A Comparison, Evaluation and Accuracy Assessment of the Fluidigm (TM) Platform in Genotyping Common and Rare variants in an Alaska Native Population

Joseph Manuel Vincent Yracheta

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2014

Committee:

Kenneth Thummel

Wylie Burke

Edward Kelly

Program Authorized to Offer Degree:

Pharmaceutics

University of Washington

Abstract

A Comparison, Evaluation and Accuracy Assessment of the Fluidigm (TM) Platform in Genotyping Common and Rare variants in an Alaska Native Population

Joseph Manuel Vincent Yracheta

Chair of the Supervisory Committee:
Dr. Kenneth Thummel
Dept. of Pharmaceutics

Indigenous people have historically experienced financial, cultural, political, social and bio-medical disadvantages. Disparities abound. The rise of genomic information and the development of rapid, low-cost testing technologies has led to the introduction of tailored and individualized medicine into clinical care; lack of access to such technology or participation in the research to develop it could be a new source of health disparity. Conversely, the historical conflicts between the federal government and tribal communities, their lack of inclusion in research development and interpretation make such an investment of group trust in research a difficult proposition. These issues demonstrate a need for a change of research approach in improving AI/AN health care.

This study with its attention to community based participatory research, a two-step procedure in SNP identification, and the use of the discrete and cost effective Fluidigm platform, represents such a change. Knowledge that can be generated with proper application of the Fluidigm platform can be empowering to the Yup'ik people and other indigenous people of the US by filling-in the gaps between gene variants in metabolizing enzymes and drug response. In addition to possible differences in AI/AN drug response, researchers who work with AI/AN communities must equally recognize the importance and historical foundation of tribal/group/and collective protections of research data. In this

investigation, the Fluidigm platform succeeds in balancing respect for genomic privacy while delivering accurate and efficient genomic testing. It conserves sample volume, it yields reliable results, it is quickly modifiable both in specific sample and assay selection and it has a high and rapid throughput. All of these characteristics make it amenable to genomic discretion in that AI/AN groups do not have to supply researchers with large or repeated collections of DNA, pilot and preliminary findings can be reported back to tribal entities for appraisal and approval, and once done can be modified for continuing studies.

DEDICATION

To my parents: Mary & Manuel Yracheta, my Grandparents: Jesus & Eladio Yracheta, Florentino & Carmen Cruz, all my ancestors, my sisters: Deborah, Leetriana, & Christine, my children Francesca, WagaCan, Makaheya, Tatiye, & Wacinyeya and most of all my wife Karen Ellen Ducheneaux

"There is just one thing I would like all of you to acknowledge: my people and I bring wood from the mountains everyday to light up our sacred fires, we make arrows and hunt for deer in order to keep the Sun, our grandfather, alive. We eat their meat in ceremony because this is how we communicate and give thanks to the spirits of heaven and of the four directions and to our mother Kuerajperi the creator. If we could not do this the fate of the world would be in danger." ~Ire Tikatame ~Purhepecha ~Relación de Michoacán

ACKNOWLEDGEMENTS

I am grateful to the advice and mentorship of Ken Thummel, Wylie Burke and Ed Kelly. You all have been tremendously patient and wholeheartedly encouraging through my time here at the University of Washington. My journey from the Rosebud Sioux Reservation to Seattle and onward has been made so much more powerful via your example and network.

I would also like to thank Yvonne Lin and Tauri Senn for first taking me under their wing when I was new in the lab. I thank the many members of the Thummel Lab such as Zhican Wang, Tim Wong and Tina Calamia. I thank the department of Pharmaceutics and the great support staff there. In addition, I would like to thank my co-workers and grant members such as, Barbara Kavanaugh, Program Manager, Allan Rettie, PhD from NWA-PGRN, Malia Fullerton, PhD, Rosalina James, PhD, Helene Stark, PhD, Kate West, Lorelei Walker, Nini Shridhar, Michelle Montgomery and Dana Gold from CGHE.

For their help with genetic analysis, I would like to thank the Fred Farin Lab with Patricia Janssen and Jesse Tsai as my most helpful lab assistants. I would also like to thank the study participants from Alaska, Bert Boyer, Scarlett Hopkins, and Jynene Black and the rest of the crew at the University of Alaska Fairbanks for all of their hard work with sample collection and community engagement.

I would like to acknowledge the friendship, encouragement, faith and employment of my former PI's and Graduate Student fellows: Mark Brownfield, PhD, Thackery Gray, PhD, Elena Bingaman Cleary, PhD, Peter Rittenhouse, PhD and Louis Van De Kar, PhD (R.I.P.). And, though I never met him, Dr. Atul Sharma of McGill University, who early on in my days at the UW gave me much advice and encouragement based on his own experience with Canadian First Nations people.

Lastly, I'd like to thank all of my extended family of mine and my wife's, especially mi Leksi Mato Sake. This strong combination of Mexican and Lakota family encouragement, love and raison pour la découverte (reason to discover). Without my desire to assist and educate about their health needs and solutions, I wouldn't have the motivation or tenacity to have finished this work. Thanks greatly to my wife Karen Ducheneaux and all of my children for whom I have the deepest love and devotion.

"But I think I have done right to save the vision in this way, even though I may die sooner because I did it; for I know the meaning of the vision is wise and beautiful and good; and you can see that I am only a pitiful old man after all." "A good nation I will make live." ~ Black Elk

INTRODUCTION

This project was designed as one of many first steps of addressing American Indian /Alaska Native (AI/AN) health care disparities via genomic technology. Through the auspices of the National Institutes of Health (NIH), the Pharmacogenomics Research Network (PGRN) was created in 2000 to contribute to better human health through the investigation of variation in genes that determine drug responses.[1] Of the many goals of the PGRN, fundamental in these investigations are the goals of creating reference sequence and variant data, with population frequencies, statistically valid phenotype-genotype correlations, mechanistic understanding of statistical associations, and dissection and systematic analysis of complex drug responses.[1]

Patient reports of adverse side effects or lack of efficacy and clinical anecdote amongst Indian Health Service (IHS) physicians and other health care providers amongst reservation and urban AI/AN communities indicate that AI/AN people may sometimes have different responses to medications that are seen among the majority white population.[234567] The Northwest Alaska Pharmacogenomics Research Network project (NWA-PGRN) was formed to pursue partnership-based research to better understand the genetic contributors to medication response among AI/AN people.[8] The NWA-PGRN represents a unique partnership between multiple academic entities (U of Washington, U of Alaska Fairbanks, U of Montana), healthcare organizations in Washington (Group Health), Alaska (Southcentral Foundation and Yukon-Kuskokwim Health Corporation) and Montana (Confederated Salish and Kootenai Tribal Health Department) and American Indian and Alaska Native people receiving care from those health organizations.

The limited research of the AI/AN genome that has been conducted to date has uncovered novel SNPs (private alleles) and unexpected patterns of genetic variation in AI/AN

populations in North America that justify further study.[9][10] With regard to pharmacogenetic variation, it was noticed that, comparatively, data for common SNP's and their applicability toward drug response or other health outcomes in AI/AN populations is very sparse. A systematic literature review conducted by Cheedy et. al. (2008) revealed very little information on cytochrome P450 (CYP) polymorphisms and even less clinically meaningful functional assessments of genetic variation in AI/AN populations of the Americas. Those papers that were relevant were found mainly in studies of Central and South American Amerinds. [11] Though the peoples of the Americas do share many common genetic components, they, like most human populations, have experienced the effects of genetic drift, regional population bottlenecks, culturgenic effects (genetic drift influenced by culture including marriage rules, dietary adaptations and religious isolation)[12][13][14], epigenetics, epistasis and the timing and make up of European admixture would caution against the generalization of any metabolizing enzyme SNP's found in the Americas. It is because of this dearth of information regarding allele frequency, epistasis, and dietary/environmental interaction that this project set out to evaluate the performance of a modifiable and intermediate scale platform called Fluidigm™ for the purpose of generating genotype data for several Alaska Native populations, including the Yup'ik people living in the Yukon-Kuskokwim Delta, who were participants in this research project.

