstation User's Manual

Biomek FX Sotfware User's Manual

Generation DNA purification from Blood Spots protocol supplement (now Qiagen

BSD-Duet600 Automatic FTA puncher documentation

Biotracker documentation (LIMS)

Thermocycler User's Manual (Thermo Electron Corporation)

Appendices

PGx- PR-013_APP1:	**Appendix 1: *Manual Punching of FTA Cards.* (not included in the Annex A)**
PGx- PR-013_APP2:	**Appendix 2: *Manual DNA Extraction of FTA cards.* (not included in the Annex A)**
PGx- PR-013_APP3:	**Appendix 3: *Naming Scheme for Héma-Québec's Project***

Figure 1: “Deck layout for “FTA cards extraction pgx 6 plates_span8_1000µL”

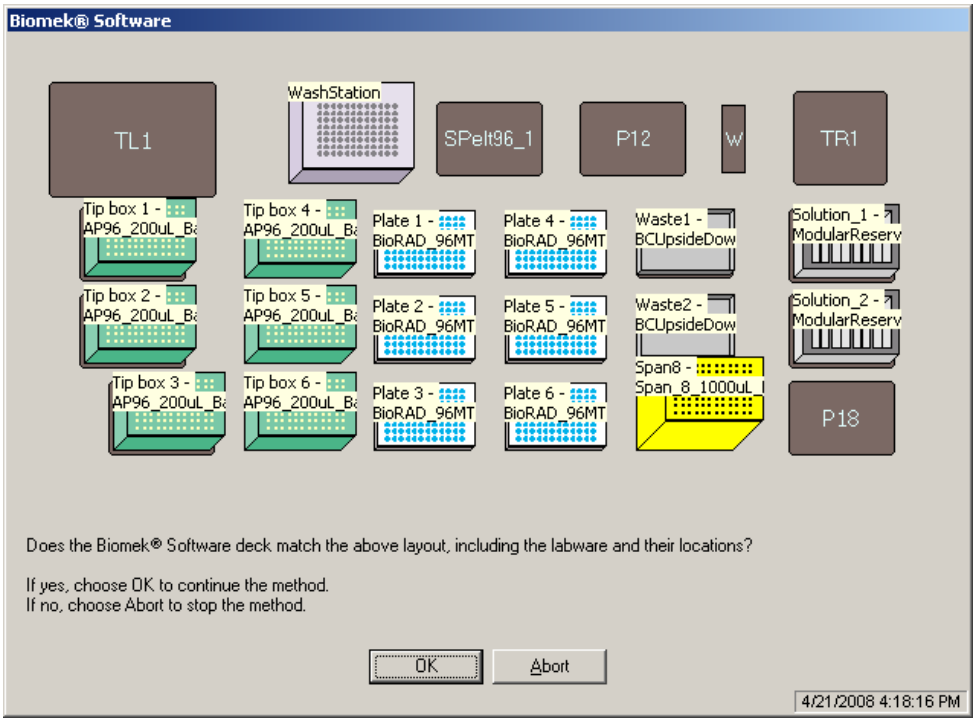
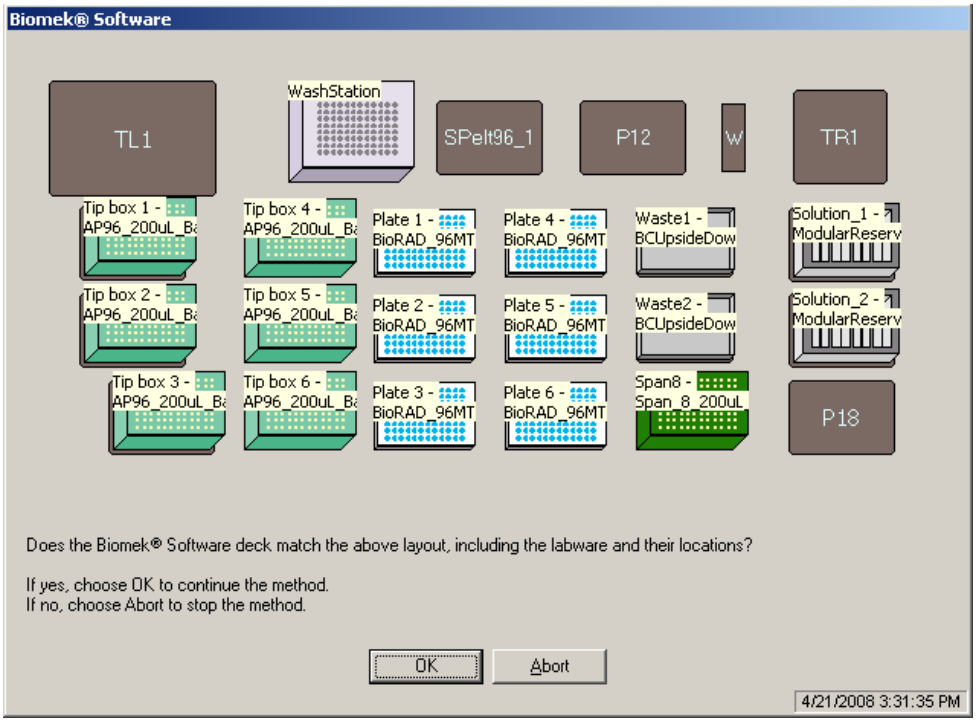
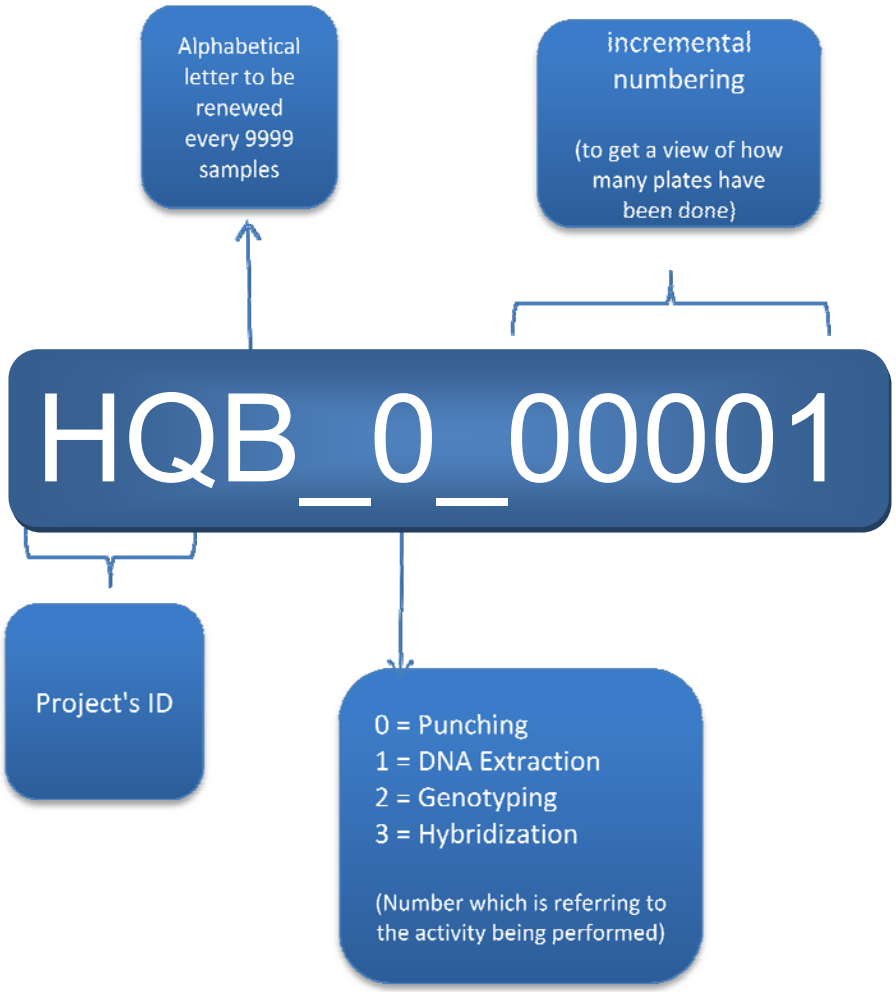


Figure 2: Deck layout for “FTA cards extraction pgx 6 plates_span8_200µL”





Examples

Punching Plates:

HQB_0_00001, HQB_0_00002...etc.

DNA Plates:

HQB_1_00001, HQB_1_00002...etc.

Genotyping Plates:

HQB_2_00001_00002,

HQB_2_00003_00004...etc.

Hybridization Plates:

HQB_3_00001_00002,

HQB_3_00003_00004...etc.

Plate Maker file:

HQB_0_00001_00002,

HQB_0_00003_00004...etc.

Annex B

This Standard Operating Procedure is currently used at the Genome Quebec and Montreal Heart Institute Pharmacogenomics Centre as reference document for the clinical genotyping of the minor blood antigens. It described all steps from the DNA transfer to the assay plate to the results reporting to Héma-Québec. This SOP is referred at the Material and Methods under the clinical application section.

Héma-Québec Minor Blood Antigen Genotyping Assay

Panel#1 on SNPstream Platform

Author Signature:

Signature	Date (dd-mmm-yyyy)

Ian Mongrain, Senior Technician

Reviewer and Approver Signature:

Signature	Date (dd-mmm-yyyy)

Lizette Mailloux, Clinical Technician

1. Purpose

1.1. This SOP defines the procedure for the detection of a panel of 11 nucleotide (22 antigens) polymorphisms (SNPs), (c/C, E/e, K/k, Fy^a/Fy^b, Jk^a/Jk^b, Kp^a/Kp^b, M/N, S/s, HPA-1a/1b, HPA-2a/2b and HPA-5a/5b) found within 9 genetically polymorphic genes ((RHCE (NM020485), KEL(NM_000420), FY(NM_002036), SLC14A1(NM_015865), GYPA(NM_002099), GYPB(NM_002100), ITGB3(NM_000212), GP1BA(NM_000173) and ITGA2(NM_002203)) located on different chromosomes using the SNPstream Genotyping Platform.

1.2. This SOP defines the Genotyping procedure for one ASSAY PLATE of 384 total samples. If more than one ASSAY plate is processed in the same plate, the volume of all reagents has to be adjusted accordingly.

2. Scope

2.1. The variants K/k, Fy^a/Fy^b, Jk^a/Jk^b, Kp^a/Kp^b, M/N, S/s, HPA-1a/1b, HPA-2a/2b and HPA-5a/5b are tested in both DNA directions for calls concordance.

2.2. Since e/E is a G/C variant, two assays need to be performed to detect the presence of G (using the SNPware extension mix G/A) and C (using the SNPware extension mix C/T). Using the analysis tool (see Appendix 8), the final genotype (G/G, G/C and C/C) can be associated for each sample.

2.3. Variants c/C needs two different assays to detect the final polymorphism. The assay c/C detects the presence of c/c, whereas the assay c/C-Intron detects the presence of C/C. All samples that are not assigned C/C or c/c are assigned c/C.

2.4. As with any hybridization-based assay, underlying polymorphisms in primer-binding regions can affect the alleles being detected and thus can subsequently adversely affect the resulting calls being made.

3. Abbrevitions

3.1	PGx	:	Pharmacogenomics
3.2	LIMS	:	Laboratory Information Management System
3.3	DNA	:	Deoxyribonucleic acid
3.4	PCR	:	Polymerase chain reaction
3.5	DCN	:	Document Control Number

Contents

5. Definitions:

5.1. **Tipmix:** Using a pipette**,** mix with up and down action.

6. Responsibilities

6.1. PGx Centre Management personnel and QA representatives.

6.1.1. Ensure that all PGx employees having to genotype HQ samples were trained on the SNPstream instrument and on the full content of this SOP.

6.1.2. Maintain and, when required, review and approve this SOP.

6.1.3. Any of these responsibilities can be delegated to the appropriate person provided he/she was trained on this SOP.

7. Safety Precaution

7.1. Apply the hygiene and gowning requirements of SOP PGx-IG-016 *Material, Sample and Personnel Flow and Hygiene*.

8. Procedure

Equipment

- GenomeLab SNPstream Genotyping System (Beckman)
- Biomek FX Laboratory Automation workstation
- Biomek Software
- Thermocycler
- Incubator (can reach 42°C)
- Ice Bucket
- Vortex
- Vacuum Pump
- Microtube Centrifuge
- Plate Centrifuge
- Multichannel Pipettes (2-20 µL)
- Multichannel Pipettes (20-200 µL)
- Pipettes (P10, P20, P200, P1000)
- Barcode Reader
- LIMS
- Egel System (Invitrogen cat. EBMO3)
- Gel documentation system

Consumables

- PCR plate: 384-well format (Greiner cat.41-785201)
- 1.2 mL 96-well plate
- 50 ml Reservoir Basin
- Compressed Air Can
- 15 mL and 50mL tubes
- 2.0 mL microtube
- 500 uL microtube
- Adhesive plastic film
- Aerosols Resistant tips for Pipettes P10, P200 and P1000

- Progene filter tips 50 uL for Biomek FX
- Progene filter tips 20 uL for Biomek FX
- Progene filter tips 10 uL for Biomek FX
- Progene filter tips 200 uL for Biomek FX
- Kimwipes

Chemicals

- Distilled water PCR grade Rnase/Dnase –free
- Exonuclease-1 (USB: 70073X)
- Qiagen HotStart taq enzyme 5 U/ µl
- Trackit 50bp Ladder (Invitrogen, 10488-043)
- Shrimp Alkaline Phosphatase (USB: 70092X)
- 70% Ethanol
- Ice
- 4% Pre-Cast Agarose Gel (Invitrogen cat. G800804)
- Héma-Québec panel#1 PCR oligos
- Héma-Québec panel#1 Extension oligos
- SNPware 12-plex T/C extension mix (10104500)
- SNPware 12-plex G/A extension mix (10104400)
- SNPstream SNPware Core reagent Kits (Kits A and B and 10 SNPstream array plates)
- SNPstream Wash Buffer #1 (Non-Stringent Buffer)
- SNPstream Wash Buffer #2 (Stringent Buffer)
- Controls DNA Coriell Samples: NA17175, NA17223, NA17275, NA17284, NA17294, NA17282
- Qiagen 10X PCR buffer
- Qiagen 25 mM $MgCl_2$
- Invitrogen 10X PCR buffer- No $MgCl_2$

8.1. Generalities

8.1.1. Delimited zones for DNA extraction, pre-PCR and post-PCR are to be used accordingly to the step being performed. This will avoid any carry-over contamination.

8.1.2. Every technical step performed has to be documented and recorded in the LIMS.

8.1.3. Every consumable, chemical and equipment has to be documented and recorded in the LIMS.

8.1.4. Lab coat and gloves should be worn at all times.

8.1.5. Visual inspection of the plates has to be done following every step of the genotyping technique. Any evaporation, observation or abnormality has to be documented in the LIMS and the Tracking Sheet.

8.1.6. Every automated step can be manually performed using a multichannel pipette.

8.2. Creating a Genotyping Run

8.2.1. Open a copy of the"Tracking Sheet" found in: KT/Shared/Héma-Québec/Tools/.

8.2.2. Open the "plate_layout" tab and fill the appropriate information using the corresponding Plate-Maker file (refer to SOP PGx-PR-013, section 9.3).

Areas to be filled
Plate Name
Plate Barcode
DNA PGx_ID column
Client Sample ID column
Sample Type

8.2.3. Save the tracking sheet to: R:\Commun\Héma-Québec\Sherpa\TrackingSheet following the naming scheme described in PGx-PR-013_App1.

8.2.4. In the LIMS, create a genotyping run in "Experiment Manager". Use the *Genotyping Assay on SNPStream Platform* protocol.

8.2.5. Verify in the LIMS which distilled water samples (e.g. WT000PGX001) and diluted Coriell DNA Controls (e.g. BA000PGX0001) are available (refer to Appendix 1 for Coriell Control Preparation).

8.2.6. Add samples, water and controls to the run. The samples will be associated to the Assay plate using the "Plate Management" option in the LIMS.

8.2.7. Open the Assay_PlateSampleOrder tab of the Plate-Maker. Fill the green cells of the Assay_Plate table with the appropriate PGx barcodes (refer to Appendix 2 for the layout).

8.2.8. Go to *AssayPlate_Import* tab and save it as a csv file in R:\Commun\Héma-Québec\Assay_Plate_Management.

8.2.9. Using "Plate Management", import the previously created file into the LIMS. Print the corresponding barcode.

8.2.10. Label a PCR 384-well plate with the previously created barcode generated by the LIMS. Record the *Assay Plate* name and its barcode number in the "Comments" section of the genotyping run.

8.3. DNA transfer to the Assay Plate

8.3.1. Retrieve from storage the DNA plates and diluted Coriell controls and let them thaw at room temperature.

8.3.2. Quick spin the *DNA Plates* (at 2000 RPM for at least 5 seconds).

8.3.3. Vortex the plates with the Biomek robot. Use the DNA_Mix_BioRad_plate protocol.

N.B.: The mixing of the samples can be done manually. To do so, tipmix gently ~5 times before performing the DNA transfer to the Assay plate (use a pipette at a volume greater than 10 µL).

8.3.4. Transfer 2 µl of each DNA samples to the Assay Plate using the Biomek Robot (refer to Appendix 2, 9 and 10 for the Héma-Québec *Assay Plate* layout, the BiomekFX deck layout and the DNA transfer BiomekFx script description respectively).

8.3.5. Add manually to the assay plate, 2 µl of water and Coriell controls to their corresponding wells (refer to Appendix 2 for the Héma-Québec *Assay Plate* layout).

8.3.6. Gently tap the PCR plate on the working bench to bring all the samples at the bottom of their well. Seal the plate using an adhesive plastic film.

8.3.7. Quick spin the *Assay Plate* (at 2000 RPM for at least 5 seconds).

8.4. **Multiplex PCR reaction**

8.4.1. Thaw one Multiplex PCR Master Mix. Vortex and quick spin the tube. (Refer to Appendix 3 for Multiplex PCR oligos preparation and Appendix 4 for Multiplex PCR Master Mix preparation).

8.4.2. Add 28.8 µl of Qiagen HotStart Taq enzyme (5 U/ µl) to the Multiplex PCR Master Mix preparation. Tipmix gently using P1000 pipette. Avoid any foaming.

8.4.3. Pour the PCR Master Mix Preparation into a reservoir basin.

8.4.4. Using a multichannel pipette, add 3 µl of the Multiplex PCR Master Mix solution to each and every well of the *Assay Plate*. Avoid any contact with the DNA samples.

N.B.: The Biomek robot can be used for the addition of the Multiplex PCR Master Mix solution. For time consuming reasons, the automated application is only optional.

8.4.5. Gently tap the Assay plate on the working bench to bring all the samples to the bottom of the wells. Seal the plate using an adhesive plastic film.