The Fluidigm company and its technology was founded in 1998 by Dr. Stephen Quake and his laboratory at the California Institute of Technology. He and other researchers in the semi-conductor industry from the 1970's realized that the potential for integrated circuitry could also be applied to fluidic circuits in the life science industry in applications where samples and volumes were at a premium because of their small size, rarity and/or cost of acquisition. Their technology, termed Integrated Fluidic Circuits (IFCs), consists of multilayered rubber channels set into an acrylic plate that is similar to the common 96 well

reaction plates. These plates and channels can be precisely choreographed for carrying out reactions on a fine scale. Designed for high throughput, the Fluidigm IFCs have the advantage of carrying out hundreds of thousands of reactions on a microscopic scale, decreased variability due to small volumes and precise geometry of layering fabrication, direct viewing of reactions through the clear elastomer, sample/reagent efficiency and cost (using up to hundreds-of-times less sample compared to microtiter plates), and most importantly for this study, versatility of assay type/number and sample number.[15]

In this study we chose to use the Biomark HD System™ with the 96 x 96 well Dynamic Array™ IFC[16] for characterization of genetic variation in AI/AN populations. This platform allows up to 96 participant samples to be placed at low volume on one side of the pressurized "chip" and up to 96 SNP genotyping assays on the other (Figure 1). The chip is then pressurized to mobilize and make ready the samples and reagents for 9,216 micro-PCR reactions with about a day's preparation and a 4-hour automated reaction and analysis time.

Three other types of genotyping technology considered for this project. The first competing technology to be considered was genotyping via direct sequencing through traditional Sanger sequencing or "next gen" Illumina sequencing of the entire genome or selected genes. Besides the lag time caused by the large amount of data generated and intensive bioinformatic analysis required, this method would be costly and slow given the high number of samples and possible SNP's across many loci we would need for our study design. Also, broad interrogation of the genome brings with it the potential for "genomic invasiveness". Briefly, "genomic invasiveness" connotes potential bioethical problems arising from privacy, social stigma and intellectual property concerns and is of a unique and specific nature in regard to AI/AN populations.

The second competing technology evaluated was a Taqman ©based Allelic Discrimination Assay with fluorochrome labeled probes and primer sequences in the standard 96 well plate. Here the cost is lower and the assay time is faster but it is still relatively slow and reagent/ sample intensive compared to Fluidigm, thereby raising the cost for a study of this size.[17] The third technology that competes more favorably with Fluidigm is a DNA-SNP microarray or "chip" technology, such as DMET™, marketed by Affymetrix©. This method has the advantage of high throughput, tailored design for SNP detection and quick assay time. DMET is capable of genotyping 1,936 high-value drug metabolism and transporter allelic variants in 225 genes.[18] However, the cost is relatively high given the size of this study.

An important platform selection issue that especially pertains to Indigenous populations worldwide and AI/AN populations of this study is the flexibility of the technology to allow targeted genotype assay additions/subtractions for specific research needs under tribal approval and consent processes. In terms of assay inventory, the DMET product insert claims that "These markers have been evaluated across a minimum of 1,200 individuals from multiple populations, including Caucasian, African and Asian"[18] with no mention of AI/AN populations. It is because of the aforementioned lack of research both in the unique and private alleles and their functional associations in these populations that made flexibility a premium attribute of the genotyping technology that we would want to employ. DMET was found wanting in this respect.

More comprehensive analysis of the genome can overcome the limitation of DMET, but it raises another important issue for AI/AN populations – that of discretion. The amount of information produced by direct whole genome or exome sequencing will capture population-specific variation, but it could be personally invasive for tribal communities, due to their small size and highly inter-related nature.

Partnership with Yup'ik people living in the Yukon-Kuskokwim Delta has been nurtured for many years by CANHR researchers, community representatives and the community as a whole.[1920] This study is an extension of that research history and has been conducted following principles of Community Based Participatory Research (CBPR), including the convening of a Community Planning Group which has provided advice and guidance to PGRN researchers. These actions, and the history of the YKHC-CANHR partnership, exemplify ethical and culturally aware research practices, in stark contrast to cases such as the Havasupai vs. Arizona State University, where researchers demonstrated a lack of cultural and historic sensitivity to the tribe's moral right and legal claim to control their own samples and research trajectory.[22]

It is because research should be mindful of the socially and historically unique aspects of the community they intend to study that the researchers in this study have endeavored to ensure consultation and respect for tribal sovereignty. Such mindfulness should and can be reflected in our choice of technologies and analyses.

In summary, after consideration of technical performance, cost, and genomic discretion, it was concluded for the purposes of the NWA-PGRN investigation that none of the alternative technologies available can efficiently compete with the Fluidigm genotyping platform. The remainder of this thesis investigation focused on the specific performance of the Fluidigm platform for simultaneous, targeted, multi-variant genotyping of DNA samples of an Alaska Native population – specifically, Yup'ik people living in coastal and inland villages of the Yukon-Kuskokwim delta. Specific genetic analysis performance characteristics interrogated were: (1) frequency of "no call" results for each SNP and in aggregate; (2) reproducibility of SNP calls after introduction of a gene targeted pre-amplification step prior to Fluidigm analysis; (3) concordance of Fluidigm calls with direct gene sequencing results in a subset of samples; (4) concordance of Fluidigm results with DMET chip results from an analysis of

DNA from archive, anonymous human livers; (5) concordance of allele frequencies between Fluidigm w/ Pre-Amp and Deep Sequencing.

MATERIALS AND METHODS

Study Participants:

All aspects of this research project were approved by the Yukon-Kuskokwim Health Corporation Executive Board of Directors and the University of Alaska Fairbanks IRB (UAF #174024-8, Pharmacogenetics in Rural and Underserved Populations). Study participants for this investigation were engaged by Dr. Bert Boyer and his team at the Center for Alaska Native Health Research (CANHR), University of Alaska Fairbanks (UAF). Participant criteria include men or women 14 years or older, non-pregnant that self-identify as Alaska Native. However, during the time period of this study, only participants 18 years and above were analyzed. DNA samples were obtained by one of two processes: from archived material of those re-contacted and re-consented participants from previous studies done with CANHR or from fresh material collected from new participants after verbal and written consent. A total of 366 DNA samples from a self-identified Alaska Native (AN) study population living in the Yukon-Kuskokwim delta were analyzed using the Fluidigm platform.

As part of the platform performance evaluation, archived DNA from a human liver bank maintained by the University of Washington School of Pharmacy was also subjected to Fluidigm-based genotyping. The tissue bank was created originally through the donation of organs for unrestricted research purposes and converted to an anonymous tissue bank after destruction of all donor identifiable information. Thus, use of DNA samples derived the liver tissues has been declared “non-human research” by the UW IRB, in accordance with federal (Office for Human Research Protections) regulations.

DNA Isolation:

A blood sample from each study participant was collected via venous blood draw and subjected to differential centrifugation. DNA was isolated from the buffy coat (White blood cells), using a DNeasy 96 Blood & Tissue Kit (Qiagen), following the manufacturer's established protocol. DNA optical densities were measured using a NanoDrop 8000 (Thermo Scientific, Wilmington, DE, USA). An aliquot of each DNA sample was diluted to normalize to under 100ng/µl. Isolation of DNA from human liver samples for the DMET array and Deep Sequencing efforts are detailed in the cited references of that section.

Overview of Genetic Analysis:

Genotyping of DNA samples collected from study participants over a period of 18 months. Of the DNA samples collected, 94 were subjected to Deep Sequencing of the genes CYP2C9, CYP2D6, CYP3A4, CYP3A5, CYP4F2, CYP4F11, GGCX and VKORC1. In a pilot test of the Fluidigm platform, 95 DNA samples were analyzed to capture 37 allelic sites (all single nucleotide polymorphisms, SNPs) and evaluated for accuracy, call rate, population-specific frequencies of known pharmacogenetic variants, and assay functionality. In the second Fluidigm run, 95 new DNA samples were analyzed to capture 44 allelic sites and evaluated in the same manner. The difference in the number of allelic sites evaluated in the second and first runs reflects the removal and addition of SNPs based on inclusion criteria and the discovery of population specific variants from the Deep Sequencing effort. A total of 359 samples and 60 allelic sites were included in the 3rd and final Fluidigm run (this is referred to later as "final AN samples set"). As described below, a PCR specific target amplification step was performed prior to this final Fluidigm analysis.

A list of the final 60 allelic sites from ten different genes (VKORC1, GGCX, CYP2C9, CYP2D6, CYP3A4, CYP3A5, CYP4F2, CYP4F11, CYP2R1, VDR) is found in Table 1. It contains all common variants found in those genes by deep sequencing of the Yup'ik DNA samples, common variants found in other populations of the world that are known to alter enzyme function, coding variants in the ten target genes found uniquely in the Yup'ik DNA samples, and those reported recently by NWA-PGRN investigators following a similar analysis of DNA samples collected from the Confederated Salish and Kootenai Tribes of western Montana.[23]

In an additional Fluidigm run, 24 Yup'ik DNA samples that had a significant amount of "No Calls" across many, but not all assays, were re-run. As a comparator and another validation of the Fluidigm call rate and reproducibility, 46 DNA samples from the University of Washington, Department of Pharmaceutics Liver Bank were included on the same plate.

Multiplex PCR Specific Target Amplification (Pre-Amp) for Fluidigm (applies to final AN sample set, liver bank and "no call" runs):

DNA samples were pre-amplified following Fluidigm's (South San Francisco, CA) Specific Target Amplification protocol[24] to increase available template DNA for genotyping reactions. A total of 60 forward and reverse primer pairs in the form of ABI's *TaqMan® Genotyping Assays* at a concentration of 0.2X TaqMan Assays (Applied Biosystems Inc., Foster City, CA, USA) were added to each PCR reaction well (i.e., multiplexed). To each well was added a ~100ng/ul aliquot of DNA for each study participant. Reactions were conducted in 5µl volumes containing 1.25µl DNA, 1.25µl of 0.2X pooled TaqMan assay mix, and 2.5µl of TaqMan PreAmp Master Mix 2X (Applied Biosystems Inc.). Reactions were thermal cycled on an Applied Biosystems 2720 as follows: 95°C hold for 10 minutes followed by 14 cycles

of 95°C for 15 seconds and 60°C for 4 minutes. Pre-amplified samples were then diluted 1:5 with DNA Suspension Buffer (10 mM Tris, pH 8.0, 0.1mM EDTA).[25,26]

SNP Genotyping by Fluidigm™ (applies to final AN sample set (366), liver bank and" no call" runs):

Multiplexed TaqMan SNP Genotyping Assays (Applied Biosystems, Inc.) were run using the Fluidigm Biomark 96.96 Dynamic Genotyping Arrays (according to the manufacturer's established protocol for BioMark 96.96 Genotyping. DNA samples were analyzed un-replicated, except for: *CYP2C9*2, *3, *11, *13, *14; CYP2D6*6, *10, *17, *41*; *CYP3A4*1G, *2, *7, *16, *18, exon-10 Splice; CYP3A5*3, *5, *7, H30Y, N398T*; *CYP4F2*3, G185V, W12G; CYP4F11 R276C, N46D*; and *VKORC1 -1639GA, -1173GT, V29L;* that were run in duplicate in order to utilize an entire plate, maximize accuracy of calls and provide full coverage of important SNPs. Dynamic Arrays were primed and loaded on the Fluidigm HX and thermal cycled on the Fluidigm FC1 controller following pre-set programs in the instruments (Thermal Mix: 50^0C-2 min, 70^0C-30 min, 25^0C-10 min; Erase: 50^0C-2 min; Hotstart: 95^0C- 10 min; PCR; 95^0C-5 sec, 60^0C-60 sec, 25^0C-10 sec x 16 cycles) End-point fluorescence was read on a Fluidigm BioMark Real-Time PCR System and analyzed using SNP Genotyping Analysis software (Fluidigm).[24] Quality Control samples were checked and the data analyzed for allele frequencies, concordance, call rate and Hardy-Weinberg equilibrium. Across 24 SNP assays (36 possible) and 190 DNA samples, Fluidigm and Fluidigm w/ Pre- Amp results were compared. No calls and concordance were counted and recorded by observation, tallied and arithmetic means calculated on percents.

Deep Sequencing vs Fluidigm Data Comparison:

Ninety-four DNA samples from the Yup'ik population were submitted for deep sequencing of eight candidate genes, as part of the larger research objectives of the NWA-PGRN. Thirty-six of these sequenced samples overlapped with samples run on the Fluidigm platform, permitting a head-head comparison of SNP calls. Briefly, exons, adjacent intronic and 5'- and 3'-flanking regions of the CYP2C9, CYP2D6, CYP3A4, CYP3A5, CYP4F2, CYP4F11, GGCX and VKORC1 genes were subjected to Sanger sequencing. PCR amplicons ~500–600 bp in size were designed to provide full coverage over all regions of interest; adjacent amplicons contained overlapping segments of ~150 bp to validate all sequences underneath primer binding sites and to rule out allele-specific amplification. The PCR primers contained a universal M13-tailed PCR sequence to standardize the sequencing reactions. Amplicons were purified, diluted, and used in conventional Sanger sequencing reactions using BigDye 3.1 chemistry under standard conditions and separated on an ABI 3730XL DNAanalyzer (Applied Biosystems, Hercules, California, USA). Please see Pharmacogenetics and Genomics 2013, 23:403–414 for additional detail.[23] Individual SNP calls for the Deep Sequencing and Fluidigm SNP platforms were recorded. Across 19 SNP assays and 36 DNA samples, Fluidigm and Deep Sequencing results were compared. "No calls" and concordance were counted and recorded by observation, tallied and arithmetic means calculated on percents.

DMET vs Fluidigm Data Comparison:

In this Fluidigm run, 24 Yupik DNA samples that had a significant amount of "no calls", across many but not all assays during previous runs, were reanalyzed. As a comparator and another validation of Fluidigm call rate and reproducibility, 46 samples from the University of Washington, Department of Pharmaceutics Liver Bank were also included on the plate.

Genomic DNA was extracted from the individual livers and genotyped using Affymetrix DMET Plus Array (Santa Clara, CA) according to the manufacturer's protocol. Individual genotypes were resolved using the Affymetrix DMET Plus console (version 1.1) (Y. Lin, K. Thummel, A. Rettie, and M. Rieder, unpublished results).[27] All Fluidigm preparation and PCR parameters were run as described in the *Materials & Methods section* on the Pre-Amp protocol. Individual SNP calls for the different platforms were recorded. Across 27 SNP assays and 46 DNA samples, Fluidigm and DMET genotyping data were compared. “No calls” and concordance were counted and recorded by observation, tallied and arithmetic means calculated on percents.

RESULTS

The results of each comparison was used to assess the performance of the Fluidigm platform for one or more of the following specific performance characteristics: (1) frequency of “no call” results for each SNP and in aggregate; (2) reproducibility of SNP calls after introduction of a gene targeted pre-amplification step prior to Fluidigm analysis; (3) concordance of Fluidigm calls with previously run samples analyzed on other platforms.

Fluidigm vs Fluidigm w/ Pre-Amp. As represented in Figure 2A, for nearly all of the allelic sites evaluated, there was good resolution of the three genotype groups expected for a bi-allelic variant. Comparison of the first two runs of Fluidigm and the third run conducted with a Pre-amp step was only possible on 36 of 60 assays. This is because the first three runs had a different constellation of SNPs per run but also some of the assays in the first run weren’t optimized and yielded uninterpretable results. Of the 36 assays, 24 were sufficiently optimized and had all 190 samples from the first two runs also on the third run. Here, Fluidigm w/ Pre-Amp had significantly fewer no calls than Fluidigm w/ out, with the average

"no call" rate (NCR) being 0.17 and 1.96 percent, respectively (Table 2). Concordance was high, at 98.9%, with the highest non-concordance rate seen with the *CYP2D6*4* assay at 6.3%. This result demonstrates a reduction in NCR with inclusion of the Pre-Amp step, and reproducibility within Fluidigm platform and between Fluidigm runs. Further reproducibility was demonstrated within plates run including the Pre-Amp step. As mentioned in the Methods section, duplicates were included for some assays to fill up the additional wells. For replicate comparisons in the final AN sample set, there was a 100% call rate (data not shown).

Fluidigm w/ Pre-Amp vs Deep Sequencing. Deep sequencing over the genes encoding CYP2C9, CYP2D6, CYP3A4, CYP3A5, CYP4F2, CYP4F11, GGCX and VKOR revealed only 19 SNP's (10 novel) of significant frequency (> 0.5%) in either this AN population (n=94) or an AI population (n=94) of an associated study.[23] The 36 samples from the Fluidigm w/ Pre-Amp run that overlapped with the 94 deep sequenced samples demonstrated that Fluidigm w/ Pre-Amp performed very well (Table 3). At almost 100% (0.994) concordance and a zero NCR, Fluidigm w/ Pre-Amp out performed Deep Sequencing, which exhibited a NCR as high as 6.3% and an average of 2.74 % across the genes and SNPs assayed. In fact, 8 SNPs picked up by deep sequencing had an NCR over 3% (range 3.2- 6.3%), some of the highest in the study. This result demonstrates a lower NCR than Deep Sequencing, likely due to the Pre-Amplification of poor quality DNA samples in the Fluidigm protocol. It also demonstrates high reproducibility via the concordance with a platform considered to be the "gold standard" of gene analysis.