8.4.6. Quick spin the *Assay Plate* (at 2000 RPM for at least 5 seconds).

8.4.7. Put the plate into a 384-well thermocycler and run the following program:

Multiplex PCR_HQA-1:

T°C	Time
95	15 min
94	30 sec
55	55 sec
72	45 sec
72	5 min
4	Hold

8.4.8. After completion of the thermocycler program, quick spin the *Assay Plate* at 2000 RPM for at least 5 seconds.

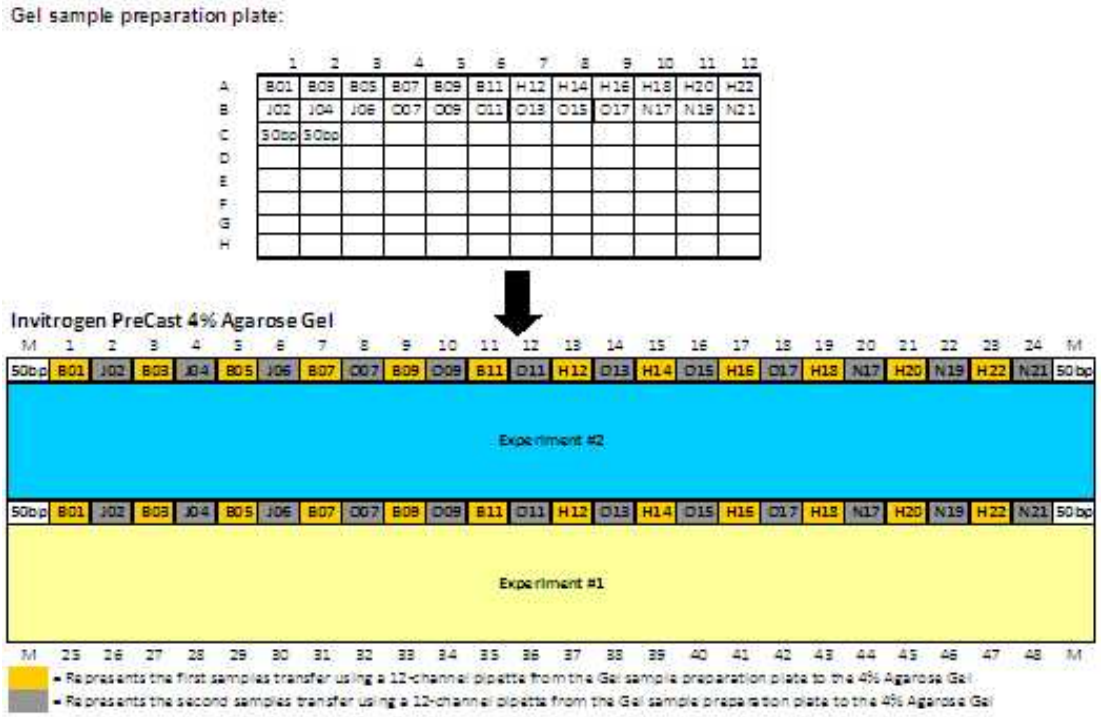
8.4.9. Do a visual check of the plate. Document and record any evaporation or sample abnormalities in the LIMS and the Tracking Sheet.

8.5. Amplification Verification

8.5.1. Remove the adhesive plastic film and transfer to a sample preparation plate, 1 µl of the following PCR samples reactions (see layout below): B01, B03, B05, B07, B09, B11, H12, H14, H16, H18, H20, H22, J02, J04, J06, O07, O09, O11, O13, O15, O17, N17, N19, N21.

8.5.2. Transfer 2 µl of the Track it 50 bp Ladder into the sample preparation plate (see layout below).

8.5.3. Add respectively 14 µl and 18 µL of distilled water (PCR grade) to each sample and ladder. Tipmix and load 15 µL of the mixture to a 4% Pre-Cast Invitrogen Agarose Gel (see layout represented in the figure below).



8.5.4. Set the time on the Egel box for 15 minutes (in EG mode).

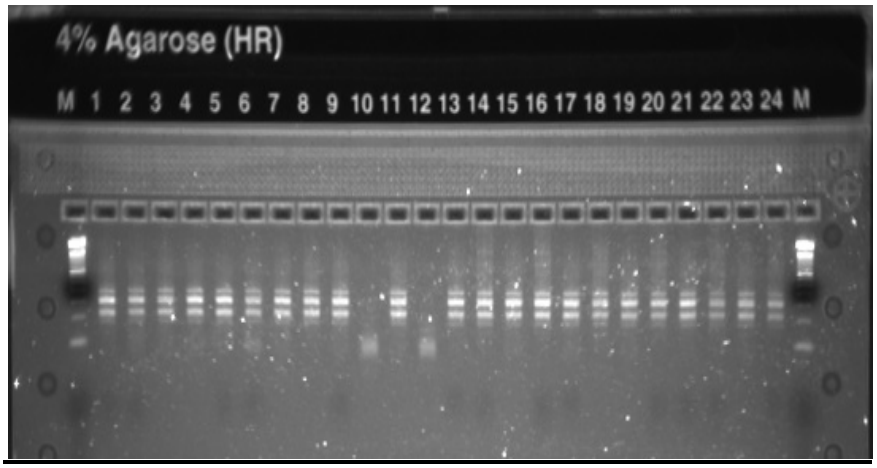
8.5.5. Place the Agarose Gel under UV light and take a picture (refer to the documentation on the Gel Picture Acquisition System for instructions).

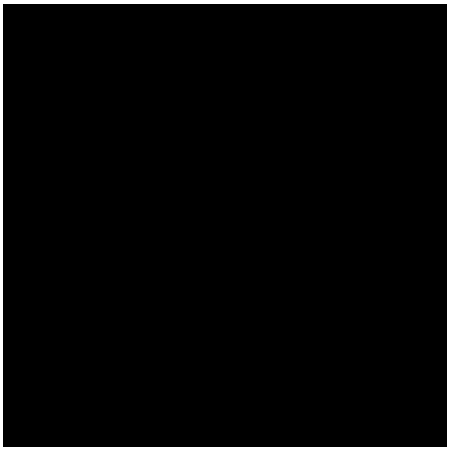
8.5.6. Verify band profiles before pursuing the technique (refer to gel picture below for an example). Refer to Table 1 for amplification acceptance criteria.

Table 1: Amplification acceptance criteria

Condition	Description	Action
Condition 1	All samples and controls show a positive amplification band profile. No contamination seen in the FTA blank (well: O09) nor in the PCR water (well: O11)	Genotyping assay to be pursued
Condition 2	No amplification band profile for more than 4 samples	Extraction and PCR on all samples and controls must be performed again.
Condition 3	No amplification band profile for more than 1 control (wells: O13, O15 and O17)	If condition 2 is not applicable, PCR on all samples and controls must be performed again. _____ If condition 2 is applicable, extraction and PCR on all samples and controls must be performed again.
Condition 4	Amplification band profile seen in FTA blank (well: O09) and not in PCR water (well: O11)	Extraction and PCR on all samples and controls must be performed again.
Condition 5	Strong amplification band profile seen in FTA blank (well: O09) and in PCR water (well: O11)	PCR on all samples and controls must be performed again.
Condition 6	Weak amplification band profile seen in PCR water and not in FTA blank.	Continue genotyping assay, but record it in the LIMS.

Gel Picture





8.6. EXO-SAP Reaction:

8.6.1. Prepare EXO-SAP Mix as described in the table below:

EXO-SAP Reaction

Reagents	Lot#	Vol/Sample	Qte/Conc	1 plate 480 Total vol
10X PCR buffer (10X) No $MgCl_2$		0.30	0.375X	144
$MgCl_2$ (25mM)		0.12	0.375mM	58
SAP (1U/µl)		0.330	0.041 U/µl	158
EXO (10U/µl)		0.067	0.084 U/µl	32
PCR grade Water		2.18		1048
	Total	3.00		1440

N.B.: In the case where more than one *Assay Plate* is being processed, the volumes described in the above table should be adjusted accordingly to the number of *Assay Plate*s.

8.6.2. TipMix the EXO-SAP solution and pour into a reservoir basin.

8.6.3. Using a multichannel pipette, add 3 µl of the EXO-SAP solution to each and every well of the *Assay Plate.* Avoid any contact with the PCR mixture.

N.B.: The Biomek robot can be used for the addition of the EXO-SAP solution. For time consuming reasons, the automated application is only optional.

8.6.4. Gently tap the Assay plate on the working bench to bring all the samples to the bottom of the wells. Seal the plate using an adhesive plastic film.

8.6.5. Quick spin the *Assay Plate* (at 2000 RPM for at least 5 seconds).

8.6.6. Put the plate into a 384-well thermocycler and run the following program:

EXO-SAP Program

T°	Time
37	30 minutes
99	10 minutes
4	Hold

8.6.7. After completion of the thermocycler program, quick spin the *Assay Plate* at 2000 RPM for at least 5 seconds.

8.6.8. Do a visual check of the plate. Document and record any evaporation or sample abnormalities in the LIMS and the Tracking Sheet.

8.7. **Extension reaction:**

8.7.1. Thaw one G/A Multiplex Extension Oligos Mix and one C/T Multiplex Extension OligosMix per *Assay Plate* (refer to Appendix 5 for Extension Oligos Mix Preparation).

8.7.2. Prepare the Multiplex G/A and C/T Extension Master Mix solutions as described in the following tables:

Table 1: GA Extension Master Mix Preparation:

HQA_Panel#1_Multiplex GA Extension Master Mix				1 Plate: 240
Reagents	Lot#	Vol (µL)/Sample	Qte/Conc	Total vol (µL)/
SNPware Extension Mix Diluent		3.77	6.7mM MgCl2 and 65.6mM Tris-Cl	905
Distilled Water, PCR Grade		2.95		709
SNPware 12-plex Extension Mix GA		0.200		48
GA Multiplex Extension Oligos Mix (5 -10µM)		0.06	0.02µM-0.04µM	14
SNPware DNA polymerase		0.017		4.08
	Total	7.00		1680 µL

Table 2: CT Extension Master Mix Preparation:

HQA_Panel#1_Multiplex CT Extension Master Mix				1 Plate: 240
Reagents	Lot#	Vol (µL)/Sample	Qte/Conc	Total vol (µL)
SNPware Extension Mix Diluent		3.77	6.7mM MgCl2 and 65.6mM Tris-Cl	905
Distilled Water, PCR Grade		2.95		709
SNPware 12-plex Extension Mix CT		0.200		48
CT Multiplex Extension Oligos Mix (5 -10µM)		0.06	0.02µM-0.04µM	14
SNPware DNA polymerase		0.017		4.08
	Total	7.00		1680 µL

N.B.[1]: The Extension Mix SNPware 12-Plex CT and GA are light sensitive. Cover the reagents with aluminum foil.

N.B.[2]: In the case where more than one *Assay Plate* is being processed, the volumes described in the above tables should be adjusted accordingly to the number of *Assay Plate*s.

8.7.3. Remove the adhesive plastic film from the *Assay Plate*.

8.7.4. Tipmix the G/A Extension Master Mix solution and pour into a reservoir basin.

8.7.5. Using a multichannel pipette, add 7.0 µl of the G/A Extension Master Mix solution to the corresponding wells (refer to Appendix 2 for the Héma-Québec assay plate layout). Avoid any contact with PCR mixture.

N.B.: The Biomek robot can be used for the addition of the G/A and C/T Extension Master Mix solutions. For time consuming reasons, the automated application is only optional.

8.7.6. Repeat steps 8.7.4. and 8.7.5. with the C/T Extension Master Mix solution.

8.7.7. Gently tap the Assay plate on the working bench to bring all the samples to the bottom of the wells. Seal the plate using an adhesive plastic film.

8.7.8. Quick spin the *Assay Plate* (at 2000 RPM for at least 5 seconds).

8.7.9. Put the plate into a 384-well thermocycler and run the following program:

Extension Program:

T°	Time	
96	3 minutes	
94	20 seconds	45 Cycles
40	11 seconds	
4	Hold	

8.7.10. After completion of the thermocycler program, quick spin the *Assay Plate* at 2000 RPM for at least 5 seconds

8.7.11. Do a visual check of the plate. Document and record any evaporation or sample abnormalities in the LIMS and the Tracking Sheet.

8.8. **Hybridization Reaction:**

8.8.1. Using "Plate Management", import the *Assay Plate* import file previously created into the LIMS (see section 8.2.) to create the *Detection Plate*. Print the corresponding barcode. (Refer to PGx-PR-013_App1 for the naming scheme).

8.8.2. Label a 12-Plex SNPstream array plate with the previously created barcode. Record the *Detection Plate* name and also the LIMS and SNPstream barcode numbers in the "Comments" section of the genotyping run.

8.8.3. Pour in a reservoir basin ~40 ml of Non-Stringent Buffer (Buffer #1) (refer to Appendix 6 for the preparation of "Non-Stringent and Stringent Buffers).

8.8.4. Transfer ~20 µl of Non-Stringent Buffer (avoid touching the plate glass) into the array plate wells. Tap gently on all sides of the plate to cover the entire well surface.

8.8.5. Remove the buffer using a vacuum pump system.

8.8.6. Repeat steps 8.8.4 and 8.8.5. two more times.

8.8.7. Prepare the Hybridization solution as described in the following table:

Hybridization Solution Preparation:

Reagents	Lot#	Vol (µL)/Sample	Qte/Conc	1 Plate: 480 Total vol (µL)
SNPware Hybridization Solution		7.56	1.64M NaCl, 164mM EDTA,	3629
SNPware Hybridization Additive		0.44	190.6mM MES	211
Total		8.00		3840 (µL)

N.B.: In the case where more than one *Assay Plate* is being processed, the volumes described in the above table should be adjusted accordingly to the number of *Assay Plate*s.

8.8.8. Using a multichannel pipette, add 8 µl of Hybridization Solution to each and every well of the *Assay Plate.* Avoid any contact with the PCR mixture.

N.B.: The Biomek robot can be used for the addition of Hybridization solution. For time consuming reasons, the automated application is only optional.

8.8.9. Using the BiomekFX robot, transfer 14 µl from the *Assay Plate* to the SNPstream 12-plex Array plate (refer to Appendix 9 for the BiomekFX deck layout and Appendix 10 for the BiomekFX Liquid transfer: Genotyping on SNPstream).

8.8.10. Gently tap the Detection Plate on each side in order to cover the entire well with liquid.

8.8.11. Place the *SNPstream 12-plex Array plate* into a hermetically sealed humid container and put it into a hybridization oven set at 42ºC for 2 hours.

8.8.12. Retrieve the plate from the oven and do a visual check of the plate. Document and record any evaporation or sample abnormalities in the LIMS and the Tracking Sheet.

8.8.13. Using the vacuum pump system, remove all mixture from the SNPStream 12-plex Array Plate.

8.8.14. Pour in a reservoir basin ~40 ml of the Stringent Buffer (Buffer #2).

8.8.15. Transfer ~20 µl of Stringent Buffer (avoid touching the plate glass) into the array plate wells. Tap gently on all sides of the plate to cover the entire well surface.

8.8.16. Remove the buffer using a vacuum pump system.

8.8.17. Repeat steps 8.8.15 and 8.8.16. two more times.

8.8.18. Dry the plate thoroughly using compressed air (if necessary).

8.8.19. Wash the back of the plate with 70% EtOH or isopropanol (if necessary).

8.8.20. Put back the Array Plate into its aluminum pouch until ready for the scanning procedure.

8.8.21. The tracking sheet should then be added to the genotyping run of the LIMS under the section: Output/File.

8.9. SNPStream Scanning Procedure and Data Acquisition

8.9.1. Refer to Appendix 7 for complete details.

8.10. Interpretation of results

8.10.1. For cluster analysis, data analysis and data reporting, refer to Appendix 8 for complete details.

9. Reference Documents

9.1. Biomek FX Laboratory Automation Workstation Quick Sart Guide

9.2. Biomek FX Laboratory Automation Workstation User's Manual

9.3. Biomek FX Sotfware User's Manual

9.4. BSD-Duet600 autoFTA puncher

9.5. Biotracker LIMS software Manual

10. Appendices

10.1. PGx_PR-012_APP-1 Coriell Control preparation

10.2. PGx_PR-012_APP-2 Héma-Québec Assay Plate layout.

10.3. PGx_PR-012_APP-3 Multiplex PCR Oligos Preparation

10.4. PGx_PR-012_APP-4 Multiplex PCR Master Mix Preparation

10.5. PGx_PR-012_APP-5 Extension Oligos Mix Preparation

10.6. PGx_PR-012_APP-6 SNPstream Stringent and Non-Stringent Buffers preparation.

10.7. PGx_PR-012_APP-7 Raw data acquisition using the SNPstream technology and software

10.8. PGx_PR-012_APP-8 Data Analysis and Data reporting

10.9. PGx_PR-012_APP-9 Robotic Liquid transfer: BiomekFX deck layout

10.10. PGx_PR-012_APP-10 BiomekFX Liquid transfer: Genotyping on SNPstream

Héma-Québec Coriell Control plate preparation: Freshly prepared after 10 freeze-thaw cycles.

1. Create in the LIMS two PGx ID for each Coriell control. This way, each Coriell control will have two different barcode numbers to facilitate data analysis.
2. Print the barcodes and label a plate.
3. Prepare Control samples following the plate layout below (figure 1).
4. Prepare 12.5 ng/µl « NA » samples using distilled water (PCR grade) as diluent. Make dilution as per the following example:
    a. Prepare 110 µl of Controls samples: if stock concentrations of « NA » samples = 300 ng/µl, pipette 4.6 µl into the appropriate wells and add 105.4 µl of PCR grade water. Tipmix the wells A01 to A06 10 times.

N.B.: Refer to DNA quantification in case the « NA » DNA samples concentrations are unknown.

5. Place a silicone mat on the storage plate.
6. The Control samples are stored at -20ºC. The plate can be freeze-thawed up to 10 times.

Figure 1 :

Control samples plate:

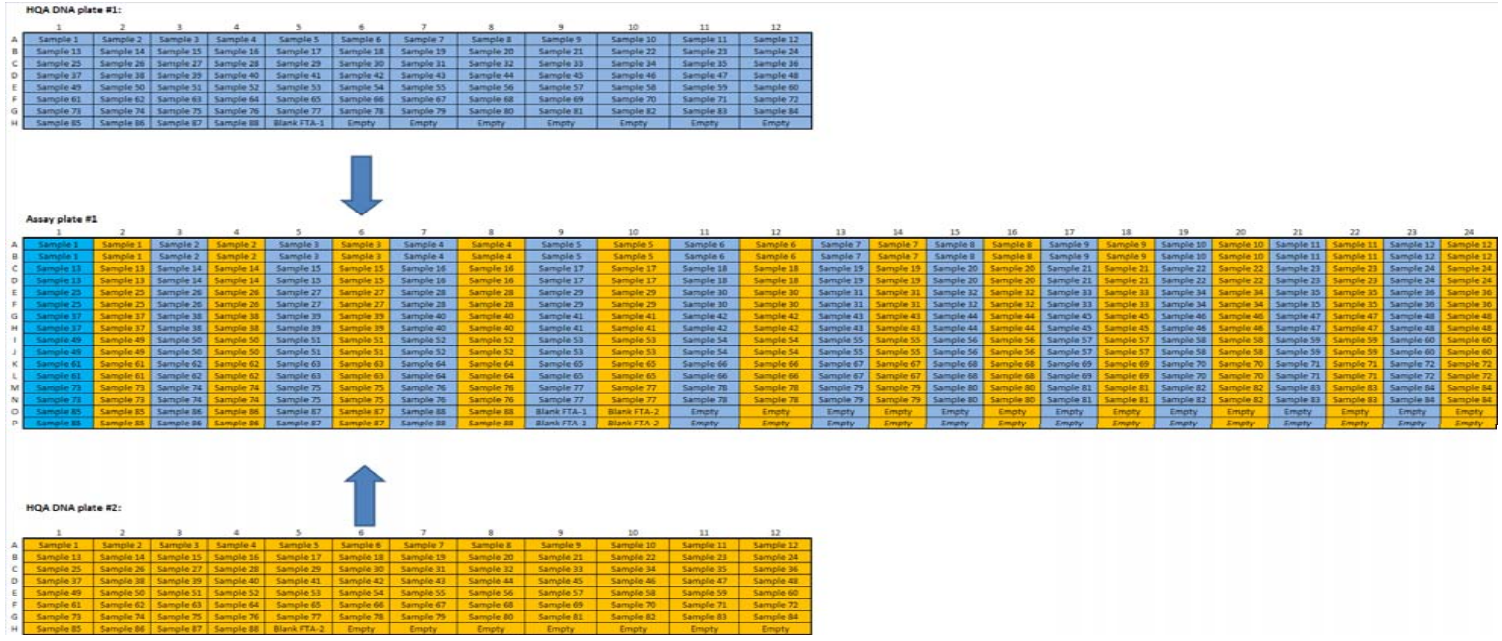
	1	2	3	4	5	6	7	8	9	10	11	12
A	NA17175	NA17223	NA17275	NA17284	NA17294	NA17282						
B												
C												
D												
E												
F												
G												
H												

Héma-Québec assay plate layout:

1. **Transfer of test DNA samples to an Assay Plate**

1.1 Transfer 2 µl of DNA samples prepared as per procedure PGx-PR-013 to the *Assay Plate* following the scheme below. These samples are referred as "SAM" for the data analysis.

1.2 Keep the wells entitled "empty" for Water and Control DNA samples (refer to appendix PGx-PR-012_APP1) for control samples preparation).



2. Transfer of Control Samples to an assay plate

2.1 Transfer 2 µl of Control samples to the *Assay Plate*, following the scheme below (refer to appendix PGx-PR-012_APP1 for control samples preparation). These samples are referred as "SYN" for the data analysis.

2.2 Transfer 2 µl of Water control samples to the *Assay Plate* following the scheme below.

Control samples plate:

	1	2	3	4	5	6	7	8	9	10	11	12
A	NA17175	NA17223	NA17275	NA17284	NA17294	NA17282						
B												
C												
D												
E												
F												
G												
H												

Assay plate #1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
B	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
C	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
D	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
E	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
F	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
G	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
H	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
I	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
J	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
K	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
L	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
M	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
N	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
O	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282
P	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282

Quadrants Layouts:

The wells A01, A03… C01, C03… E01, E03… represent quadrant #1

The wells A02, A04… C02, C04… E02, E04… represent quadrant #2

The wells B01, B03… D01, D03… F01, F03… represent quadrant #3

The wells B02, B04… D02, D04… F02, F04… represent quadrant #4

3. Transfer of PCR Master Mix Preparation to *Assay Plate*

3.1. Refer to the Quadrants layouts to understand transfer technique.

3.2. Transfer 3.0 µl of PCR Master Mix preparation to the entire *Assay Plate* (Quadrants 1, 2, 3 and 4).

Assay plate #1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
B	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
C	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
D	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
E	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
F	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
G	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
H	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
I	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
J	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
K	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
L	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
M	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
N	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
O	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282
P	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282

4. Transfer of Extension Master Mix Preparation to *Assay Plate* (see Figure below for visual details)

4.1. Refer to the Quadrant layout to understand transfer technique

4.2. Transfer 7 µl of the Extension GA Master Mix preparation to the Quadrants #1 and #2

4.3. Transfer 7 µl of the Extension CT Master Mix preparation to the Quadrants #3 and #4

Assay plate #1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
B	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3	Sample 4	Sample 4	Sample 5	Sample 5	Sample 6	Sample 6	Sample 7	Sample 7	Sample 8	Sample 8	Sample 9	Sample 9	Sample 10	Sample 10	Sample 11	Sample 11	Sample 12	Sample 12
C	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
D	Sample 13	Sample 13	Sample 14	Sample 14	Sample 15	Sample 15	Sample 16	Sample 16	Sample 17	Sample 17	Sample 18	Sample 18	Sample 19	Sample 19	Sample 20	Sample 20	Sample 21	Sample 21	Sample 22	Sample 22	Sample 23	Sample 23	Sample 24	Sample 24
E	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
F	Sample 25	Sample 25	Sample 26	Sample 26	Sample 27	Sample 27	Sample 28	Sample 28	Sample 29	Sample 29	Sample 30	Sample 30	Sample 31	Sample 31	Sample 32	Sample 32	Sample 33	Sample 33	Sample 34	Sample 34	Sample 35	Sample 35	Sample 36	Sample 36
G	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
H	Sample 37	Sample 37	Sample 38	Sample 38	Sample 39	Sample 39	Sample 40	Sample 40	Sample 41	Sample 41	Sample 42	Sample 42	Sample 43	Sample 43	Sample 44	Sample 44	Sample 45	Sample 45	Sample 46	Sample 46	Sample 47	Sample 47	Sample 48	Sample 48
I	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
J	Sample 49	Sample 49	Sample 50	Sample 50	Sample 51	Sample 51	Sample 52	Sample 52	Sample 53	Sample 53	Sample 54	Sample 54	Sample 55	Sample 55	Sample 56	Sample 56	Sample 57	Sample 57	Sample 58	Sample 58	Sample 59	Sample 59	Sample 60	Sample 60
K	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
L	Sample 61	Sample 61	Sample 62	Sample 62	Sample 63	Sample 63	Sample 64	Sample 64	Sample 65	Sample 65	Sample 66	Sample 66	Sample 67	Sample 67	Sample 68	Sample 68	Sample 69	Sample 69	Sample 70	Sample 70	Sample 71	Sample 71	Sample 72	Sample 72
M	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
N	Sample 73	Sample 73	Sample 74	Sample 74	Sample 75	Sample 75	Sample 76	Sample 76	Sample 77	Sample 77	Sample 78	Sample 78	Sample 79	Sample 79	Sample 80	Sample 80	Sample 81	Sample 81	Sample 82	Sample 82	Sample 83	Sample 83	Sample 84	Sample 84
O	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282
P	Sample 85	Sample 85	Sample 86	Sample 86	Sample 87	Sample 87	Sample 88	Sample 88	Blank FTA-1	Blank FTA-2	HQA-Water-1	HQA-Water-2	NA17175	NA17175	NA17223	NA17223	NA17275	NA17275	NA17284	NA17284	NA17294	NA17294	NA17282	NA17282

Assay Plate + GA and CT Extension Master Mixes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
B	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
C	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
D	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
E	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
F	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
G	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
H	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
I	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
J	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
K	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
L	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
M	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
N	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
O	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
P	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT

Héma-Québec SNPstream panel#1 PCR oligos preparation:

The calculation is based on a production of 70 *Assay plates*.

1. Prepare the Multiplex PCR oligos in a 15 mL tubes as described on the Table 1 below:

Table 1: PCR Oligos Preparation:

HQA_Panel#1_Multiplex PCR Oligos Mix

Mutation	Forward Primer ID	Forward Oligos Sequence	Stock Conc (uM)	Forw. Oligos Vol (µl)	Reverse Primer ID	Reverse Oligos Sequence 5' - 3'	Stock Conc (uM)	Reverse oligos Vol. (µl)	H2O vol (µl)	Total vol (µl)	Final conc (µM)	PCR Lenght (Pb)
kK	KelEx6K1_2-F-5	CRGATCCTTATGCTCAGCCC	200	50	KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	200	50			1	157
Kapa	KelExon8Kpa_bF	AGCAAGGTGCAAGAACACT	200	50	KelExon8Kpa_b-R-1	AGAGCTTGCCCTGTGCCC	200	50			1	101
sS	GYPBEx4s_S-F-6	CTGGTACAGTGAAACGATGG	200	50	GYPBEx4s_S-R-6	GCACATGTCTTWCTTATTTGG	200	50			1	170
HPA2ab	GPIBHPA-2a_b-F-1	CTCAGTCAAGTTGTTGTTAGCCAGAC	200	50	GPIBHA-2a_bR-S2	CTCTACCTGAAAGGCAATGAGCT	200	50			1	105
HPA1ab	GPIII3HPA-1a_b-F-1	ATTCTGGGGCACAGTTATCC	200	50	GPIII3HPA-1a_b-R-5	ATAGYTCTGATTGCTGGACTTC	200	50			1	114
Fyab	FYex2Fya_b-F-5	AGWCATCCAGCAGGTTACAGG	200	50	FYex2Fya_BR	AAGATGTATGGAATTCTTCCTATGG	200	50	8800	10000	1	129
HPA5ab	GPIaHPA5-F-1	gctcttggtaggtgcaccaatgt	200	50	GPIaHPA5_R	ttccaaatgcaagttaaattaccag	200	50			1	148
Jkab	JKa_b-F-1	CCTGCTAACTTTCAATCCCACCCTC	200	50	JKa_b_R-s3_1	TGCCAGGTGAGCGCCATGAAC	200	50			1	160
eE	RHCE_e-F-1	TGGATGTTCTGGCCAAGTG	200	50	RHCE_e-R-S2	CTGTCACCACACTGACTGCTAG	200	50			1	107
cC	RHCEc_C-F-1	tccccctcctccttctca	200	50	RHCEc_C-R-1	GCCAGCATGAAGAGGTTGAA	200	50			1	143
cC-intron	RHCintron-F-1	cagggccaccaccatttgaa	200	50	RHCintron_R	tggtagcaggcgtctgtaaaaa	200	50			1	151
MN	GYPAex2M_N-F-1	TGAGGGAATTTGTCTTTTGCA	200	50	GYPAex2M_N-R-4	TCTTTGTGACTGAAGAAGAR	200	50			1	98

N.B: The needed PCR oligos volumes may change depending of the measured Oligos concentration.

2. Mix by vortexing (~5 secondes) the PCR Oligos Preparation.
3. Aliquot two 4300 µl into two 15 mL tubes. Identify the aliquot tubes as: Oligos_PCR_HQA-A and B.
4. Stored the aliquots at -20ºC until used for the Multiplex PCR Master Mix Preparation. These preparations are single use.

Héma-Québec SNPstream panel#1 Multiplex PCR Master Mix Preparation:

N.B.: The calculation is based on a production of 70 Assay plates. Two preparations of this PCR Master Mix will be necessary to cover the 70 Assays Plates.

1. Please refer to PGx-PR-012_APP3 for instructions on the PCR oligos Mix Preparation.

5. Prepare the Multiplex PCR Master Mix in a 50 mL tubes as described in the Table 1 below.

Table 1: (All volumes are in µl)

HQ_Panel#1_Multiplex PCR Master Mix				35 plates
				16800
Reagents	Lot#	Vol/Sample	Qte/Conc	Total vol
10X Qiagen PCR buffer (10X)		0.50	1X, Contain 1.5mM MgCl2	8400
MgCl2 (25mM)		0.60	3.0 mM (Final: 4.5mM)	10080
dNTPs (10mM)		0.038	0.075mM	630
Multiplex PCR Oligos Mix (1µM)		0.25	0.05mM	4200
PCR grade Water		1.55		26082
Total		2.94		49392

3. Mix the 50 mL or the 2.0 mL tube by inversion.

4. Aliquot precisely 1411.2 µl into 35 microtubes of 2.0 mL. Label all tubes with a LIMS barcode.

5. Stored all PCR Master Mix Tubes at -20ºC.

Héma-Québec SNPstream panel#1 Extension Oligos Mix Preparation:

N.B.: The calculation is based on a production of 70 *Assay plates*.

6. Prepare the Multiplex Extensions oligos in a 15 mL tube as described on the Table 1 below:

HQA_Panel#1_Multiplex GA Extension Oligos Mix: SNPstream Assay #: 328

Mutation	GA Extension oligos ID	GA Extension Oligos Sequences 5'-3'	SNPstream 12-plex Array Positions	Stock Conc (uM)	Oligos Vol. (µl)	Vol H2O (µl)	Final Vol. (µl)	Final Conc. (uM)
kK	KelEx6K1_2-GA-U5	GCGGTAGGTTCCCGACATATACTCATCAGAAGTCTCAGC	3_2	100	120			10
Kapa	KelExon8KGA_U1	ACGCACGTCCACGGTGATTTGCCTCAGAAACTGGAACAGCC	2_1	100	60			5
MN	GYPAex2M_NGA-U2_2	GGATGGCGTTCCGTCCTATTGTGCATTGCCACMYCAGTGGTACTT	3_1	100	60			5
sS	GYPBs_S-GA-U7	AGGGTCTCTACGCTGACGATAAACGATGGACAAGTTGTCCC	1_3	100	60			5
HPA2ab	GPIBex2HPA2GA-U10-s3	AGATAGAGTCGATGCCAGCTGAGCTTCTCCAGCTTGGGTGTGGGC	4_3	100	60			5
HPA1ab	GPIIIAex3HPA-1GAU8	GTGATTCTGTACGTGTCGCCCTTCAGGTCACAGCGAGGTGAGCCC	2_3	100	60	360	1200	5
eE	RHCEex5RhE_eCTU3-s3_1	CGTGCCGCTCGTGATAGAATCCTTTGGATTGGACTTCTCAGCAGAG	1_2	100	60			5
cC	RHCEEx2Rhc_CGA-U9	GACCTGGGTGTCGATACCTACCCTTGGCTTGGGCTTCCTCACCTC	3_3	100	60			5
Fyab	FYex2Fya_b-U11	AGAGCGAGTGACGCATACTAGATTCCTTCCCAGATGGAGACTATG	2_4	100	60			5
HPA5ab	GPIaHPA5-GA-U6	GGCTATGATTCGCAATGCTTGTCTACCTGTTTACTATCAAA	4_2	100	60			5
Jkab	JKa_b-U4_1	AGCGATCTGCGAGACCGTATCTCAGTCTTTCAGCCCCATTTGAG	2_2	100	60			5
cC-intron	RHCintron-Gins-U12	CGACTGTAGGTGCGTAACTCGGTGCCCTTTGTCACTTCCCA	3_4	100	120			10

HQA_Panel#1_Multiplex CT Extension Oligos Mix SNPstream Assay #: 329

Mutation	CT Extension oligos ID	CT Extension Oligos Sequences 5'-3'	SNPstream 12-plex Array Positions	Stock Conc (uM)	Oligos Vol. (µl)	Vol H2O (µl)	Final Vol. (µl)	Final Conc. (uM)
kK	KelEx6K1_2U7	AGGGTCTCTACGCTGACGATTGGACTTCCTTAAACTTTAACCGAA	1_3	100	60			5
Kapa	KelExon8Kpa_bU12	CGACTGTAGGTGCGTAACTCTTCCTTGTCAATCTCCATCACTTCA	3_4	100	60			5
MN	GYPAex2M_NCTU11-s3_2	AGAGCGAGTGACGCATACTAACAGRAATTGTGAGCATATCAGCAT	2_4	100	60			5
sS	GYPBEx4s_S-CT-U2-s2	GGATGGCGTTCCGTCCTATTTTGAAATTTTGCTTTATAGGAGAAA	3_1	100	60			5
HPA2ab	GPIBex2HPA-2a_bCTU9	GACCTGGGTGTCGATACCTAGACCCTGCCCCCAGGGCTCCTGA	3_3	100	60			5
HPA1ab	GPIIIAex3HPA1CTU8-s2	GTGATTCTGTACGTGTCGCCCTCTTTGGGCTCCTGTCTTACAGGCCCTGCCTC	2_3	100	60	420	1200	5
eE	RHCEex5RhE_eCTU1-s2_1	ACGCACGTCCACGGTGATTTCCTTTGGATTGGACTTCTCAGCAGAG	2_1	100	60			5
cC	RHCEEx2Rhc_CctU6-s2	GGCTATGATTCGCAATGCTTTGCTCCAGCTGTGTCTCCGGAAA	4_2	100	60			5
Fyab	FYex2Fya_bU4	AGCGATCTGCGAGACCGTATGGGGGCAGCTGCTTCCAGGTTGGCA	2_2	100	60			5
HPA5ab	GPIaHPA5-CT-U3	CGTGCCGCTCGTGATAGAATTTAGTTTATTTTTTTTTTTTTACCT	1_2	100	60			5
Jkab	JKa_b-CT-U5-s2_1	GCGGTAGGTTCCCGACATATGAAACCCCAGAGTCCAAAGTAGATGT	3_2	100	60			5
cC-intron	RHCintron-CT-U10	AGATAGAGTCGATGCCAGCTGGCAGAGGCTGCAATGAGCTATGATTGTAC	4_3	100	120			10

7. Mix by vortexing the Extension Oligos Preparation and quick spin.

8. Aliquot 17 µl in thirty five 500 µl microtubes and ~605 µl into one 2.0 mL microtube. When the first thirty five aliquots are used, freeze-thaw the aliquot of 605 µl and re-aliquot 17 µl in another thirty five 500 µl microtubes. Label all tubes with a LIMS barcode. Store the aliquots at -20ºC until used for the Multiplex Extension Master Mix Preparation. These preparations are single use.

SNPstream Non-Stringent and Stringent Buffers Preparation:

N.B.: Both preparations should be stored at -4ºC for no more than three months.

1. Prepare in 1.0 L bottles as listed in the following tables:

Table 1:

Non-Stringent Buffer:

Reagents	Vol (mL)
SNPStream 20 X Non-Stringent Buffer:	50
ddWater	950
Total	1000

Table 2:

Stringent Buffer:

Reagents	Vol (mL)
SNPStream 64 X Stringent Buffer:	15.6
ddWater	984.4
Total	1000

2. Invert 10 times for complete reagent re-suspension.

Raw data acquisition using the SNPstream technology and software

1. SNPstream Input Excel Tool

N.B.: As guidance, every green cell in the SNPstream Input excel tool corresponds to a cell where information needs to be entered by the user. Other cells are automatically filled or locked.

1.1. From Knowledge Tree, under /Shared/Héma-Québec/Tools/, download a copy of *SNPstream_inputFile_Format3_Template.xls*.

1.2. In the *Sheet 1* tab, fill out the process group which should correspond to the Detection plate name (e.g. HQB_3_00001_00002). Write down number 3 for the format. Refer to *Format_Viewer* tab for format details.

1.3. Open the corresponding plate-maker.

1.4. Copy the BA PGx_Barcode columns of the *FTA_BA-SamplesOrder* tab and paste them to the corresponding PGxSampleID columns of the SNPstream Input *Plate_Layout* tab.

1.5. Copy the ClientSampleID columns of the *Plates_List* tab of the same plate-maker and paste them to the corresponding ClientSampleID columns of the SNPstream Input *Plate_Layout* tab.

1.6. Save the file with the name of the hybridization plate in Knowledge Tree under /Shared/Héma-Québec/Final_SNPstream input/.

1.7. Go to the *Maker* tab and save it as a text (.txt) file in Knowledge Tree under /Shared/Héma-Québec/Final_SNPstream input/.

2. Importation of the Input file in Plate Explorer

2.1. Open the Plate Explorer application of the SNPstream software.

2.2. Click on the *Upload* icon.

2.3. Select the Input (txt) file previously created in section 1.

2.4. Click *Open* and *Upload*.

2.5. In the *Process group* list box, select the process group corresponding to your plate's name. Copy the process group name and paste it in the *PCR plate name*.

2.6. Enter the SNPware plate barcode number in the *SNPware Plate* section.

2.7. Select the following marker panels for the following segments in *Marker Panel(s)* section:

Seg	Panel	Name
1	328	GA: HQA_GA_01OCT2007
2	329	CT: HQA_CT_01OCT2007

2.8. Click the *Save to DB* icon and exit Plate Explorer software.

3. Data Acquisition

3.1. Open the SNPstream instrument.

3.2. Open the Run Manager application. The instrument will automatically initialize. If not, click on *Initialize*.

3.3. Once the initialization is completed, select *Manual* in the *Run* drop down menu.

3.4. Choose *12-Plex* in the Plate type section.

3.5. Choose the *Twocoloradjusted_6Mai2008* in the 12-plex Method section.

3.6. Click on *Eject plate*.

3.7. Insert the hybridization SNPstream plate making sure that the SNPstream barcode is facing the mirror and click on *Insert plate*.