Average Call Count and NCR across all permutations from all DNA samples. Results presented in Table 4 show the average number of samples called across 41 SNP assays of 366 samples from the final Fluidigm run (Avg. = 348.1 range 332-362). This average call count is compared to the NCR over all Fluidigm runs (19 common SNPs assayed for 424 samples). Also shown for comparison is the SNP call rate for just the 94 samples subjected to deep sequencing. As our goal was to accurately genotype 350 participants and we analyzed 366 participant samples via the Fluidigm platform, the average successful call count represents a reasonable outcome. By comparing this average call count to our other measure (NCR) across study platforms (Fluidigm and deep sequencing) and in aggregate, it can be seen that performance over all was consistent and no platform stood out as a culprit in "no call" events. This result demonstrated reproducibility by showing that the NCR is comparable between the two sample sets (n=366 and n=424), and only ~2% higher than the n=94 deep sequencing sample set alone.

DMET comparison to Fluidigm w/Pre-Amp. In the Fluidigm w/ Pre-Amp versus the DMET© comparison (Table 5), Fluidigm w/ Pre-Amp performed almost perfectly and best overall. Over 27 of the same SNP assays conducted on both platforms and for 46 liver samples, there was only 1 "no call" (Fluidigm) and one assay that demonstrated less than 100% concordance (91.8% - *CYP2C9*2*). However, it should be noted that 15 of the assays performed on the Fluidigm platform were duplicate runs, with *CYP2C9*2* being one of them. It called the sample T:C twice, while the DMET chip called it as TT. It was the only such anomaly in the experimental results. Overall this result was remarkable by producing a negligible NCR and non-concordance rate across the two different platforms. Again, this demonstrates reproducibility and comparable accuracy of the Fluidigm platform.

Another validation of Fluidigm w/ Pre-Amp via Deep Sequencing. The allele frequency and count table in Table 6 shows that Fluidigm w/ Pre-Amp performed reasonably well in achieving our objective of 350 sample results per SNP assay. A total of 21 out of 60 (35%) of the assays fell slightly below the n =350 objective due to "no calls". All assays, except for the *CYP2C9 M1L* variant (6% vs 10%, respectively) echoed the minor allele frequencies (MAF) found via Deep Sequencing in this AN population. The difference in the MAF of the *CYP2C9 M1L* variant could have been the result of a difference in sample selection. Deep sequencing was performed on samples from participants that were as unrelated as possible, whereas the Fluidigm analysis was performed on a random cross-section of samples collected during the first two years of recruitment. Nonetheless, the Fluidigm w/ Pre-Amp performed well against Deep Sequencing and the Affymetrix DMET array and demonstrates reproducibility across platforms.

DISCUSSION:

As noted in the results section, the Fluidigm and the ABI Taqman Genotyping assays performed very well overall, but suffer from five minor drawbacks – 3 technical and 2 expository: 1) The single plate system precludes adjustment for varied assay conditions; 2) insufficient overlap and adjustment controls when superimposing many runs of data; 3) ABI's probe design precludes accurate calls of SNPs that are too close together (~5 bp before and after the target SNP); 4) they do not allow for elucidation of new exomal SNPs or regulatory SNP in the intronic regions; 5) positive controls can be nullified by competing assay conditions between assays.

In the first run of the Fluidigm platform, several of the 37 original SNP assays clearly were suffering from optimization problems. Of the assays that did work, some, like *CYP2C9*2* and

*3, had several “no calls”. Some of these calls were evaluated in parallel by direct Sanger sequencing, along with a control from each allele combination (data not shown). What was elucidated by this QC process was that some DNA samples were corrupted enough that calls were not possible even by direct sequencing and that some assays required different PCR conditions. Another benefit from this evaluation process was the realization that many SNPs described and observed for the three genogeographic groupings of European, Asian and African were not found at any frequency amongst the AN population of this investigation. Accordingly, other possibly salient SNP assays, based on direct sequencing results, were added and irrelevant SNP assays were removed from subsequent runs.

After the second run (comprised of 44 SNP assays), where certain SNP assays had been swapped out for assays that ran well under uniform PCR conditions, many improvements in platform performance were achieved. However, one problem that remained. This was, that for certain assays, the delineation between allele groupings on the scatterplots was less than optimal (Figure 2B), resulting in a “no man’s land” of ambiguous allelic assignment. In this case, Sanger sequencing was used to confirm allele calls and reclaim as many “no calls” as possible.

It was shortly after this experiment that it was determined by the research group that Deep Sequencing could serve as a QC step for some of the minor alleles that we were seeing, in addition to guiding our SNP selection process and the discovery of novel gene variants in the Y-K delta Yup’ik population. Once sequencing was completed, the Fluidigm platform was modified to reflect this population’s allele constitution. However, we still had to contend with poor separation of allele groupings in certain assays and the reclamation of “no calls”. These technical difficulties prompted us to seek and adopt the Multiplex PCR Specific Target Amplification protocol. In addition, for added reliability, we decided to run SNP assays of particular import for the study, or SNP assays that were previously problematic, in

duplicate on the Fluidigm plate. This strategy was successful and all assays demonstrated sufficient spread and spacing between groups on their respective scattergram to provide confidence with the accuracy of calls. Where there were still some questions in sample calls, usually the duplicate answered the question and reclaimed “no calls”.

As can be seen from the results, the Pre-Amp step resulted in significant improvement from the first two Fluidigm runs. Furthermore, concordance and NCR were very comparable across platforms and in the statistical recapitulation of MAFs in this population. The performance of Fluidigm compared to Affymetrix’s DMET chip was also reassuring, and provided confirmatory support for the validity of the full scale AN sample set study results. That the technologies are significantly different and spaced far apart chronologically lends much credence to the reliability of the Fluidigm platform.

Concordance across platforms provided confidence in the accuracy and reproducibility of the Fluidigm w/Pre-Amp’s performance, yet the “no calls” that were observed were still worrisome. The problem may be with the SNP assays themselves; that is, trying to analyze a particularly difficult sequence of the genome. The fact that the Deep Sequencing analysis also yielded a number of “no calls” hints at this possibility. Comparing across platforms, it was either the same assays or assays in the same gene that demonstrated a high percentage of “no calls”. This could be due to other unseen variants near the probe SNP, a possibility that was made aware to us by ABI’s technical service team when trying to design probes for non-stocked SNP assays. In addition to seeing “no calls” for the same or similar SNP assays across platforms, it was also observed that a small minority of samples received “no calls” at every SNP site that was probed, even after re-genotyping and with a different SNP detection platform (data not shown). This is likely the result of poor quality DNA.

That the “no call” rate was very much improved after the addition of the Pre-Amp step also lend credence to the rehabilitation of low quality or corrupted genetic samples. In a paper by Matt Smith et. al. 2010, where they genotyped over 760 museum collected and historic fish scales, the Pre-Amp step decreased their “no call” rate from 100% to an average of 1.13% as reflected in their improved the separation of their allele groupings where there had been none previously.[26] As a control, they also genotyped high quality isolations of DNA from contemporary fish scales and compared it to diluted samples (1:20) of the same. The “no call” rates were similar at 0.66% and 0.69%, respectively. Their two main concerns were allelic drop out from poorly preserved scales yielding low quality of low concentration DNA and allelic drop in from over amplification. They, like us, saw evidence of neither with the employment of the Pre-Amp protocol.

The lack of genomic exposition demonstrated by Fluidigm is not surprising, but it is an advantage provided by the other platforms, such as DMET©. In contrast, whole genome sequencing has the potential to uncover variation in the genes of immediate interest as well as all other genes in our genome, some that might become of new research interest to investigators (particularly if results are shared) but not approved for investigation by the participating tribal community. However, it is important for the purpose of pharmacogenetic applications to identify all population significant variants (0.5-1%) in the pre-selected genes of interest, as was done by other members of the NWA-PGRN team. Deep sequencing can uncover CNV’s of various kinds such as indels, transversions, repeats, splices and also functionally different, non-synonymous variants like *CYP2C9 L1M*, which was identified in this AN population. Utilizing the Fluidigm platform, you sacrifice this advantage and SNP identification and selection must be performed separately and thus progress moves at a slower pace. However it is far less costly overall to take a two-stage approach and perform deep sequencing on the genes of interest in a sample set large enough to capture variants

present at a preset threshold allele frequency (eg, 1 %), and then utilize an informed Fluidigm platform thereafter ascertain accurate population frequencies and for genotype-phenotype association studies.

Lastly, both the DMET and deep sequencing option have a more reliable assurance of positive controls than does the Fluidigm platform. For DMET, it is built into the chip and for deep sequencing, it is read-depth. In the case of the Fluidigm, platform, the clustering of sample signals for different genotype groups (i.e., the power of scatterplots of genotype calls for all samples interrogated on the plate) provide a unique element of platform control, along with the potential for individual sample replication. In this study, we also had the advantage of positive control samples identified via deep sequencing and could track those samples across the scatterplots.

Cost and Time Comparison

A comparison between Fluidigm and DMET platforms yielded an estimated cost of $3-6 and $230 –$300 per sample, respectively, not including technician time and labor.[26,28,29] Estimates of labor time are ~12 hrs per 190 samples for sample prep and processing using the Fluidigm platform and ~64hrs from DNA to results for assaying by the DMET chip for the same number of samples. It has been estimated in one paper that the Pre-Amp step for Fluidigm would add a 5% total increase to costs for a given project.[26] In addition, in our study, some time was devoted to optimizing and selecting assays for the final run which undoubtedly increased cost compared to both DMET and Deep Sequencing because both of those technologies are fixed with regard to quality control (QC) and assay conditions. Moreover, Fluidigm has the disadvantage of applying a single set of PCR conditions (on a single plate) for multiple variants, with the high likelihood that the conditions are not optimal

for all variants. However, the Pre-Amp step reduced the impact of this problem for all the SNP assays conducted in this project. Overall, we find the Fluidigm platform to be superior in terms of cost with other available technologies.

Although we made no comparisons here to Taqman Assays run on the ABI 7900, one estimate of cost by this method per sample and allelic site was $11.38.[30] An analysis of 60 different SNP sites, as performed here, would be quite expensive, in comparison to Fluidigm costs. Moreover, the ABI7900 uses considerably more sample, always an important consideration, especially in regard to remote and vulnerable populations like AI/AN groups. Additionally, while deep sequencing has costs attached and those costs drop every day, the nature and scope of deep sequencing is altogether too different to make a meaningful comparison here. It should be mentioned however, that as costs for deep sequencing drops, the temptation for researchers in AI/AN communities to consider this method first will become much greater. If that is the case, it will be important that all involved (tribal community, heathcare providers and research investigators) understand at the outset the full scope of data collection, how it will be stored and who will have access, both for the immediate investigation and in the future. One might anticipate agreements that assign control of such genome-wide data with tribal authorities, and that are re-evaluated by both parties as the study proceeds, with the option of terminating the study.

One additional time/labor expense consideration occurs in the analysis of data. Affymetrix provides a program to analyze the DMET data[31] and sequencing labs generally have a fleet of researchers continually updating and sourcing programs to quickly perform QC checks and interpret their data. Fluidigm has no such data analysis program. There are few toggles within the display program to provide minor allele frequencies, “no call” rates and genotype counts, but the task of QC for each SNP assay must be conducted by the experimenter. In this regard, there can at times be difficulty in delineating borders between SNP genotype

groups (Figure 2B, for example). In addition, the comparison and tabulation between runs has to be performed by hand. The company claims that data from different runs can be superimposed upon one another, but because of vagaries of fluorescence signals (as recorded in “no template” wells), we found this often not possible. But once again, the Pre-Amp step reduced much of the impact of this inherent limitation with the platform. As Fluidigm with Pre-Amp and appropriate replicates (for problematic assays) performed very well when compared to the “gold standard” of Deep Sequencing, and at low cost and run time, we conclude that it is an excellent choice for pharmacogenetic research with AI/AN populations.

THESIS RESEARCH IN CONTEXT:

Indigenous people have historically experienced financial, cultural, political, social and bio-medical disadvantages. Disparities abound. The rise of genomic information and the development of rapid, low-cost testing technologies has led to the introduction of tailored and individualized medicine into clinical care; lack of access to such technology or participation in the research to develop it could be a new source of health disparity.[32] However, the historical conflicts between the federal government and tribal communities, their lack of inclusion in research development and interpretation make such an investment of group trust in research a difficult proposition. These issues demonstrate a need for a change of research approach in improving AI/AN health care.

This study with its attention to community based participatory research, a two-step procedure in SNP identification, and the use of the discrete and cost effective Fluidigm platform, represents such a change. Knowledge that can be generated with proper application of the Fluidigm platform can be empowering to the Yup’ik people and other

indigenous people of the US by filling-in the gaps between gene variants in metabolizing enzymes and drug response. As stated before, and it cannot be overemphasized, awareness and the education of health providers in AI/AN communities about these potential associations can be invaluable to the improvement of health in these communities.

In addition to possible differences in AI/AN health care, researchers who work with AI/AN communities must equally recognize the importance and historical foundation of tribal/group/and collective protections of research data. In this investigation, the Fluidigm platform succeeds enormously in the goal of balancing respect for genomic privacy while delivering accurate and efficient genomic testing. It conserves sample volume, it yields reliable results, it is quickly modifiable both in specific sample and assay selection and it has a high and rapid throughput. All of these characteristics make it amenable to discretion in that AI/AN groups do not have to supply researchers with large or repeated collections of DNA, pilot and preliminary findings can be reported back to tribal entities for appraisal and approval, and once done can be modified for continuing studies.

Table 1. List of SNPs and corresponding genes included in the "final AN sample set" Fluidigm™ analysis.

AssayNumber	Gene	Variant	AssayNumber	Gene	Variant
1	CYP 2C9	*2(rs1799853)	31	CYP 4F2	(rs2189784) g72220026G
2	CYP 2C9	*3(rs1057910)	32	CYP 4F2	*3/V433(rs2108622)
3	CYP 2C9	*8(rs7900194)	33	CYP 4F2	Splice C>G 22312
4	CYP 2C9	*13(rs72558187)	34	CYP 4F2	Gly**185**Val (rs3093153)
5	CYP 2C9	*11(rs28371685)	35	CYP 4F2	Met**519**Leu (rs3093200)
6	CYP 2C9	*14(rs72558189)	36	CYP 4F2	Gly**12**Tryp (rs3093105)
7	CYP 2C9	Leu**1**Met T>A			
8	CYP 2C9	Iso**218**Asp T>A	37	CYP 4F11	Gly**12**Arg
9	CYP 2C9	Thre**279**ProA>C	38	CYP 4F11	Arg**276**Cys (rs8104361)
			39	CYP 4F11	Asn**446**Asp (rs1060463)
10	CYP 2R1	rs10741657			
11	CYP 2R1	rs2060793	40	GGCX	rs11676382(intron 14)
12	CYP 2R1	rs11023374	41	GGCX	Arg**325**Gln (rs699664)
13	CYP 2R1	rs1993116	42	GGCX	Ala**421**Gly
14	CYP 3A4	*1G (rs 22422480)	43	VKORC1	1639 G>A (rs9923231)
15	CYP 3A4	*2 (rs55785340)	44	VKORC1	1173 G>T (rs9934438)
16	CYP 3A4	*7 (rs56324128)	45	VKORC1	Asp**29**Leu (rs28940302)
17	CYP 3A4	*16 (rs12721627)	46	VKORC1-L	rs4128574
18	CYP 3A4	*18 (rs28371759)			
19	CYP 3A4	*20(rs67666821)	47	VDR APA	rs7975232
20	CYP 3A4	*22 (rs35599367)	48	VDR BSM	rs1544410
21	CYP 3A4	Exon10Splice (rs4646438)	49	VDR FOK	rs2228570
			50	VDR Taq	rs731236
22	CYP 3A5	*3 (rs776746)			
23	CYP 3A5	*6 (rs10264272)	51	VDR-GATA	rs4516035 (A-1012G)
24	CYP 3A5	*7 (rs41303343)			
25	CYP 3A5	His**30**Thr(rs28383468)	52	VDR-Cdx-2	rs1156820
26	CYP 3A5	Ile**488**Thr (rs28365085)	53	VDR-Cdx-2	rs11574010
27	CYP 3A5	Asn**398**Thr (rs28365083)			
28	CYP 3A5	Del7797A	54	CYP 2D6	*3 (rs35742686)
29	CYP 3A5	Pro**484**Leu	55	CYP 2D6	*4 (rs3892097)
30	CYP 3A5	G**458**del	56	CYP 2D6	*6 (rs5030655)
			57	CYP 2D6	*10 (rs1065862)
			58	CYP 2D6	*17 (rs28371706)
			59	CYP 2D6	*41 (rs28371725)
			60	CYP 2D6	Arg**25**Tryp