3.8. Once the lasers are warmed up, click *Run.*

3.9. At the end of the raw data acquisition, click on *Eject Plate* and retrieve the hybridization plate. Place it back in its aluminum pouch.

3.1.10. Click on *Insert Plate*.

4. Automatic clustering

4.1. Open the SNPAdmin application of the SNPstream software.

4.2. Click on *Run* and select in the drop down menu *Start* and then *Auto Run*.

The automatic raw data analysis and clustering will start.

Description of the analysis and reporting flow using three different software tools:

1- Cluster analysis using the SNPstream GetGenos
2- Data analysis and formatting using Sherpa software©
3- Data reporting using Area51©

1. Cluster analysis using the SNPstream GetGenos software

N.B.: Two different users are required to go through cluster analysis (User1 and User2).

1.1. Open the *GetGenos* application. Open the *View* tab and then select *QCReview*.

1.2. In the *Qc Review* section, click on *New* and select *SNPplate Number* as the category.

1.3. Select *Plate within 5 days* from the "Select by SNP Plates Number" drop down menu.

1.4. In the drop down menu of "Select Qc plate from list", select the SNPstream barcode of the plate being analyzed. Choose the first segment (GA).

1.5. The software will prompt you to adjust any grid shift if needed. If not, the analysis will resume at step 1.4. In the presence of grid shifts:

1.5.1. Click Yes. Select the appropriate SNPware batch. The two grids from the two color lasers analysis will appear.

1.5.2. Realign the shift grids by using the plate control spots as references (see figure 1 for Assay variation positions on the 12-plex Array plate). Click on *Manual Adjust*.

1.5.3. Click on Update DB.

1.5.4. Repeat steps 1.3.2. and 1.3.3. for each sample showing grid shifts.

1.5.5. Select Segment 2 (CT) and verify if grid shifts are detected. If so, repeat steps 1.3.1. to 1.3.4.

1.5.6. Reprocess the plate.

1.5.6.1. Select *GetGenos from DB* in the *Run* dropdown menu.

1.5.6.2. In the "Run on DB" tab, choose the *Reprocess plate with #* option.

1.5.6.3. Enter the SNPware barcode and click on *Retrieve plate list*.

1.5.6.4. In the "Reprocess a plate" section, click *All segments* if segments 1 and 2 have grid shifts or click on *Selected Segments* and choose the segment that needs to be reprocessed.

1.5.6.5. Click on *Run GetGenos*.

N.B.: Refer to the SNPstream Grid Shift documentation for complete instructions.

1.6. Go back to segment 1 (GA) and perform a manual search for grid shifted samples (the software could have overlooked some grid shifts):

1.6.1. Look for any cluster outlier for XX, XY and YY controls (see figure 1 for Assay variation positions on the 12-plex Array plate).

1.6.2. Select all outlier samples and click on ‘’View Image’’. Make sure the correct SNPstream barcode file is open.

1.6.3. For every observed grid shifted sample, click on ‘’Mark this well as grid shifted’’. If there are no grid shifts, close the images and resume the procedure at step 1.7.

N.B.: Outliers can also be caused by dust particles on the array spot. If dust particles are visible, the decision to fail or pass the assay for that particular well must be taken by the most experience person between User1 and User2 (refer to figure 2 for an example of a dust particle on the array spot).

1.6.4. When all grid shifted samples have been marked as so, click on “Fix Grid Shifted well in this segment” and follow steps 1.5.1. to 1.5.6.

1.6.5. Repeat steps 1.6.1. to 1.6.4. for segment 2 (CT).

1.7. Make a visual evaluation of the plate’s status by clicking on “View whole plate status” (refer to figure 3 for an example). This is an informative tool that could link failed samples to an assay plate pattern (e.g. genotypes missing for a complete row, column failure for A and G results etc.). If a plate pattern is present, refer to the lab manager before pursuing the analysis.

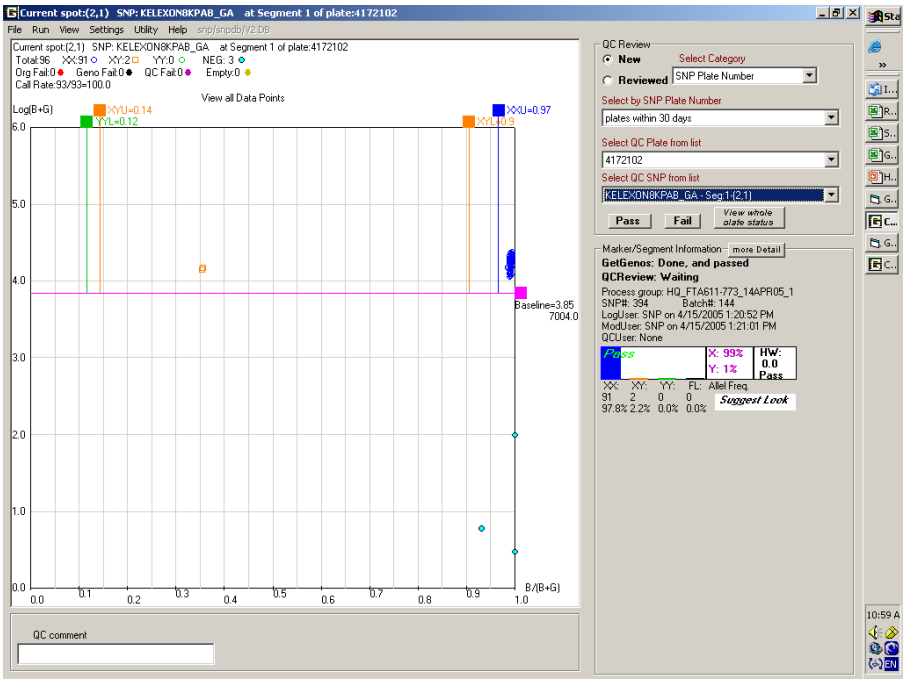
1.8. Perform a cluster analysis as per cluster profiles shown below.

Standard Cluster Analysis:

Segment 1 (GA analysis) :

a. Kapa :

Normal Setting for Cluster Analysis: Polar View: YY => 0.00 to 01 XY => 0.2 to 0.6 XX => 0.9 to 1



b. MN

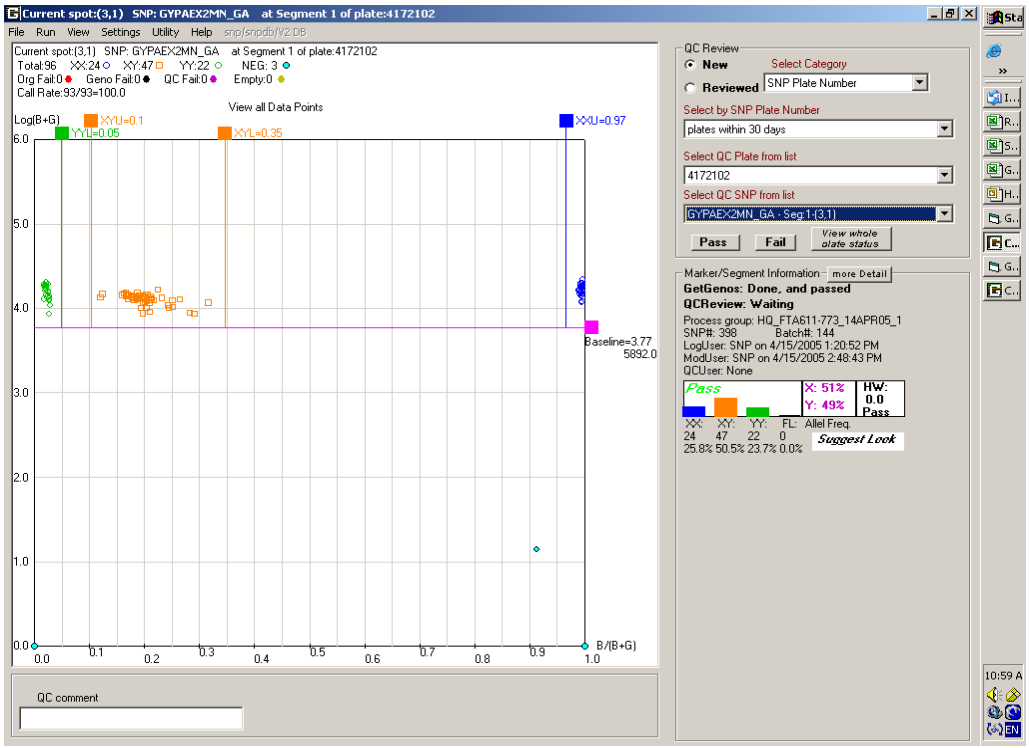
Normal Setting for Cluster Analysis:

Polar View:

YY => 0.00 to 0.05

XY => 0.10 to 0.45

XX => 0.9 to 1



c. eE

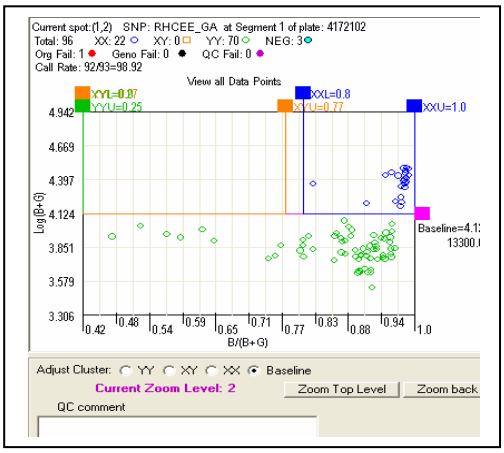
Normal Setting for Cluster Analysis:

Cartesian

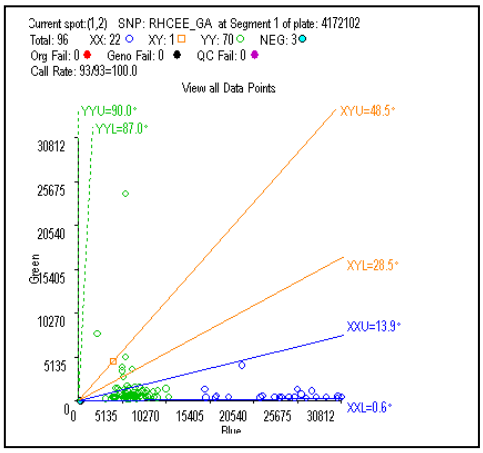
YY < 11000 (AA = ee)

XX > 15000 (GG = eE or EE)

Polar View:



Cartesian:



d. Jkab

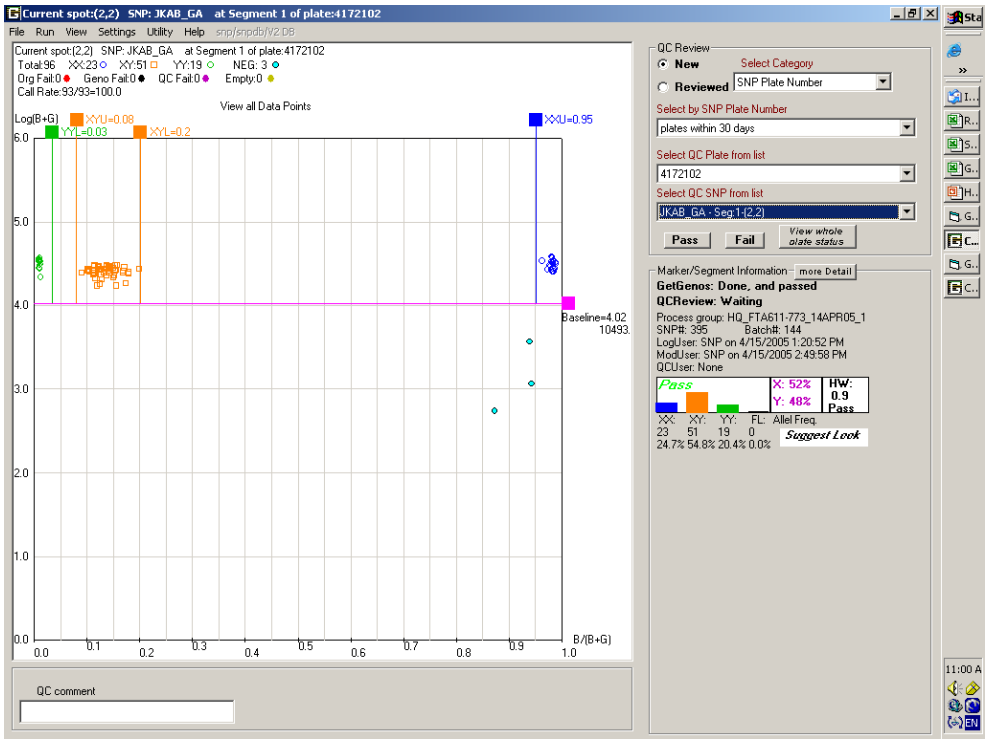
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.05

XY => 0.08 to 0.35

XX => 0.9 to 1



e. kK

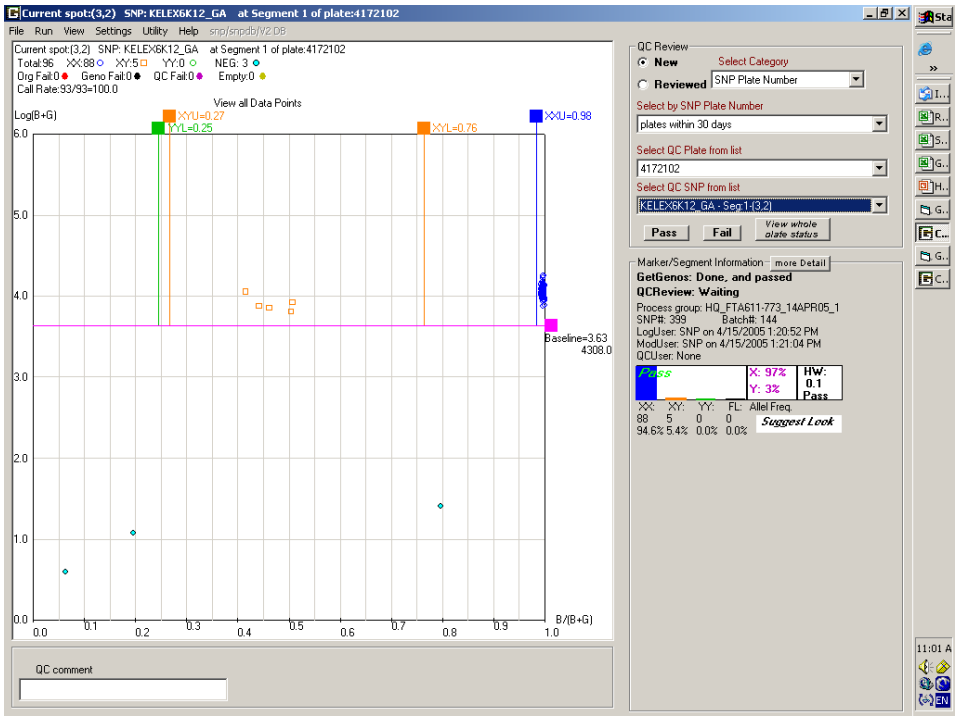
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.10

XY => 0.30 to 0.75

XX => 0.9 to 1



f. HPA5ab:

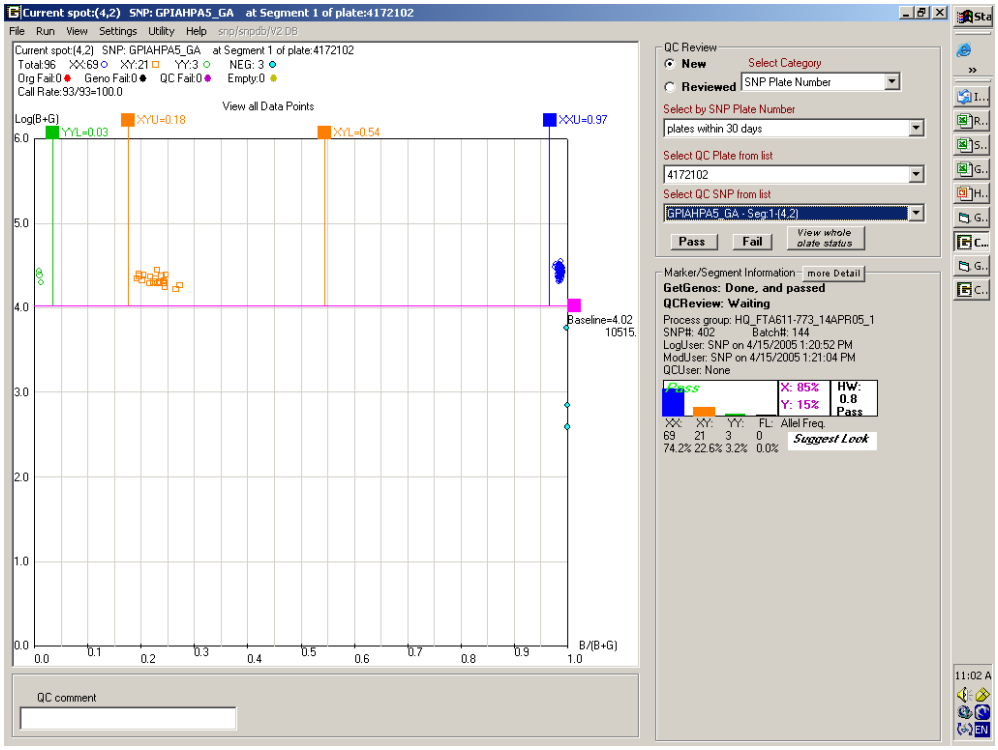
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.10

XY => 0.15 to 0.50

XX => 0.9 to 1



g. sS:

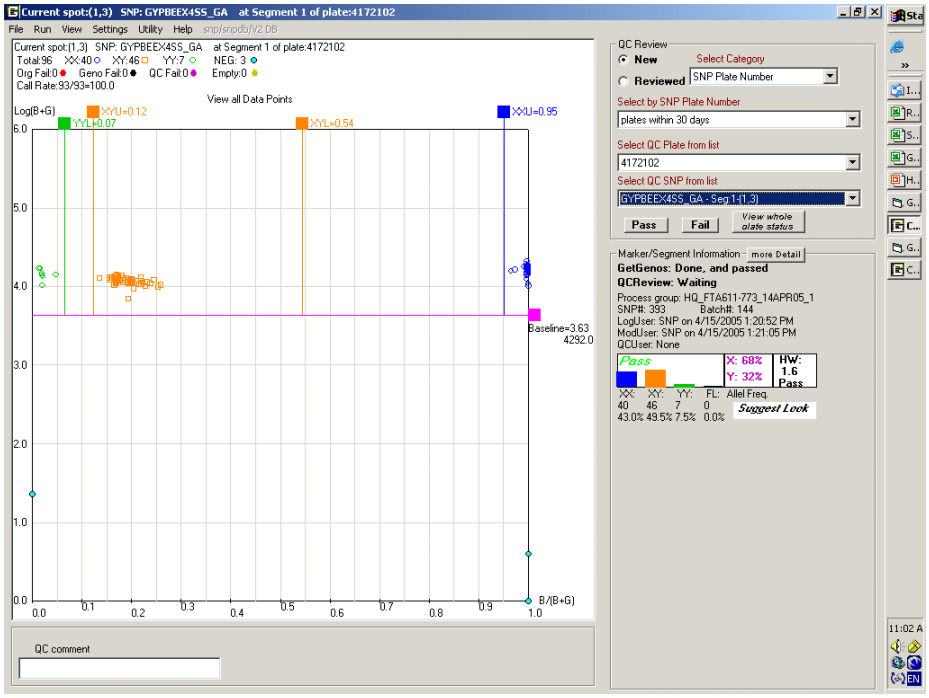
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.05