Table 2. Fluidigm™ w/out Pre-Amp vs Fluidigm™ w/ Pre-Amp

Name	CYP2C9*2	CYP2C9*8	CYP2C9*13	CYP2C9*11	CYP2C9*14	CYP2D6*4	CYP2D6*6	CYP2D6*10
rs#	rs1799853	rs7900194	rs74052158/rs72558187	rs28371685	rs72558189	rs3892097	rs5030655	rs1065862
FluPre-Amp NCR	0	0	0	0	0	0	1.5	1.5
Flu w/o NCR	1.1	0.5	1.1	2.1	1.6	3.2	1.1	2.1
Non Concordance	0	0	0	0	11.1	0	6.3	0
Name	CYP2D6*17	CYP2R1	CYP2R1	CYP3A4*1G	CYP3A4*7	CYP3A4*16	CYP3A4*18	CYP3A5*3
	rs28371706	rs10741657	rs11023374	rs2242480	rs56324128	rs12721627	rs28371759	rs776746
FluPre-Amp NCR	0.5	0	0	0	0	0	0	0
Flu w/o NCR	4.2	1.1	1.6	2.6	2.6	2.6	2.6	1.1
Non Concordance	0	1.1	1.6	1.1	0	0	0	1.1
Name	CYP3A5*6	CYP3A5*7	CYP4F2g72220026G	CYP4F2*3	GGCX-intron14	VKORC1-1639GA	VKORC1-1173GT	VKORC1V29L
	rs10264272	rs41303343	rs2189784	rs2108622	rs11676382	rs9923231	rs9934438	rs28940302
FluPre-Amp NCR	0	0	0	0.5	0	0	0	0
Flu w/o NCR	2.6	2.1	2.1	2.6	1.6	1.1	1.6	2.1
Non Concordance	0	1.6	1.1	0.5	0	0.5	0.5	0

FluPre-Amp NCR	x = 0.17
Flu w/o NCR	x = 1.96
Concordance	x = 98.9

Table 3. Concordance between Fluidigm™ and Deep Sequencing results.

Name	CYP2C9*2	CYP2C9*3	CYP2C9L1M	CYP2C9I218D	CYP2C9T279P	CYP2D6*4	CYP2D6*10
rs#	rs1799853	rs1057910	CYP2C9L1M	CYP2C9I218D	CYP2C9T279P	rs3892097	rs1065862
Fluidigm (pre-amp)NCR	0	0	0	0	0	0	0
Exome Seq NCR	0	3.2	2.1	1.1	1.1	1.1	5.3
Allele Concordance	100	98.9	100	100	100	100	100

Name	CYP2D6*41	CYP2D6R25W	CYP3A4*1G	CYP3A5-458del	CYP4F2*3	CYP4F2 Splice	CYP4F2W12G
rs#	rs28371725	CYP2D6R25W	rs2242480	CYP3A5-G-del	rs2108622	CYP4F2SplCG	rs3093105
Fluidigm (pre-amp)NCR	0	0	0	0	0	0	0
Exome Seq NCR	2.1	4.3	5.3	3.2	6.3	1.1	0
Allele Concordance	97.2	100	100	100	100	100	100

Name	CYP4F11G12R	CYP4F11R276C	CYP4F11N446D	GGCX R325Q	GGCX-intron14	19 SNPs
rs#	CYP4F11G12R	rs8104361	rs1060463	rs699664	rs11676382	94 Subject Sequenced
Fluidigm (pre-amp)NCR	0	0	0	0	0	36 Overlapping Subjects
Exome Seq NCR	2.1	1.1	2.1	5.3	5.3	2.74% Avg. No Call Rate
Allele Concordance	100	100	100	100	100	99.8% Concordance

Table 4. Average number of samples called from 41 assays with percentage of “No Calls” comparison.

Total	348.1
NCR366	4.466667
NCR424	4.710526
NCR94	2.742105

Table 5. Fluidigm™ - DMET© concordance and "no calls"

Name	CYP2C9*2	CYP2C9*3	CYP2C9*13	CYP2C9*11	CYP2C9*14	CYP2D6*6	CYP2D6*10	CYP2D6*17	CYP2D6*41
rs#	rs1799853	rs1057910	rs74052158/rs7	rs28371685	rs72558189	rs5030655	rs1065862	rs28371706/rs1	rs28371725
Concordance	91.8	100	100	100	100	100	100	100	100
No Calls	0	0	0	0	0	0	0	0	0
Name	CYP3A4*1G	CYP3A4*2	CYP3A4*7	CYP3A4*16	CYP3A4*18	CYP3A4Ex10Sp	CYP3A5*3	CYP3A5*6	CYP3A5H30Y
rs#	rs2242480	rs55785340	rs56324128	rs12721627	rs28371759	rs4646438	rs776746	rs10264272	rs28383468
Concordance	100	100	100	100	100	100	100	100	100
No Calls	0	0	0	0	0	0	0	0	0
Name	CYP3A5N398T	CYP4F2*3	CYP4F2G185V	CYP4F2W12G	CYP4F11R276C	CYP4F11n446D	VKORC1-1639G	VKORC1-1173G	VKORC1V29L
rs#	rs28365083	rs2108622	rs3093153	rs3093105	rs8104361	rs1060463	rs9923231	rs9934438	rs28940302
Concordance	100	100	100	100	100	100	100	100	100
No Calls	0	0	0	0	0	1	0	0	0

Table 6. Allele frequencies, counts, No Calls and Hardy-Weinberg p-values for 57assays and 10 Genes

Cdx-2	rs11568820					Cdx-2	rs11574010				
CC	TC	TT	Total	C freq	T freq	CC	TC	TT	Total	C freq	T freq
322	28	4	354	0.949153	0.050847	338	0	0	338	1	0
12 No Calls				HWE p=	0.0007	28 No Calls				HWE p=	X
CYP2C9*2R144C	rs1799853					CYP2C9*3I359L	rs1057910				
CC	TC	TT	Total	C freq	T freq	CC	AC	AA	Total	C freq	A freq
408	2	3	413	0.990315	0.009685	1	22	379	401	0.029925	0.972569
11 No Calls				HWE p=	0	20 No Calls				HWE p=	0.35
CYP2C9*8H150R	rs7900194					CYP2C9*13P90L	rs74052158/rs72558187				
GG	GA	AA	Total	G freq	A freq	CC	TC	TT	Total	C freq	T freq
354	0	0	354	1	0	0	0	356	356	0	1
12 No Calls				HWE p=	X	10 No Calls				HWE p=	X
CYP2C9*11R335\rs28371685						CYP2C9*14	rs72558189				
CC	TC	TT	Total	C freq	T freq	GG	GA	AA	Total	G freq	A freq
353	0	0	353	1	0	361	0	1	362	0.997238	0.002762
13 No Calls				HWE p=	X	4 No Calls				HWE p=	0
CYP2C9L1M	No rs#					CYP2C9I218D	No rs#				
AA	AT	TT	Total	A freq	T freq	AA	TA	TT	Total	C freq	T freq
4	43	358	405	0.062963	0.937037	373	25	0	398	0.968593	0.031407
12 No Calls				HWE p=	0.0437	19 No Calls				HWE p=	0.52
CYP2C9T279P	No rs#					CYP2D6*3	rs35742686				
CC	AC	AA	Total	C freq	A freq	Del	Tdel	TT	Total	Del freq	T freq
331	16	1	348	0.974138	0.025862	1	0	331	332	0.003012	0.996988
11 No Calls				HWE p=	0.1	27 No Calls				HWE p=	0
CYP2D6*4 rs3892097□						CYP2D6*6 rs5030655□					
GG	AG	AA	Total	G freq	A freq	AA	Adel	Del	Total	A freq	Del freq
335	55	13	403	0.899504	0.100496	336	0	0	336	1	0
21NoCalls				HWE p=	0	30 No Calls				HWE p=	X
CYP2D6*10P34S	rs1065862					CYP2D6*17T107I	rs28371706/rs16947				
TT	TC	CC	Total	T freq	C freq	GG	AG	AA	Total	G freq	A freq
20	78	297	395	0.149367	0.850633	336	0	0	336	1	0
29 No Calls				HWE p=	0	30 No Calls				HWE p=	X
CYP2D6*41	rs28371725					CYP2D6R25W	No rs#				
CC	TC	TT	Total	C freq	T freq	CC	CT	TT	Total	C freq	T freq
392	4	1	397	0.992443	0.007557	389	2	0	391	0.997442	0.002558
27 No Calls				HWE p=	0	33 No Calls				HWE p=	0.95
CYP2R1	rs10741657					CYP2R1	rs2060793				
AA	AG	GG	Total	A freq	G freq	AA	AG	GG	Total	A freq	G freq
128	166	53	347	0.608069	0.391931	120	163	56	339	0.594395	0.405605
19 No Calls				HWE p=	0.947	20 No Calls				HWE p=	0.895
CYP2R1	rs11023374					CYP2R1	rs1993116				
CC	TC	TT	Total	C freq	T freq	GG	GA	AA	Total	G freq	A freq
21	112	220	353	0.21813	0.78187	53	166	126	345	0.394203	0.605797
13 No Calls				HWE p=	0.19	14 No calls				HWE p=	0.89
CYP3A4*1G	rs2242480					CYP3A4*2P222S	rs55785340				
GG	AG	AA	Total	G freq	A freq	AA	ADel	Del	Total	C freq	T freq
370	36	0	406	0.955665	0.044335	347	0	3	350	0.991429	0.008571
18 No Calls				HWE p=	0.35	14 No Calls				HWE p=	0
CYP3A4*7D56G	rs56324128					CYP3A4*16T185S	rs12721627				
CC	TC	TT	Total	C freq	T freq	GG	TG	TT	Total	G freq	T freq
346	0	0	346	1	0	356	0	0	356	1	0
20 No Calls				HWE p=	X	10 No calls				HWE p=	X

CYP3A4*18P293I	rs28371759					CYP3A4*20	rs67666821				
AA	AG	GG	Total	A freq	G freq	CC	TC	TT	Total	C freq	T freq
352	0	0	352	1	0	0	1	340	341	0.001466	0.998534
14 No Calls				HWE p=	X	17 No Calls				HWE p=	0.98
CYP3A4 (*22)	rs35599367					CYP3A4Exon10Splice		rs4646438			
GG	AG	AA	Total	G freq	A freq	Del	Del T	TT	Total	Del freq	T freq
342	1	0	343	0.998542	0.001458	343	0	0	343	1	0
16 No Calls				HWE p=	0.98	16 No Calls				HWE p=	X
CYP3A5*3	rs776746					CYP3A5*6K208K	rs10264272				
CC	TC	TT	Total	C freq	T freq	CC	TC	TT	Total	C freq	T freq
320	28	1	349	0.95702	0.04298	351	0	0	351	1	0
17 No Calls				HWE p=	0.64	15 No Calls				HWE p=	X
CYP3A5*7	rs41303343					CYP3A5H30Y	rs28383468				
Del	Del A	AA	Total	Del freq	A freq	GG	GA	AA	Total	G freq	A freq
354	0	0	354	1	0	344	0	0	344	1	0
12 No Calls				HWE p=	X	15 No Calls				HWE p=	X
CYP3A5T488I	rs28365085					CYP3A5N398T	rs28365083				
AA	AG	GG	Total	A freq	G freq	GG	TG	TT	Total	G freq	T freq
343	0	0	343	1	0	344	0	0	344	1	0
16 No Calls				HWE p=	X	16 No Calls				HWE p=	X
CYP3A5del7997A	No rs#					CYP3A5P484L	No rs#				
Del	Del A	AA	Total	Del freq	A freq	CC	TC	TT	Total	C freq	T freq
0	0	350	350	0	1	343	0	0	343	1	0
9 No Calls				HWE p=	X	16 No Calls				HWE p=	X
CYP3A5-458-G-d	No rs#					CYP4F2g72220026G		rs2189784			
CC	Del C	Del	Total	C freq	Del freq	CC	TC	TT	Total	C freq	T freq
407	1	0	408	0.998775	0.001225	139	162	51	343	0.641399	0.38484
9 No Calls				HWE p=	0.98	14 No Calls				HWE p=	0.73
CYP4F2*3M433V	rs2108622					CYP4F2SpliceCG	CYP4F2SpliceCG				
GG	AG	AA	Total	G freq	A freq	CC	GC	GG	Total	C freq	G freq
114	196	84	394	0.538071	0.461929	0	4	401	405	0.004938	0.995062
30 No Calls				HWE p=	0.989	12 No Calls				HWE p=	0.92
CYP4F2G185V	rs3093153					CYP4F2M519L	rs3093200				
CC	AC	AA	Total	C freq	A freq	CC	AC	AA	Total	C freq	A freq
342	3	0	345	0.995652	0.004348	350	0	0	350	1	0
14 No Calls				HWE p=	0.94	9 No Calls				HWE p=	X
CYP4F2W12G	rs3093105					CYP4F11G12R	No rs#				
GG	TG	TT	Total	G freq	T freq	CC	GC	GG	Total	C freq	G freq
2	23	372	397	0.034005	0.965995	0	15	386	401	0.018703	0.981297
20 No Calls				HWE p=	0.019	16 No Calls				HWE p=	0.7
CYP4F11R276C	rs8104361					CYP4F11N446D	rs1060463				
CC	TC	TT	Total	C freq	T freq	GG	AG	AA	Total	G freq	A freq
391	7	0	398	0.991206	0.008794	9	79	294	382	0.126963	0.873037
18 No Calls				HWE p=	0.86	35 No Calls				HWE p=	0.19
GATA-A1012G	rs4516035					GGCX-intron14	rs11676382				
CC	TC	TT	Total	C freq	T freq	CC	GC	GG	Total	C freq	G freq
54	161	138	353	0.38102	0.61898	407	3	0	410	0.996341	0.003659
13 No Calls				HWE p=	0.53	14 No Calls				HWE p=	0.94
GGCX R325Q	rs699664					GGCXG421A	No rs#				
GG	AG	AA	Total	C freq	T freq	GG	GC	CC	Total	G freq	C freq
108	198	84	390	0.530769	0.469231	343	0	0	343	1	0
27 No Calls				HWE p=	0.7	16 No Calls				HWE p=	X

VKORC1-1639GA	rs9923231				
CC	TC	TT	Total	C freq	T freq
23	91	240	354	0.193503	0.806497
12 No Calls				HWE p=	0.0009

VKORC1-1173GT	rs9934438				
AA	GA	GG	Total	A freq	G freq
240	89	23	352	0.808239	0.191761
14 No Calls				HWE p=	0.0005

VKORC1V29L	rs28940302				
GG	TG	TT	Total	G freq	T freq
353	0	0	353	1	0
13 No Calls				HWE p=	X

VKORC1-L	rs4128574				
GG	GC	CC	Total	C freq	G freq
345	0	0	345	1	0
16 No Calls				HWE p=	X

VDR APA	rs7975232				
CC	AC	AA	Total	C freq	A freq
121	168	61	350	0.585714	0.414286
16 No Calls				HWE p=	0.84

VDR BSM	rs1544410				
CC	TC	TT	Total	C freq	T freq
330	19	1	350	0.97	0.03
15 No Calls				HWE p=	0.2

VDR FOK	rs2228570				
GG	GA	AA	Total	G freq	A freq
110	169	76	355	0.547887	0.452113
11 No Calls				HWE p=	0.46

VDR Taq	rs731236				
GG	GA	AA	Total	G freq	A freq
2	14	338	354	0.025424	0.974576
12 No Calls				HWE p=	0.001

Figure 1. Illustration of Biomark HD System™ with 96 x 96 well Dynamic Array IFC™.