XY => 0.10 to 0.35

XX => 0.9 to 1



h. HPA1ab :

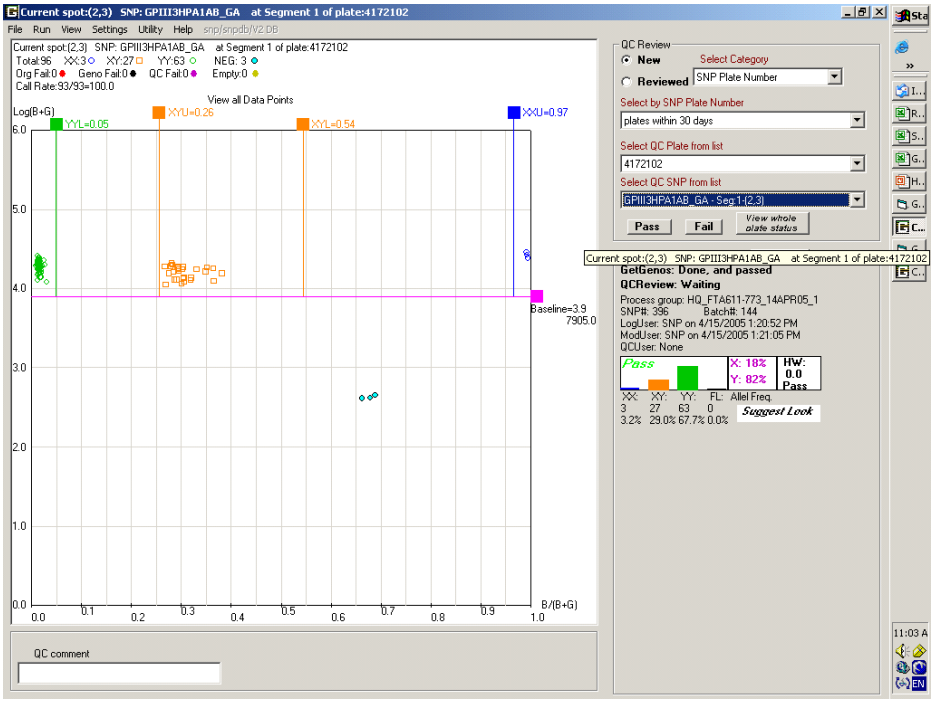
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.10

XY => 0.20 to 0.55

XX => 0.9 to 1



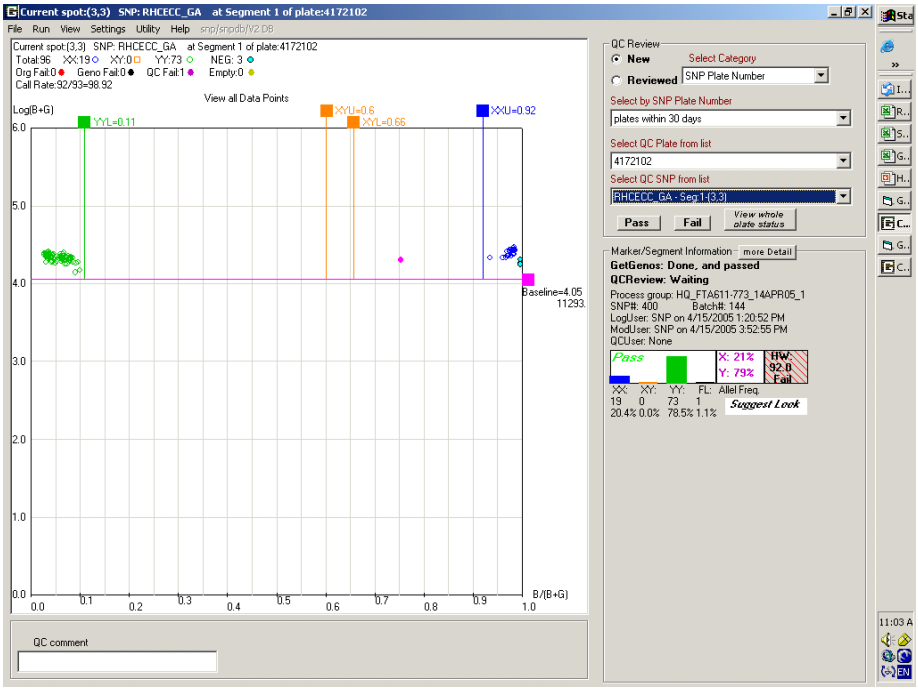
i. cC :

Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.20 (AA= cc or cC)

XX => 0.5 to 1 (GG = CC)



j. HPA2ab :

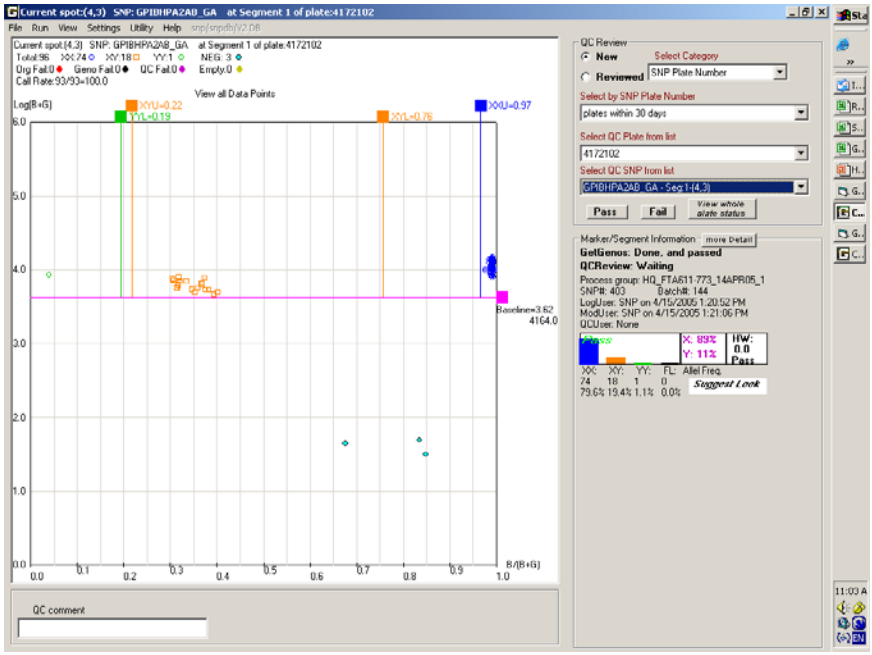
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.05

XY => 0.16 to 0.75

XX => 0.9 to 1



k. Fyab :

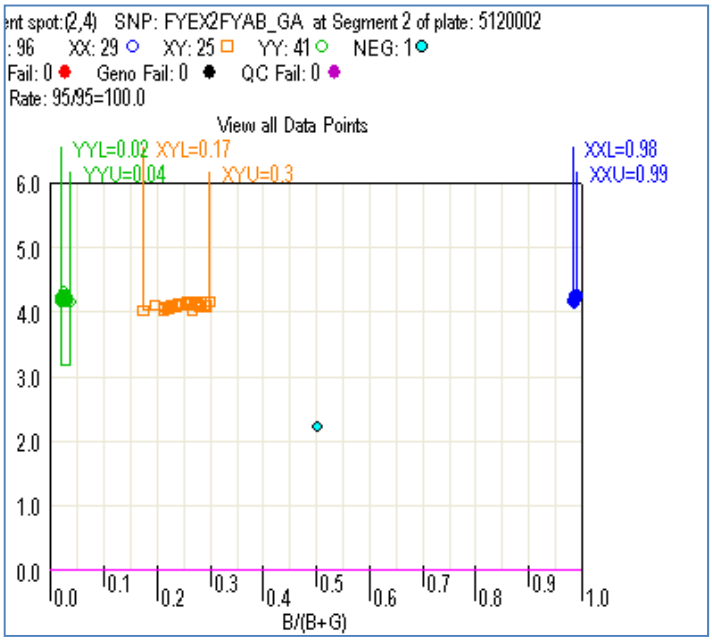
Normal Setting for Cluster Analysis:

Polar View

YY => 0.00 to 0.05

XY => 0.08 to 0.4

XX => 0.9 to 1



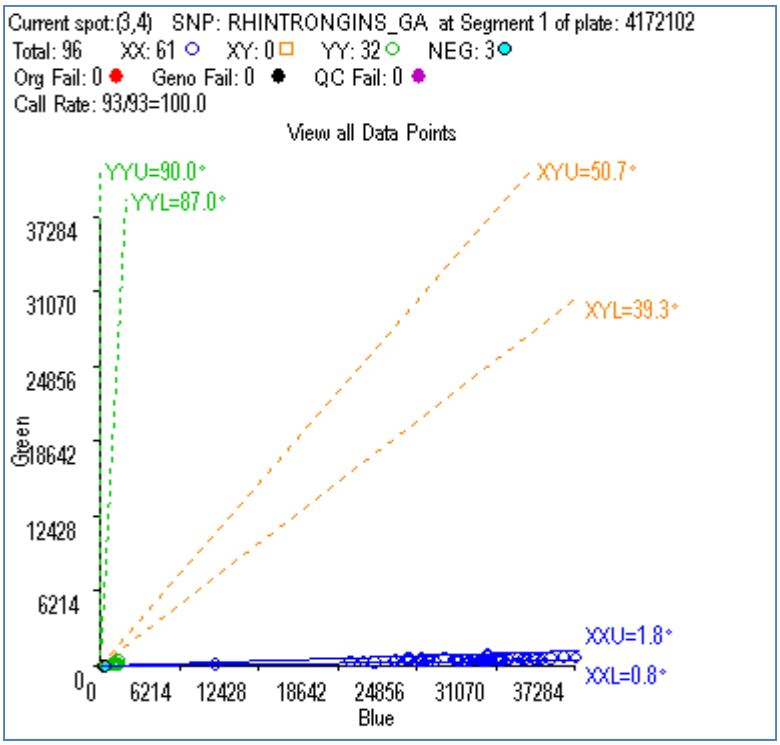
I. cC-intron :

Normal Setting for Cluster Analysis:

Cartesian View

YY < 8000(AA= cc)

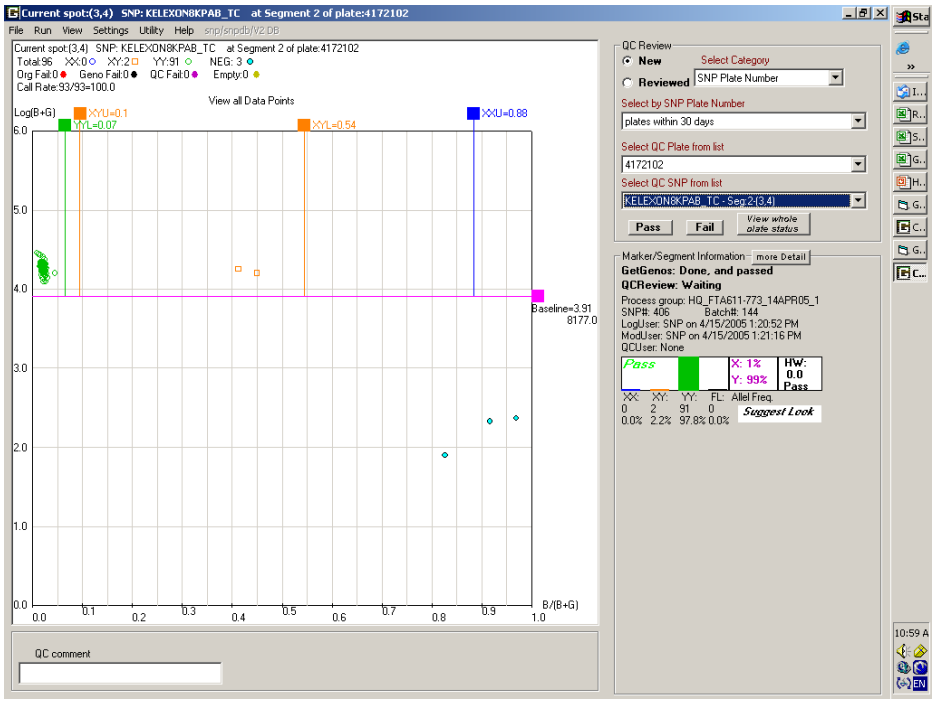
XX > 11000 (GG= (cC or CC)



Segment 2 (CT analysis) :

a. Kapa

Normal Setting for Cluster Analysis:
Polar view YY => 0.00 to 0.1 XY => 0.3 to 0.75 XX > 0.9 to 1



b. MN :

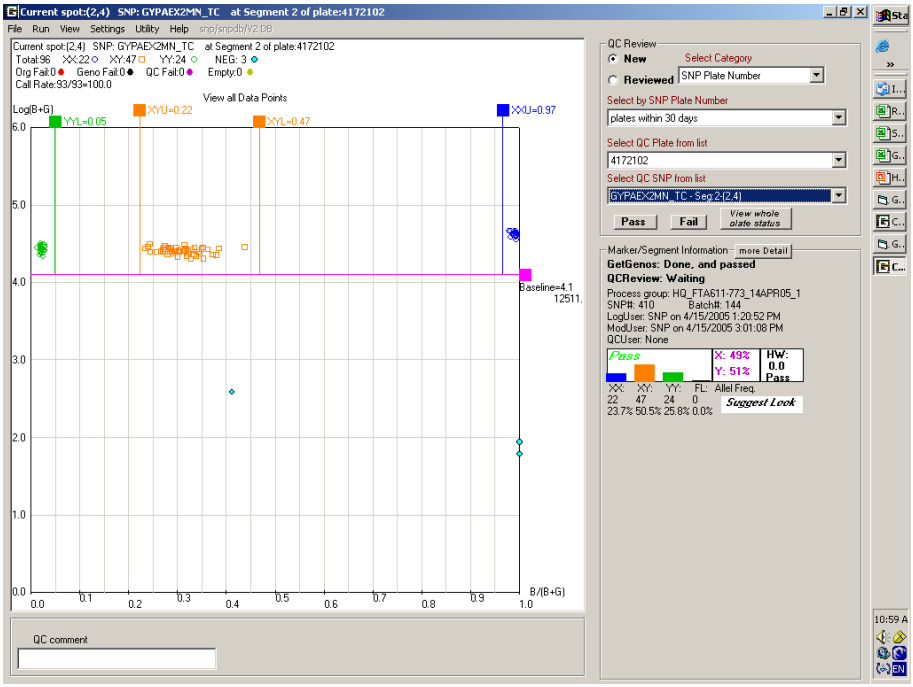
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.15

XY => 0.22 to 0.70

XX > 0.9 to 1



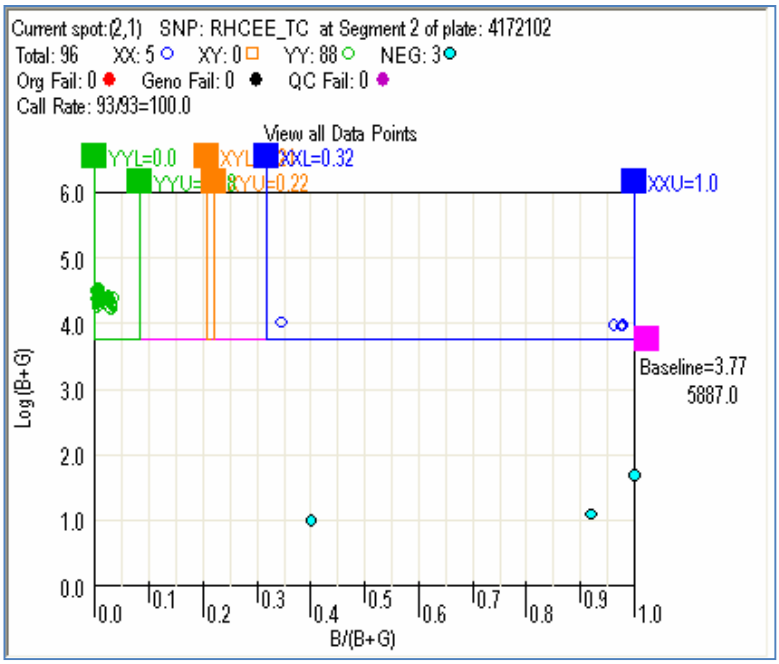
c. eE :

Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.1(CC = ee or eE)

XX > 0.3 to 1 (TT = EE)



d. Jkab :

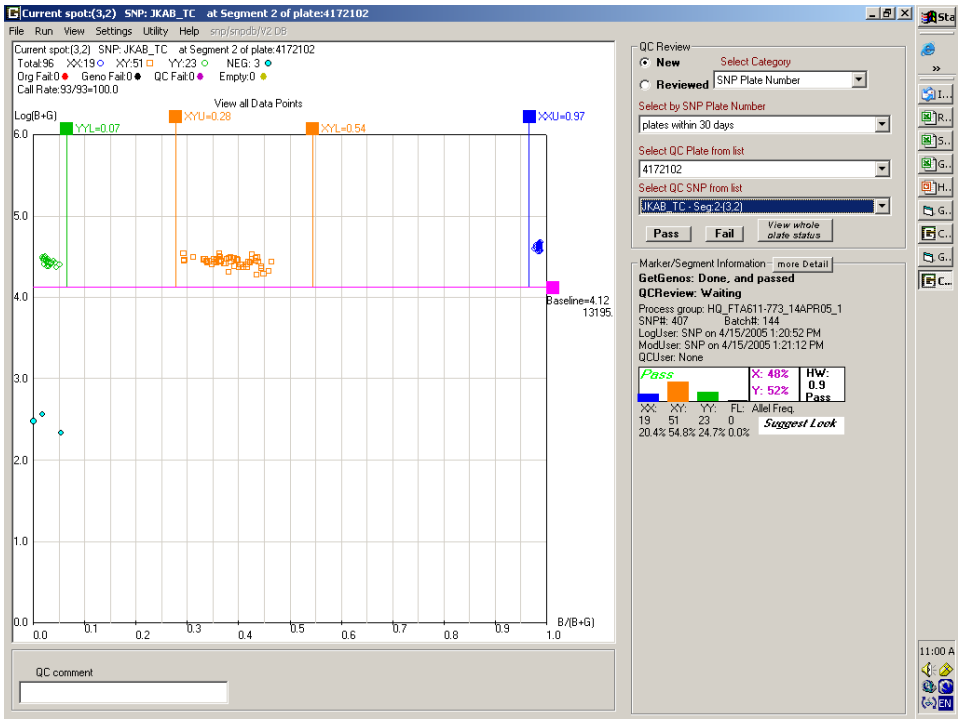
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.12

XY => 0.25 to 0.70

XX > 0.9 to 1



e. kK:

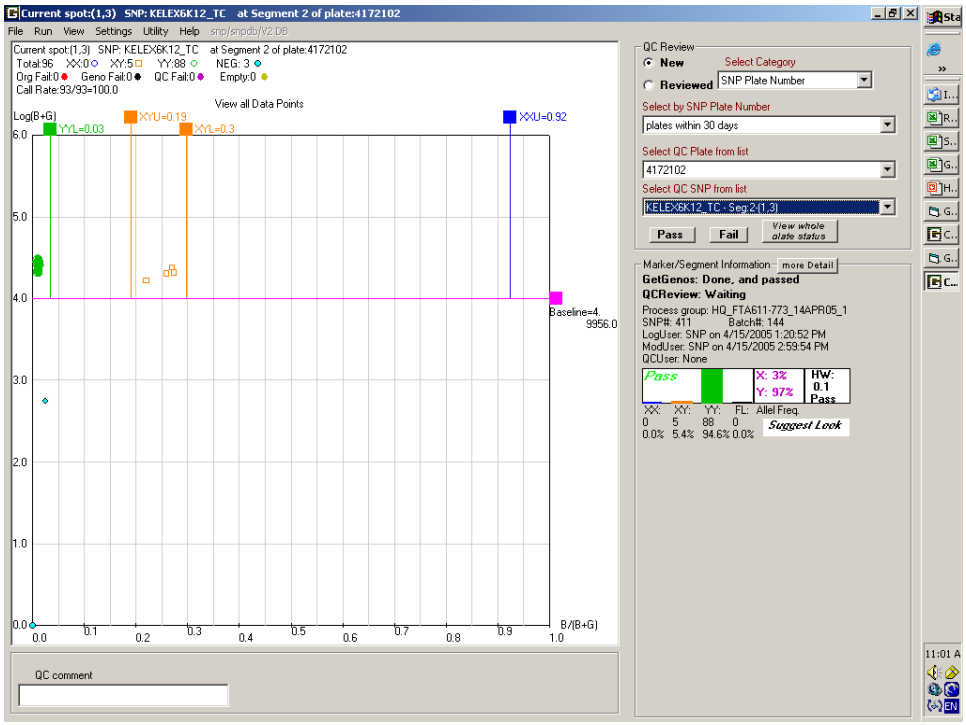
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.10

XY => 0.20 to 0.60

XX > 0.9 to 1



f. HPA5ab:

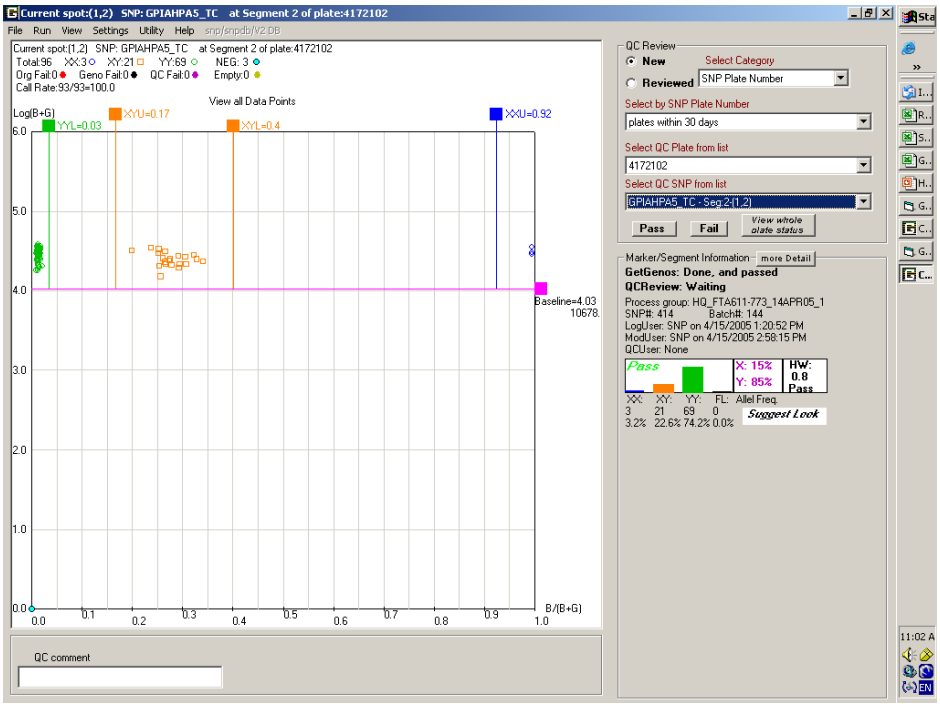
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.10

XY => 0.20 to 0.65

XX > 0.9 to 1



g. sS :

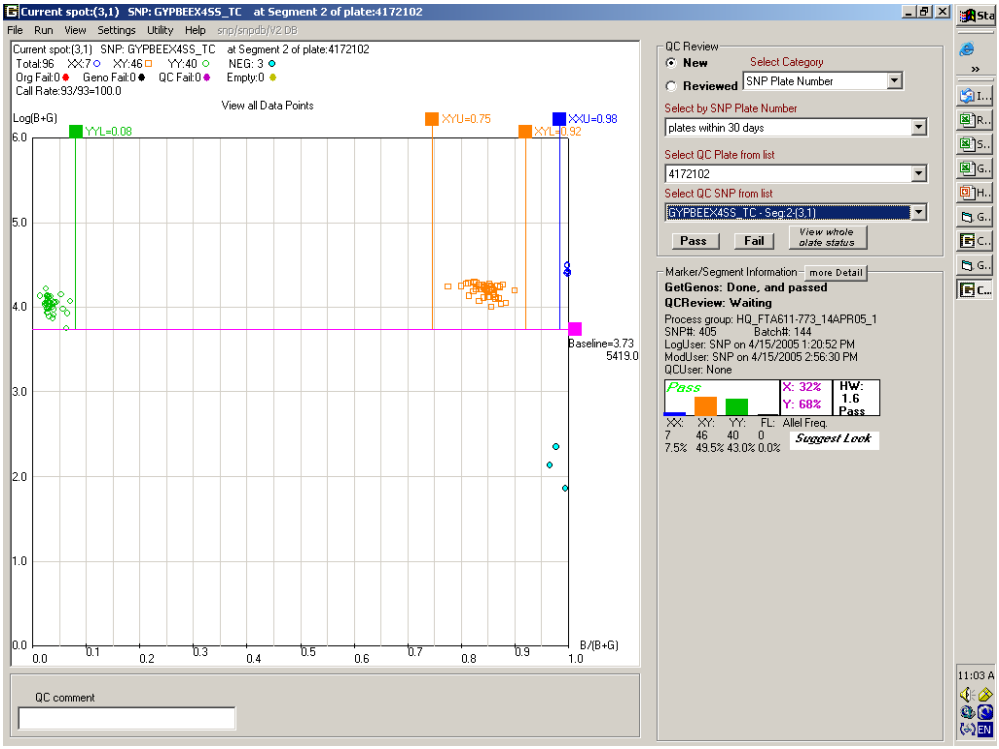
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.15

XY => 0.75 to 0.95

XX > 0.98 to 1



h. HPA1ab:

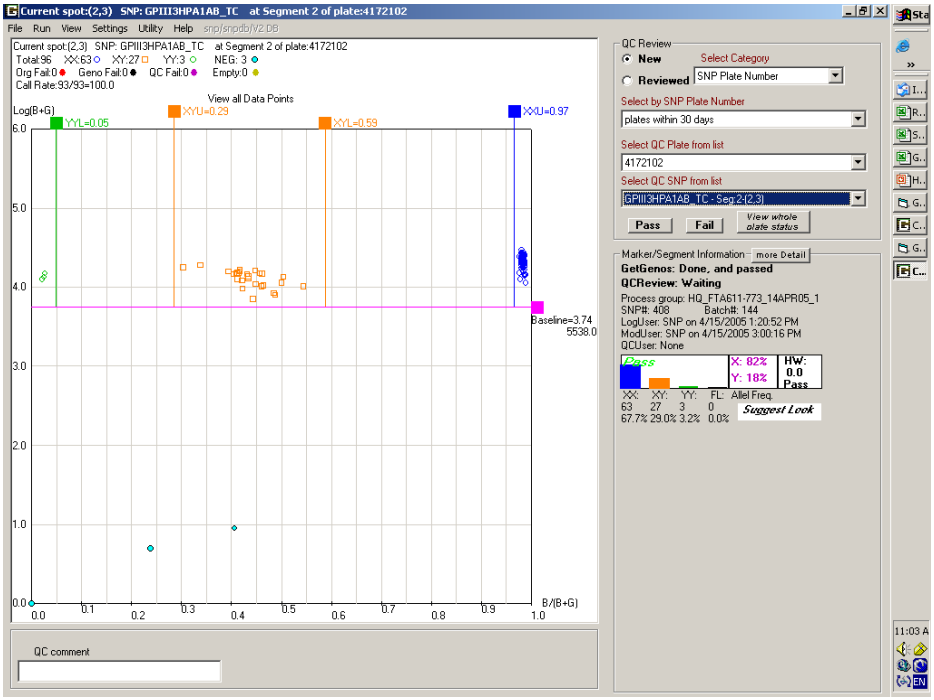
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.10

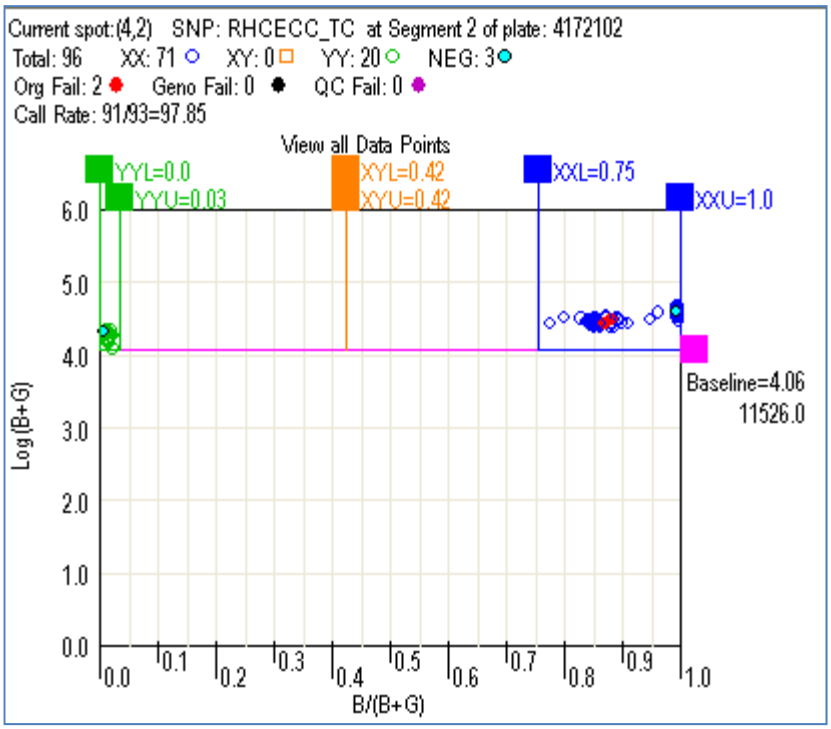
XY => 0.30 to 0.65

XX > 0.9 to 1



i. cC:

Normal Setting for Cluster Analysis: Polar view YY => 0.00 to 0.50 (CC = CC) XX > 0.8 to 1 (TT = cc or cC)



j. HPA2ab:

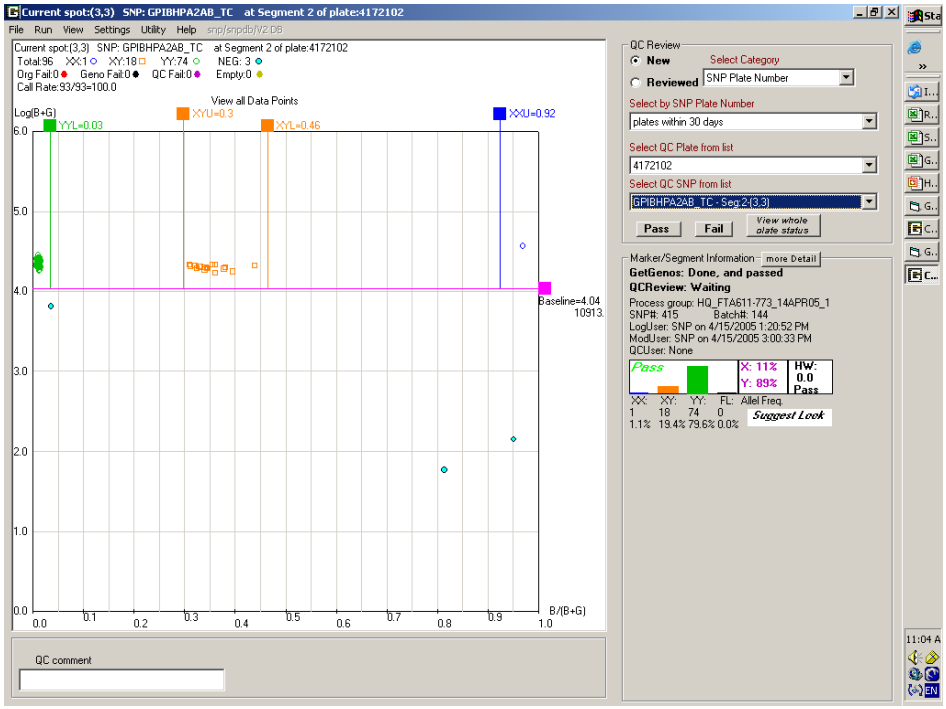
Normal Setting for Cluster Analysis:

Polar view

YY => 0.00 to 0.08

XY => 0.25 to 0.65

XX > 0.9 to 1



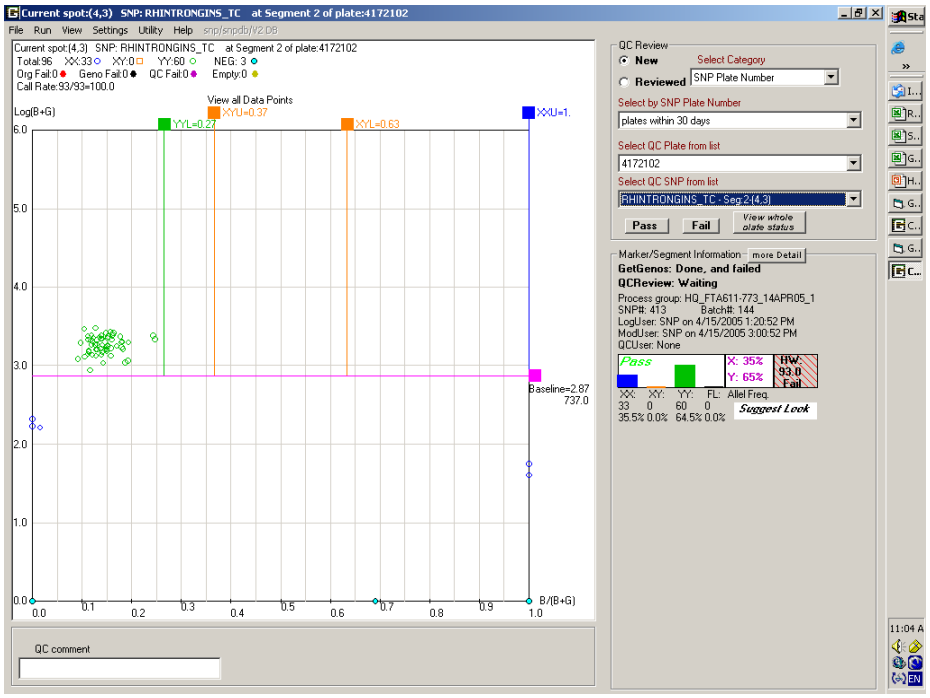
k. cC-Intron:

Normal Setting for Cluster Analysis:

Polar view

YY > 19000 (CC = cC or CC)

XX < 13000 (TT= cc)



I. Fyab:

Normal Setting for Cluster Analysis: Polar view YY => 0.00 to 0.08 XY => 0.25 to 0.65 XX > 0.9 to 1

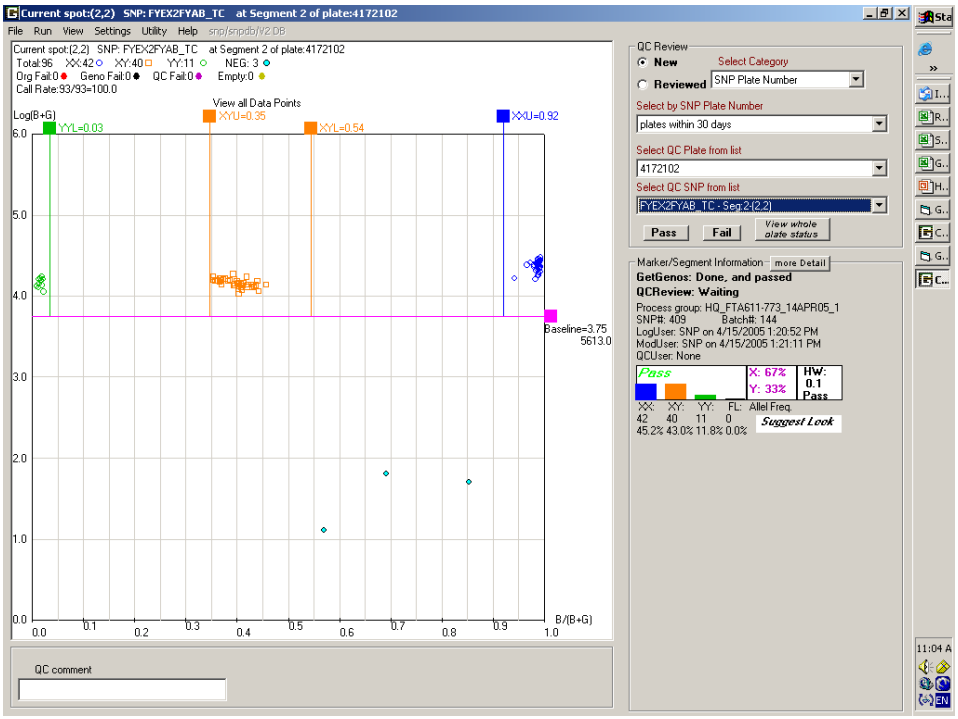


Figure 1: Position of Héma-Québec Assay variations on the 12-plex array plate:

GA:

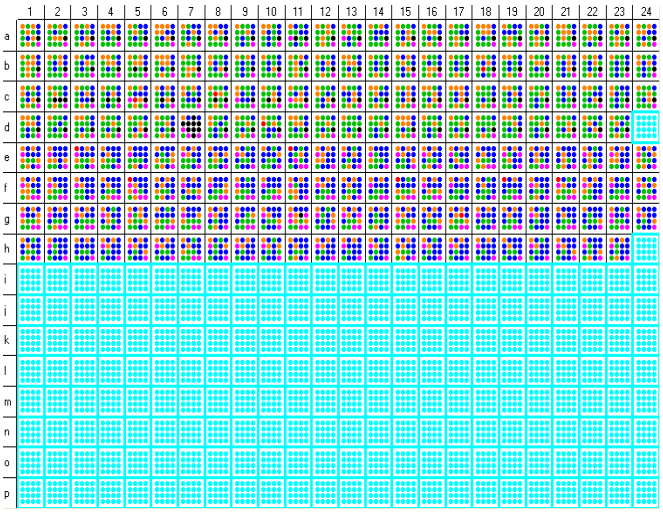
XY	Kapa	MN	XX
eE	Jkab	kK	HPA5
sS	HPA1	cC	HPA2
YY	Fyab	cCint	NEG

CT:

XY	eE	sS	XX
HPA5	Fyab	Jkab	cC
kK	HPA1	HPA2	cCint
YY	MN	KAPA	Neg

Figure 2: Example of a dust particle on the array spot:

Figure 3: Whole Plate Status View**:**



The figure represents an example of the whole plate status view. A clear problem for wells D07, D24 and H24 can be observed. Also, all samples located in the second half of the plate (I01 to P24) represent a clear assay problem. This problem has to be addressed and requires the Lab Manager approval before resuming analysis.

2. Data analysis and formatting using Sherpa software©

2.1. Importing the results to Sherpa:

N.B.: Two different users have to go through this section (User1 and User2).

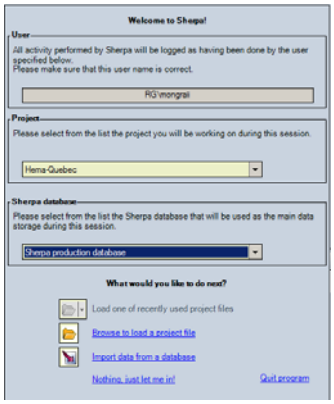
2.1.1. Login with your user name.

2.1.2. Open the Sherpa software.

N.B.: The user name listed in the Sherpa's welcoming page should match the user ID that has performed the cluster analysis.

2.1.3. Select *Héma-Québec* as the project.

2.1.4. Select "Sherpa production Database"



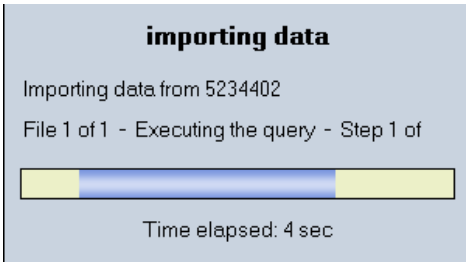
2.1.5. Click on "Import Data from a Database".

2.1.6. In the "Import Plates" window, select "SNPstream Machine Database".

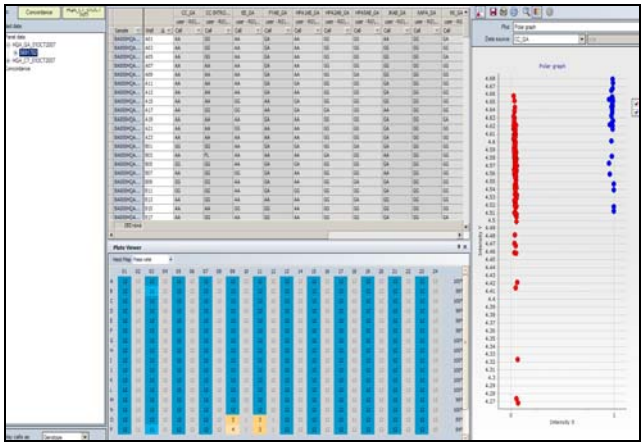
2.1.7. Click on "Load plate list" and enter the SNPware plate barcode in the "plate name" cell.

2.1.8. Check the box corresponding to the plate and click "Import".

2.1.9. While the plate is being imported, the following message should appear:

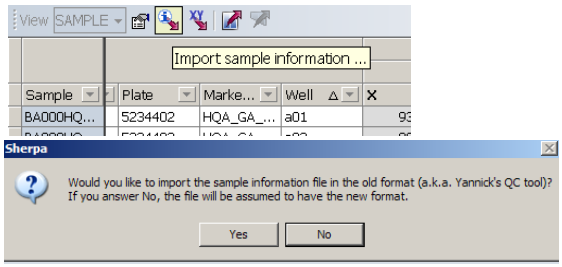


2.1.10. After the plate has been correctly imported, the following table should become available:



2.1.11. Click on the “Import Sample ID Information” icon and answer “Yes” to the pop-up question (the sample information comes from the Tracking Sheet file previously created in section 8.2. of SOP PGx-PR-012):

N.B.: Once User1 has imported the Sample ID Information, there is no need for User2 to repeat this step.



2.1.12. Click on the “Save to database” icon 

2.1.13. After saving the data, reinitialize the genotyping clusters on the GetGenos (SNPstream) software.

N.B.: Only User1 is required to perform this step.

2.2. **Concordance between users**

N.B.: This section applies to User 2 only, being a more experienced and knowledgeable technician than User 1.

2.2.1. Login to your user name.

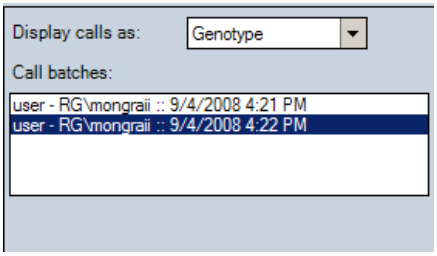
2.2.2. Open the Sherpa software.

N.B.: The user name listed in the Sherpa’s welcoming page should match the user ID that has performed the cluster analysis.

2.2.3. Select the ‘’Héma-Québec project’’.

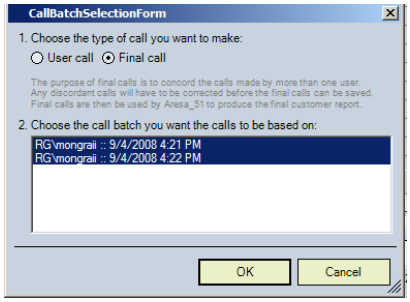
2.2.4. Click on the Data menu and select “Sherpa production database”.

2.2.3. In the left lower side of the Sherpa screen, two users’ call should now be available.



2.2.4. Click on “start call”.

2.2.5. Select the “Final call” option and highlight the two users’ call batch. Then click OK.



2.2.6. The user-to-user discordances are now displayed. Click on the filter icon  to sort down all user-to-user discordant calls.

2.2.7. Find out the discordant calls and evaluate the best clustering analysis.

2.2.8. Click on each discordant call and make the final call.

2.2.9. Give a reason and click “OK” (refer to table 1 for Sample Failure Criteria).

2.2.10. Repeat these steps for all discordant calls.

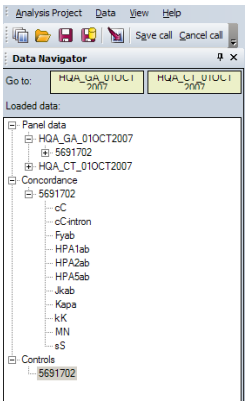
2.2.11. Once all discordant calls have been evaluated, click “Save Call” and “save database” to update the database.

Table 1: Sample failure criteria

Criteria	Description	Sample status
A	Sample failed over 2 different markers (evaluation made with Area51)	**Failed**
B	1 discordant genotyping results between both DNA strands	**Failed**
C	EXO-SAP issue pattern (refer to SNPstream documentation)	**Failed**
D	Complete well evaporation during the experiment	**Failed**
E	Clear plate pattern problem (missing arrays, non-uniform array spots on the SNPware plate).	**Failed**
F	Low intensity signal or sample outside normal clustering values (refer to cluster profiles)	**Failed**
G	Major dust particles issue onto the hybridization plate	**Failed**
H	**No signal for the SNPstream plate controls (XX, XY and YY)**	**Failed**

2.3. Control genotypes

2.3.1. On the left end side of Sherpa screen, click on “control”.



2.3.2. Look at the concordance table and check if controls match criteria described in table 2.

Table 2: Control failure criteria:

Controls	Description	Control status
DNA controls	Option[1]: Exact same criteria described in Table 1. Option[2]: No concordance between theoretical and actual genotypes.	Failed
Negative	Genotyping results are generated	Failed

2.3.3. Possible explanations for a majority of controls failing genotypes:

2.3.3.1. Possible evaporation or pipetting errors. Look at the *Plate Validation* tab in the corresponding Tracking Sheet for any comment that could justify the problem.

2.3.3.2. Possible grid shift that was not flagged previously.

2.3.3.3. Possible presence of dust particle in the array plate.

2.3.4. Possible explanations for a majority of controls discordant from expected genotypes:

2.3.4.1. Wrong sample ID and/or control ID from the SNPstream Input file.

2.3.4.2. Wrong sample ID and/or control ID from the Plate Maker file.

2.3.4.3. Possible DNA contamination. Refer to the gel picture in the *Plate Validation* tab in the corresponding Tracking Sheet.

2.3.5. If no explanation can be found, a new dilution of controls has to be done and the whole assay plate has to be repeated.

3. Data reporting using Area51©:

N.B.: The Area51 software analyzes the final calls (previously saved in the Sherpa database) as per acceptance criteria described in tables 1 to 4 (see above for tables 1 and 2; see below for tables 3 and 4).

Table 3: Marker failure criteria

Controls	Description	Control status
A	Call rate lower than 85%	Failed
B	Unacceptable clustering patterns	Failed

Table 4: Plate failure criteria

Controls	Description	Control status
Sample	More than 30% of wells show no signals for the XX, XY and YY SNPstream hybridization plate controls.	Failed
Sample	More than 30% of wells show poor EXO-SAP activity	Failed
Marker	More than 30% of markers show less than 85% of call rate	Failed
DNA controls	More than 10% of DNA controls genotypes are missing	Failed
DNA controls	More than 5% of DNA controls genotypes show discordant calls between theoretical and actual genotypes	Failed
Negative controls	More than half of markers show positive genotyping results (refer to Agarose gel for more details on possible contamination).	Failed

3.1. Type the following address in the internet browser: https://my.pgx.ca/session/new

3.1.1. Login.

3.1.2. Select Area51.

3.1.3. Select the desired plate to be reported.

3.1.4. Generate a technical report as well as a client report.

3.1.5. If Area51 generates a “Plate Failure” for any reason listed in table 4, please address the problem to the Lab Manager or to the appropriate authority.

Annex C

High throughput molecular profiling of blood donors for minor red blood cell and platelet antigens

Alexandre Montpetit[1], Michael S. Phillips[1], Ian Mongrain[1], Réal Lemieux[2] and Maryse St-Louis[2]

[1] Mc Gill University and Genome Quebec Innovation Centre, Montréal (Québec) Canada, [2] Héma-Québec, Recherche et développement, Sainte-Foy (Québec) Canada

corresponding author :

Maryse St-Louis, Ph.D.

Héma-Québec, Recherche et développement

1009, route du Vallon, Sainte-Foy (Québec) G1V 5C3 CANADA

Phone : (418) 780-4362 ext. 3254

Fax : (418) 780-2091

Running title : Molecular profiling of blood donors

The authors declare no competing financial interest

Reprints will not be available

Abstract

Background. ABO and Rh(D) phenotyping of both blood donors and transfused patients is routinely performed by blood banks to ensure compatibility. These analyses are done by antibody-based agglutination assays. However, blood is not tested for minor blood group antigens on a regular basis because of cost and time constraints. This can result in alloimmunization of the patient against one to several minor antigens and may complicate future transfusions. **Study design and Methods.** To address this problem, we have generated an assay on the GenomeLab SNPstream genotyping system to test simultaneously polymorphisms linked to 22 different blood antigens using donor's DNA isolated from minute amounts of white blood cells. **Results.** The results showed that both the error rate of the assay, as measured by the strand concordance rate, and the no-call rate were very low (0.1%). The concordance rate with the actual red blood cell and platelet serology data varied from 97 to 100%. Experimental or database errors as well as rare polymorphisms contributing to antigen conformation could explain the observed differences. However, these rates are well above requirements since phenotyping and cross-matching will always be performed prior to transfusion. **Conclusion.** Molecular profiling of blood donors for minor red blood cell and platelet antigens will give blood banks instant access to many different compatible donors through the set-up of a centralized data storage system.

Key words : SNPs, blood groups, high-throughput screening, minor antigen

Introduction

Blood banking facilities routinely test for the presence of the major antigens ABO and Rh(D). These analyses are done for every blood donor by antibody-based agglutination assays using automated equipment. However, minor antigens are not currently tested on a regular basis. This limitation may result in alloimmunization of the patient and complicate future transfusions.[1] Part of the problem is that the procurement of extensively phenotyped blood components to specific patients using agglutination-based assays is costly and labour intensive especially when the patient's serum contains antibodies against several minor blood group antigens. Furthermore, it is often impossible to test some antigens because of reagent shortages. The more frequent use of supportive transfusion therapies and a better survival of patients indicate that the need to transfuse alloimmunized individuals will increase in the future. For all these reasons, the procurement of compatible blood represents a significant operational problem for blood banks.

Cloning of blood group genes and knowledge of their molecular basis made possible the development of molecular typing methods, leading to identification of new mutations, polymorphisms and alleles. The first application of these assays has been for antenatal diagnosis of maternal-fetal compatibility to predict the risk of development of hemolytic disease of the newborn.[2] Since then, genetic tests have been described for most of the clinically important red blood cell and platelet antigens. [3,4] Despite these successes, these assays are usually not suitable for massive testing. With the human genome project came great improvements in high-throughput technologies (increased number of parallel reactions, automated systems

to facilitate the data production and better tools for the analysis), resulting in a dramatic reduction in genotyping cost.

These commercially available genotyping systems which can analyze multiple SNPs in parallel include Sequenom's MassARRAY system,[5] Beckman's GenomeLab SNPstream Genotyping System,[6] ABI's SNPlex Genotyping System, ParAllele Molecular Inversion Probe Technology,[7] Affymetrix' GenFlex Tag Arrays[8] and Illumina Bead-Array technology.[8] In this work, we have tested the possibility of using a semi-automated system, the GenomeLab SNPstream Genotyping System from Beckman Coulter, to genotype DNA samples from blood donors. This system uses multiplex PCR and single-base extension to genotype up to 12 SNPs per sample simultaneously and can perform up to 600 000 genotypes per day. Most of the other high-throughput platforms use a fixed set of SNPs or need a few hundreds or thousands of SNPs to be cost-effective.

Many recent publications have showed use of high-throughput systems to genotype multiple antigens at once, including a study using also the Genome Lab SNPstream genotyping system.[10-13] Here, we present the development of an assay that can be used to genotype polytransfused patients and regular blood donors for 22 different antigens. It improves on previous reports as it needs only minute amounts of blood and its low no-call rate and error rate makes it suitable for use in clinic in order to build a registry to accelerate the quest for compatible blood.

Materials and Methods

Blood collection and DNA extraction

Peripheral blood samples were collected from randomized volunteers from the province of Quebec after signature of an informed consent form approved by Héma-Québec's Research Ethics Committee. 618 samples consisted of a blood drop applied on an FTA card. All samples were anonymized and the FTA cards were sent to the Genome Quebec Innovation Centre in Montreal for genotyping. DNA was then extracted from the FTA cards using the GENERATION Capture Card Kit (Gentra Systems, Minneapolis, MN), according to the manufacturer's specifications. A good proportion of the blood samples were phenotyped at Héma-Québec for the most usual minor blood group antigens using standard serology techniques. The 96 samples that were used to optimize the panels consisted of DNA isolated at Héma-Québec by the R&D group using the Qiagen QiaAmp DNA Blood mini kit (Qiagen, Mississauga, ON) according to the manufacturer's protocol. These samples are routinely used as controls.

Genotyping

Primer selection

SNP flanking sequences were tested for the presence of repeats or duplicated regions and masked if needed using the BLAT program (http://www.genome.ucsc.edu). In the first round, PCR and extension primers were designed using the Autoprimer program (http://www.autoprimer.com). The program selects PCR primers that will generate products ranging between 90 and 180 bp and an optimized single base-pair extension primer 5' to the

SNP site and assemble them into panels of 12 SNPs of the same type.[6] In subsequent rounds, modifications were made to the primers and the panel composition (see Tables 2 and 3).

PCR

Twelve-plex PCR reactions were performed in 384-well plates (MJS BioLynx) in a 5 µL volume using 6 ng of DNA, 75 µM dNTPs, 0.5 U of AmpliTaq Gold (Perkin-Elmer), and the 24 PCR primers at a concentration of 50 nM each in 1 X PCR buffer. Thermal cycling was performed in GeneAmp PCR system 9700 thermal cyclers (Applied Biosystems) using the following program : initial denaturation at 95°C for 5 min followed by 40 cycles of 95°C for 30 sec, 55°C for 55 sec, 72°C for 30 sec. After the last cycle, the reaction was held at 72°C for 7 min. Following PCR, plates were centrifuged briefly and 3 µL of a mixture containing 0.67 U Exonuclease I (Amersham Pharmacia) and 0.33 U Shrimp Alkaline Phosphatase (Amersham Pharmacia) were added to each well. The plates were sealed and incubated for 30 min at 37°C and at 95°C for 10 min.

Extension and hybridization

Extension reactions and hybridizations to the Orchid UHT microarray plates were carried out as described by Bell *et al.*,[6] using the SNPware UHT reagent kit and the appropriate extension mix kit containing two dideoxynuclotides labelled with either Bodipy-Fluorescein or TAMRA dye (Beckman Coulter, Fullerton, CA). Finally, the plates were read using the GenomeLab SNPstream Array Imager and fluorescence intensity was measured with help

of the UHTImage software. Intensity was plotted and genotypes were called by the UHTGetGenos software. After visual inspection of the clusters, manual adjustments were made for some of the assays.

Alternate genotyping assays

All samples showing discrepancies between phenotype and genotype results were retested using different genotyping assays [14-16] or by sequencing.

Results

On the GenomeLab SNPstream system, panels of 12 SNPs can be genotyped in parallel for the same alleles. SNPs were prioritized based on their importance for the blood bank, their frequency in the Quebec population (we prioritized antigens with a frequency above 5%) and their allele type. Twelve antigen pairs were thus selected: C/c, E/e, M/N, S/s, $Kp^{a/b}$, $Fy^{a/b}$, $Jk^{a/b}$, K/k, and some platelet antigens, HPA-1a/1b, HPA-2a/2b, HPA-4a/4b and HPA-5a/5b. For these antigens, the main causative polymorphism have been conclusively identified [3,4]. For Rh(C/c), polymorphisms in strong linkage disequilibrium with the causative SNP were selected because of difficulties in generating good assays (see below). Panels for both strands were designed to increase confidence in the genotyping calls. Ninety-six DNA samples with known phenotypes (for about half of the antigens studied) were used to optimize the panels.

One obstacle in the creation of these panels was that the E/e antigens, which were considered essential to the panel, are caused by a C/G SNP. Since the 11 other antigen pairs are caused by C/T or A/G SNPs, we decided to develop two assays which test only one allele at a time (i.e. the C allele in the C/T panel or the G allele in the G/A panel; see Figures 1B and 2B). Another major issue that came up after a few rounds of testing was the fact that no primers could be designed to efficiently amplify the *RHCE* gene over the *RHD* gene near any C/c discriminating SNP. Two assays were then developed: one discriminates for the presence of the c antigen over C/C individuals (assay *RHCE* exon 2) and the other assay looks for the presence of an insertion in intron 2 of *RHCE* which was previously shown to be strongly associated with the *RH(C)* allele but absent from either the *RH(c)* allele or the *RHD* gene (assay

RHC intron 2; see Figure 1A).[17,18] 11 antigen pairs are included in the final panels (HPA-4a/4b was removed; see Table 4).

To evaluate the panels and the feasibility of the proposed approach, DNA from 618 regular Hema Quebec's blood donors were used. The phenotypes were not known in advance by the genotyping group in Montreal. In addition, 32 duplicate samples were present, but their identity was not transmitted to the genotyping group until after results were obtained. Examples of the genotyping results are shown in Figure 2. The call rate, as measured by combining data from both strand assays, was very high at 99.9%, indicating that the assays are extremely robust considering the variability in the DNA quantity that could be extracted from the FTA cards (Table 4). The assays for antigens E/e had the lowest call rate at 99.2% which is more than acceptable since in that case genotype calls on both strands were needed to make a final call. The accuracy of the assays, as measured by the strand correlation or the duplicate samples reproducibility, was also very high with both measurements being above 99.7%. Only one discordant call was observed when comparing data from each strand (in the Rh(C/c) assays) and two discordant calls when comparing calls from each of the duplicate samples (in the Rh(C/c) and the Rh(E/e) assays).

The concordance rate between the genotypes and the serology data was also very good for 18 out of 19 antigens, and varied from 97% to 100%. One exception was for antigen HPA-1b where the concordance was only 50% (3 correct calls out of 6 tests). Sequencing confirmed the initial genotyping result. All other discrepancies (12 samples) were verified by another approach. The initial genotype was confirmed for

ten of the twelve samples (which translates to a final error rate of 0.1%). The two discrepancies were in the RHCE gene, one for the E and one for the C antigen, and by looking at the raw data they could easily be explained. No serology data was available for antigens HPA-2a, HPA-2b and Kp^b, but all observed frequencies are in agreement with the literature.[19]

Discussion

In this study, we have developed an assay that can simultaneously test the presence of 22 different minor antigens (C/c, E/e, K/k, $Fy^{a/b}$, $Jk^{a/b}$, S/s, M/N, $Kp^{a/b}$, HPA-1a/1b, HPA-2a/2b and HPA-5a/5b) in a DNA sample. Most of these antigens are common in the general population (in the population studied here, all are above 15%, except for K and Kp^{a} at 6% and 4% respectively) and if they are not taken in consideration in transfusion medicine, major complications may arise. At the moment, blood banks must test dozens of different blood samples for specific antigens to find a compatible donor. These tests are very expensive and a compatible donor can be hard to find for some rare antigen combination. The molecular profiling of all donors and the creation of a centralized data system would give blood banks instant access to many different compatible donors decreasing the cost. Using the GENTRA system for DNA isolation from FTA cards and the Beckman Coulter GenomeLab SNPstream platform, the entire procedure is very fast as it takes only about three working days and human intervention is minimal. The use of FTA cards has many great advantages for implementation at a blood bank facility. Blood is very easy to collect as only one drop is needed. The collection can conveniently be done at the same time as the measurement of hemoglobin at the registration step. Thus existing functioning protocols need only to be modified slightly and no additional blood tubes, which would add unnecessary time and cost (reagents, shipping, etc.) need to be collected. Also DNA is very stable on FTA cards (it can last for years) which is very convenient for handling, shipping and long-term storage at room temperature.[20]

The missing call rate on this genotyping system has been previously evaluated at 5% and the error rate at 0.5%.[6] In a clinical setting, any genotype information will be confirmed with a agglutination-based assay before the blood is transfused to the patient. Thus any phenotype concordance higher than about 90% would be acceptable. But since many factors can play a role (call rate, genotyping errors, phenotyping errors, additional polymorphisms, database entry errors, etc.), we considered that obtaining the highest genotype quality was essential. Thus, we developed panels on both strands in order to reduce both the no-call rate and error rate as two "no calls" are unlikely (provided the DNA sample is of good quality) and that two discordant calls will not be registered as valid. The data obtained from this study showed the strong reliability of the technique having an accuracy of > 99.7% and a call rate of more than 99%. All genotype-phenotype correlations, except one, were above 97%, which indicates that the panel will be suitable for clinical use. The only exception was for antigen HPA-1b, which showed very poor correlation (50%, but only 6 comparisons could be made with the phenotype database). However, all discrepancies between genotype and serology data except two (both in *RHCE* gene) were confirmed by another genotyping method or by sequencing indicating that the quality of the results generated by our assays is very high.

At the time of writing this paper, a comparable study using the GenomeLab SNPstream platform was published.[11] Although it validates the use of this platform, we had a very different approach using results from both strands to increase confidence in the genotype calls. Their assays had overall lower concordance rates especially for the C and e antigens (87.7% and 73.9% respectively) which would preclude their use in clinic. They explain that the problem originated from the fact

that the RHD gene was co-amplified with RHCE during the process. As described above, our panels avoided this problem for the C/c antigens by using primers specific for the RH(c) allele and by designing an additional assay that test the presence of an insertion in intron 2 of RHCE which is associated with the C antigen.[17,18] As for the E/e antigens we designed primers that successfully discriminate between both genes using adjacent polymorphisms.

There could be many factors that could contribute to differences between the observed genotype and serological data. First, more than one polymorphism can contribute to the protein conformation or expression. For example, it is well known that a polymorphism in the promoter of the *FY* can block its expression.[21,22] This polymorphism is rare in white individuals, but extremely frequent in individuals of African descent. We confirmed that two of the three samples that differed with the FY serological information had that mutation. In addition, several different varieties of each antigen can exist (weak and partial variants) caused by additional polymorphisms (for example the Fy^x antigen is present at a frequency of 1.5 to 2.5 % in the Caucasian population).[23,24] These usually show a very different reactivity in a clinical test. Thirdly, because of the high homology between *RHCE* and *RHD* in the sequence in and around exon 2, we had to use a second assay which doesn't test directly any C antigen causing polymorphism, but which relies on the presence of an insertion in cis with the C allele. This is mostly true for Caucasians, but Faas *et al.*[25] and Daniels *et al.*[26] have showed that for up to 25% of individuals of African descent the presence of a D-CE-D hybrid gene containing exons 1-3 of *RHD* and expressing a weak C phenotype will induce false negative results. Finally, in the case of HPA 1b,

since no additional modifying polymorphism is known [3], the most plausible explanation would be an error in the phenotype database.

We are aware that many false positives or false negatives for most of the antigens studied will be observed. Testing for all the rare polymorphisms is not feasible at the moment as it would make the cost prohibitive, and also because new polymorphisms are identified on a regular basis. Also, testing for ABO and RH(D) was not a priority for our panel since these tests are done automatically for every collected blood unit and is extremely cheap to perform. In this first panel, we decided to concentrate on frequent polymorphisms in our population, but eventually, they could be expanded with additional antigens as a drop in genotyping costs and improvements in the multiplexing capacities of the systems will occur (indeed our SNPstream system has been recently upgraded to accommodate 48-plex arrays). On the other hand, specialized panels for *ABO*, *RHD* and *RHCE* genes, which exhibit the most variations, could be created to facilitate clinical analysis and support serology data.

Since the call rate and error rate obtained in this study were well above requirements, and because phenotyping and cross matching will always be performed prior to transfusion, such a procedure will definitely save both time and money. The molecular profiling of blood donors for all these antigens is much less expensive than the serological typing, not to mention that some tests are very labor-intensive and complicated to perform (platelets) or that the reagents for some antigens are in very short supply or extremely expensive. At the present time, blood banks must test on average 5-10 times more blood bags that what is required to find a compatible donor.

The process can be very time-consuming for rare combinations of antigens. The creation of a centralized data storage system would give blood banks instant access to many different compatible donors. Finally, since much more information will be known about each donor, many future complications such as alloimmunization of patients receiving multiple transfusions will be reduced.

Acknowledgments

We thank Louis Thibault, Claudine Côté, Annie Beauséjour and Gerry Boucher for the donor recruitment. We also thank Alya'a Sammak for her technical expertise in genotyping on the SNPstream platform and Josée Perreault and Martine Richard for the DNA extraction and the alternate genotyping assays to validate the results.

References

1. Seltsam A, Wagner FF, Salama A, et al. Antibodies to high-frequency antigens may decrease the quality of transfusion support: an observational study. Transfusion 2003;43: 1563-6.

2. Bennett PR, LeVanKim C, Colin Y, et al. Prenatal determination of fetal RhD type by DNA amplification. N Engl J Med 1993;329:607-610.

3. Storry JR, Olsson ML. Genetic basis of blood group diversity. Br J Haematol 2004;126:759-771.

4. Hurd CM, Cavanagh G, Schuh A, et al. Genotyping for platelet-specific antigens: techniques for the detection of single nucleotide polymorphisms. Vox Sang. 2002 Jul;83(1):1-12.

5. Jurinke C, Van-den-Boom D, Cantor CR, Koster H. Automated genotyping using the DNA MassArray technology. Methods Mol Biol 2001;170:103-116.

6. Bell PA, Chaturvedi S, Gelfand CA, et al. SNPstream UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. BioTechniques 2002;32:S70-S77.

7. Hardenbol P, Banér J, Jain M, et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotech 2003;21:673-678.

8. Matsuzaki H, Dong S, Loi H, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat Methods 2004;1:109-111.

9. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 2003;68:69-78.

10. Bugert P, McBride S, Smith G, et al. Microarray-based genotyping for blood groups: comparison of gene array and 5'-nuclease assay techniques with human platelet antigen as a model. Transfusion 2005;45:654-659.

11. Denomme GA, Van-Oene M. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. Transfusion 2005;45:660-666.

12. Beiboer SHW, Wieringa-Jelsma T, Maaskant-Van-Wijk PA, et al. Rapid genotyping of blood group antigens by multiplex polymerase chain reaction and DNA microarray hybridization. Transfusion 2005;45:667-679.

13. Hashmi G, Shariff T, Seul M, et al. A flexible array format for large-scale, rapid blood group DNA typing. Transfusion 2005;45:680-688.

14. St-Louis M, Perreault J, Lemieux R. Extended blood grouping of blood donors with automatable PCR-ELISA genotyping. Transfusion 2003;43:1126-1132.

15. Lyou J-Y, Chen Y-J, Hu H-Y, et al. PCR with sequence-specific primer-based simultaneous genotyping of human platelet antigen-1 to -13w. Transfusion 2002;42:1089-1095.

16. Meyer O, Hildebrandt M, Schulz B, et al. Simultaneous genotyping of human platelet antigens (HPA) 1 through 6 using new sequence-specific primers for *HPA*-5. Transfusion 1999;39:1256-1258.

17. Poulter, M, Kemp, TJ, Carritt, B. DNA-based rhesus typing: Simultaneous determination of RHC and RHD status using the polymerase chain reaction. Vox Sang 1996;70: 164-168

18. Tax MG, van der Schoot CE, van Doorn R, et al. PA.RHC and RHc genotyping in different ethnic groups. Transfusion 2002;42:634-644.

19. Denomme GA, Rios M, Reid ME. Molecular protocols in transfusion medicine. San Diego: Academic Press 2000.

20 Belgrader P, Del Rio SA, Turner KA, Marino MA, Weaver KR, Williams PE. Automated DNA purification and amplification from blood-stained cards using a robotic workstation. Biotechniques 1995;19:426-432.

21. Tournamille C, Colin Y, Cartron JP, Le-Van-Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 1995;10:224-228.

22. Iwamoto S, Sugimoto N, Okuda H, Kajii E. Characterization of the Duffy gene promoter: evidence for tissue-specific abolishment of expression in Fy(a-b-) of black individuals. Biochem Biophys Res Commun 1996;222:852-859.

23. Olsson ML, Smythe JS, Hansson C, et al. The Fy(x) phenotype is associated with a missense mutation in the Fy(b) allele predicting Arg89Cys in the Duffy glycoprotein. Br. J. Haematol 1998;103:1184-1191.

24. Tournamille C, Le Van Kim C, Gane P, et al. Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy(x) individuals. Blood 1998;92:2147-2156.

25. Faas BHW, Beckers EAM, Wildoer P, et al. Molecular background of VS and weak C expression in Blacks. Transfusion 1997;37:38-44.

26. Daniels GL, Faas BHW, Green CA, et al. The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. Transfusion 1998;38:951-958.

Table 1. PCR primers used in the panels

Assay	Primer name	Sequence (5'-3')	Size
RHCE exon 5	RHCE_e-F-s2	TGGATGTTCTGGCCAAGTG	107
	RHCE_e-R-s2	CTGTCACCACACTGACTGCTAG	
RHC intron 2	RHCintron-F	CAGGGCCACCACCATTTGAA	151
	RHCintron-R	TGGTAGCAGGCGTCTGTAAAAA	
RHCE exon 2	RHCEc_CF-s3	TCCCCCTCCTCCTTCTCA	143
	289-RHCEas-HD9135	GCCAGCATGAAGAGGTTGAA	
KEL exon 6	KelEx6K1_2F-s2	GGATCCTTATGCTCAGCCCC	157
	KelEx6K1_2R-s2	AGGATGAGGTCCTAGGTAGGCTCTG	
FY exon 1	FYex2Fya_bF	AGTCATCCAGCAGGTTACAGG	129
	FYex2Fya_bR	AAGATGTATGGAATTCTTCCTATGG	
JK exon 9	JKa_b_F-s3_1	CCTGCTAACTTTCAATCCCACCCTC	150
	JKa_b_R-s3_1	TGCCAGGTGAGCGCCATGAAC	
GYPB exon 4	GYPBEx4s_SF-s2	TTGTCAAATATTAACATACCTGGTACAGT	170
	GYPBEx4s_SR-s2	AATATGATTAAGAAAAGGAAACCCG	
GYPA exon 2	285-GYPAs-HD9131_1	CAACTTCTATTTTATACAGA	98
	287-GYPAas-HD9133_1	TCTTTGTGACTGAAGAAGAA	
KEL exon 8	KelExon8Kpa_bF	AGCAAGGTGCAAGAACACT	101
	KelExon8Kpa_bR	AGAGCTTGCCCTGTGCCC	
GPIIIA exon 3	GPIIIAex3HPA-1a_bF	ATTCTGGGGCACAGTTATCC	114
	GPIIIAex3HPA-1a_bR	ATAGCTCTGATTGCTGGACTTC	
GPIB exon 2	GPIBHPA-2a_bF-s2	CTCAGTCAAGTTGTTGTTAGCCAGAC	105
	GPIBHPA-2a_bR-s2	CTCTACCTGAAAGGCAATGAGCT	
GPIA	GPIaHPA5_F	GCTCTTGGTAGGTGCACCAATGT	148
	GPIaHPA5_R	TTCCAAATGCAAGTTAAATTACCAG	

Table 2. Probes used in the GA panel

Antigen	Primer name	Sequence (5'-3')
C	RHCintron-Gins-U12	CGACTGTAGGTGCGTAACTCGGTGCCCTTTGTCACTTCCCA
c	RHCEEx2Rhc_CGA-U9	GACCTGGGTGTCGATACCTACCCTTGGCTTGGGCTTCCTCACCTC
e	RHCE_eCTU3-s3_1	CGTGCCGCTCGTGATAGAATCCTTTGGATTGGACTTCTCAGCAGAG
K/k	KelEx6K1_2-GA-U5	GCGGTAGGTTCCCGACATATACTCATCAGAAGTCTCAGC
Fy^a/Fy^b	FYex2Fya_b-U11	AGAGCGAGTGACGCATACTAGATTCCTTCCCAGATGGAGACTATG
Jk^a/Jk^b	JKa_b-U4_1	AGCGATCTGCGAGACCGTATCTCAGTCTTTCAGCCCCATTTGAG
S/s	GYPBs_S-GA-U7	AGGGTCTCTACGCTGACGATAAACGATGGACAAGTTGTCCC
M/N	GYPAex2M_NGA-U2	GGATGGCGTTCCGTCCTATTGTGCATTGCCACCYCAGTGGTACTT
Kp^a/Kp^b	KelExon8KGA_U1	ACGCACGTCCACGGTGATTTGCCTCAGAAACTGGAACAGCC
HPA-1a/1b	GPIIIAex3HPA-1GAU8	GTGATTCTGTACGTGTCGCCCTTCAGGTCACAGCGAGGTGAGCCC
HPA-2a/2b	GPIBex2HPA2GA-U10-s3	AGATAGAGTCGATGCCAGCTGAGCTTCTCCAGCTTGGGTGTGGGC
HPA-5a/5b	GPIaHPA5-GA-U6	GGCTATGATTCGCAATGCTTGTCTACCTGTTTACTATCAAA

Table 3. Probes used in the CT panel

Antigen	Primer name	Sequence (5'-3')
C	RHCintron-CT-U10	AGATAGAGTCGATGCCAGCTGGCAGAGGCTGCAATGAGCTATGATTGTAC
c	RHCEEx2Rhc_CctU6-s2	GGCTATGATTCGCAATGCTTTGCTCCAGCTGTGTCTCCGGAAA
E	RHCE_eCTU3-s3_1	CGTGCCGCTCGTGATAGAATCCTTTGGATTGGACTTCTCAGCAGAG
K/k	KelEx6K1_2U7_CT	AGGGTCTCTACGCTGACGATTGGACTTCCTTAAACTTTAACCGAA
Fy^a/Fy^b	FYex2Fya_bU4	AGCGATCTGCGAGACCGTATGGGGGCAGCTGCTTCCAGGTTGGCA
Jk^a/Jk^b	JKa_b-CT-U5-s2	GCGGTAGGTTCCCGACATATACCCCAGAGTCCAAAGTAGATGT
S/s	GYPBEx4s_S-CT-U2-s2	GGATGGCGTTCCGTCCTATTTTGAAATTTTGCTTTATAGGAGAAA
M/N	GYPAex2M_NCTU11-s3_1	AGAGCGAGTGACGCATACTAACAGAAATTGTGAGCATATCAGCAT
Kp^a/Kp^b	KelExon8Kpa_bU12_CT	CGACTGTAGGTGCGTAACTCTTCCTTGTCAATCTCCATCACTTCA
HPA-1a/1b	GPIIIAex3HPA1CTU8-s2	GTGATTCTGTACGTGTCGCCCTCTTTGGGCTCCTGTCTTACAGGCCCTGCCTC
HPA-2a/2b	GPIBex2HPA-2a_bCTU9	GACCTGGGTGTCGATACCTAGACCCTGCCCCCAGGGCTCCTGA
HPA 5a/5b	GPIaHPA5-CT-U3	CGTGCCGCTCGTGATAGAATTTAGTTTATTTTTTTTTTTTACCT

Table 4 Genotyping results and correlation with phenotype

Antigen	Call rate	Concordance DNA strands	Frequency	Reproducibility	Concordance Pheno/Geno	N
C	99.8 %	99.8 %	62.2 %	97 %	99.2 %	131
c			81.5 %	100 %	100 %	323
E	99.2 %	100 %[1]	22.9 %	97 %	99.5 %	196
e			97.1 %	100 %	99.5 %	181
K	100 %	100 %	6.1 %	100 %	100 %	22
k			99.8 %	100 %	100 %	344
Fy^a	100 %	100 %	62.1 %	100 %	99.2 %	130
Fy^b			85.7 %	100 %	97.3 %	75
Jk^a	100 %	100 %	74.4 %	100 %	98.6 %	146
Jk^b			75.2 %	100 %	98.9 %	97
S	100 %	100 %	58.2 %	100 %	100 %	70
s			88.6 %	100 %	98.5 %	66
M	100 %	100 %	80.4 %	100 %	98.7 %	76
N			71.0 %	100 %	100 %	26
Kp^a	100 %	100 %	4.1 %	100 %	100 %	26
Kp^b			99.8 %	100 %	-	-
HPA-1a	100 %	100 %	97.1 %	100 %	100 %	30
HPA-1b			30.2 %	100 %	50.0 %	6
HPA-2a	100 %	100 %	99.5 %	100 %	-	-
HPA-2b			16.0 %	100 %	-	-
HPA-5a	100 %	100 %	98.6 %	100 %	100 %	9
HPA-5b			21.8 %	100 %	100 %	9
Total	**99.9 %**	**99.9 %**	**-**	**99.7 %**	**99.3 %**	**1963**

[1] No heterozygote calls are registered for this assay. N: number of samples for which serology data was available.

Legends to Figures

Figure 1. Detailed strategy for the detection of the two non-standard SNPs.

A) *RHCE* insertion in intron 2. This insertion is present almost uniquely in C positive individuals. The probes in the G/A panel and the C/T panel are indicated in bold. The base interrogated in each panel is different **B)** *RHCE* exon 5 C/G SNP. The same probe is used in the C/T and the G/A panel but in each only one terminator nucleotide can be incorporated. However, in the C/T panel, small amounts of ddT seem to be incorporated in the extension product.

Figure 2. Selected examples of genotyping results. Polar and cartesian representations are presented.

A) HPA-2a/2b in the G/A panel. **B)** E/e in the C/T panel. The SNP tested is a C/G. Since no G is present in the extension mix, the T dideoxynucleotide is incorporated relatively well in the extension product and one can clearly see the distinction between C/C and C/"G" genotypes in the polar view.