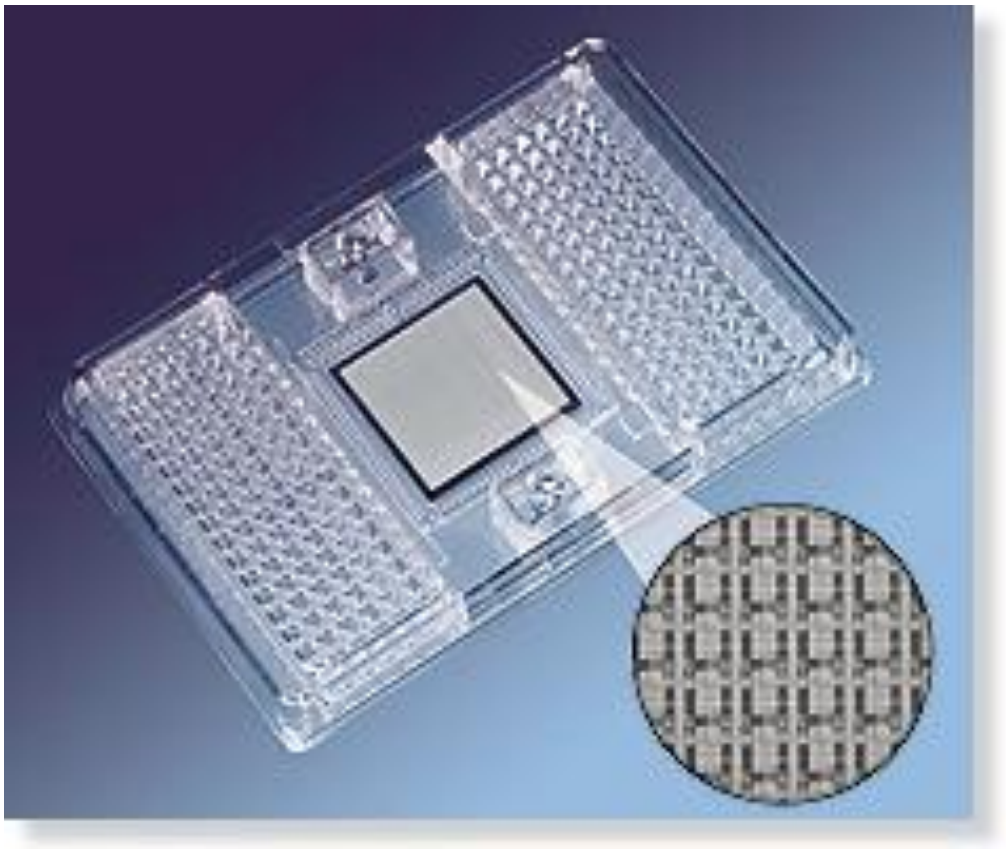
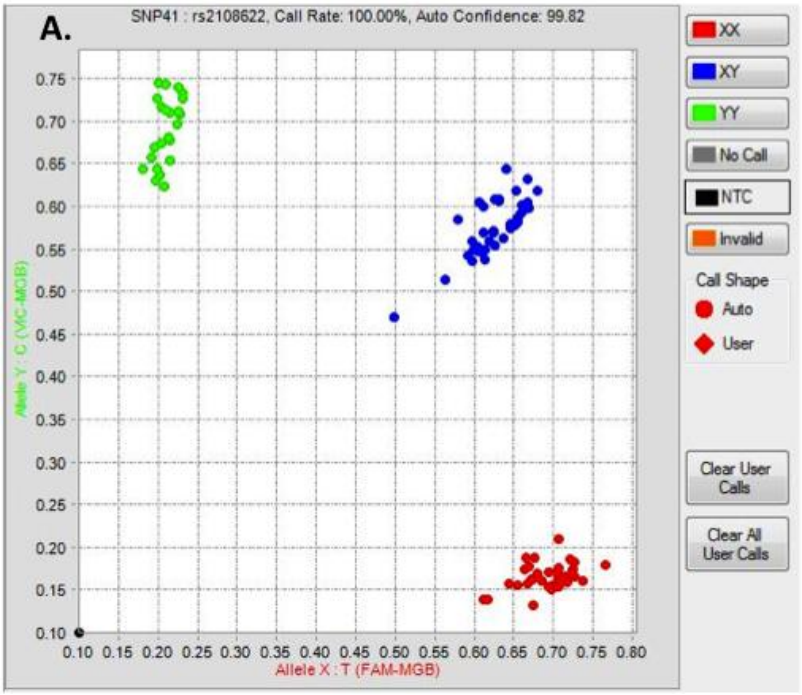
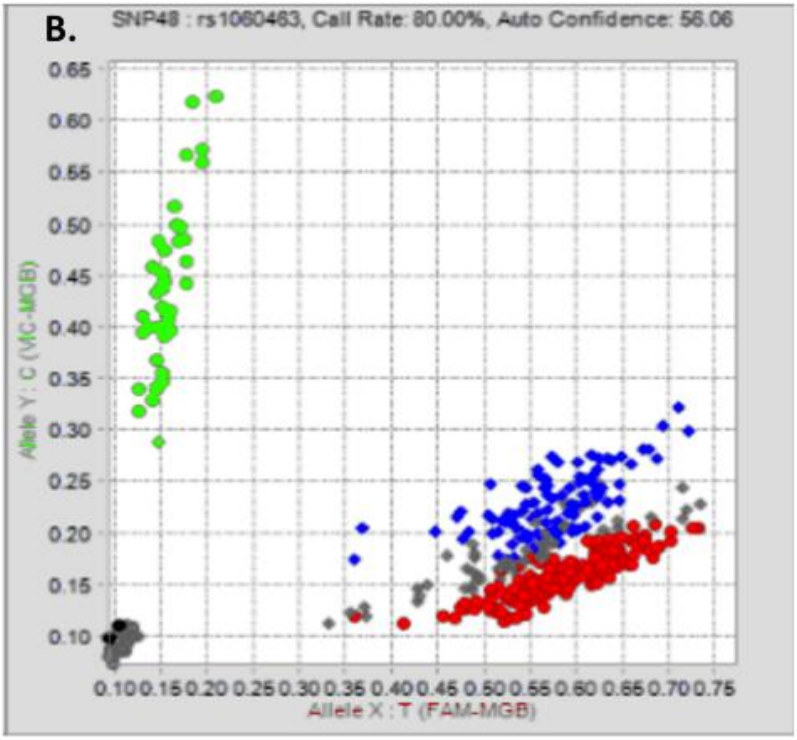


Figure 2.

Fluidigm fluorescence scattergrams that illustrate good (A) and poor (B) genotype resolution. *CYP4F2*3* and *CYP4F11*N446D (*rs1060463) are shown in (A) and (B), respectively.





REFERENCES:

1. Sciences, N. I. of G. M. Goals for the Pharmacogenetics Research Network, 2003-2005. at <http://www.nigms.nih.gov/Research/FeaturedPrograms/PGRN/Background/>
2. Adverse Reactions to Drugs and Metabolic Problems Perceived in Northern ... - Otto Schaefer - Google Books. at <http://books.google.com/books/about/Adverse_Reactions_to_Drugs_and_Metabolic.html?id=Y7m7YgEACAAJ>
3. Ethnic psychopharmacology: the Hispanic... [Psychopharmacol Bull. 1991] - PubMed - NCBI. at <http://www.ncbi.nlm.nih.gov/pubmed/1813894>
4. Ethnicity and psychopharmacology - Springer. at <http://link.springer.com.offcampus.lib.washington.edu/article/10.1007/BF00156581>
5. Cultural diversity and medication safety. at <http://www.ismp.org/newsletters/acutecare/articles/20030904.asp>
6. Is ethnicity associated with morphine's side effects in children? morphine pharmacokinetics, analgesic response and side effects in children having tonsillectomy. at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3366036/>
7. JAMA Network | JAMA Surgery | Postoperative Pain Assessment and Analgesic Administration in Native American Patients Undergoing Laparoscopic Cholecystectomy. at <http://archsurg.jamanetwork.com.offcampus.lib.washington.edu/article.aspx?articleid=1558114>
8. About | Pharmaceutics. at <http://sop.washington.edu/pgrn>
9. Malhi, R. S. *et al.* Distribution of Y chromosomes among Native North Americans: A study of Athapaskan population history. *Am. J. Phys. Anthropol.* **137,** 412–424 (2008).
10. Silva-Zolezzi, I. *et al.* Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc. Natl. Acad. Sci.* **106,** 8611–8616 (2009).
11. Jaja, C., Burke, W., Thummel, K., Edwards, K. & Veenstra, D. L. Cytochrome P450 Enzyme Polymorphism Frequency in Indigenous and Native American Populations: A Systematic Review. *Community Genet.* **11,** 141–149 (2008).
12. Alper, J. S. & Lange, R. V. Lumsden-Wilson theory of gene-culture coevolution. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 3976–3979 (1981).
13. Zerjal, T. *et al.* Geographical, Linguistic, and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations. *Mol. Biol. Evol.* **18,** 1077–1087 (2001).
14. Does Culture Prevent or Drive Human Evolution? « On the Human. at <http://onthehuman.org/2009/12/does-culture-prevent-or-drive-human-evolution/>
15. Fluidigm technology - Nanoflex valves, Integrated Fluidic Circuits. at <http://www.fluidigm.com/technology.html>
16. BioMark™ HD System | Fluidigm Corporation. at <http://www.fluidigm.com/biomark-hd-system.html>
17. ABI PRISM® 7900HT Sequence Detection System User's Manual - Life Technologies. at <http://www.lifetechnologies.com/order/catalog/product/4317596>
18. Drug Metabolism | Drug Toxicology | Affymetrix. at <http://www.affymetrix.com/estore/browse/products.jsp?productId=131412#1_3>
19. Hoeft, T. J. *et al.* Building Partnerships in Community-Based Participatory Research: Budgetary and Other Cost Considerations. *Health Promot. Pract.* (2013). doi:10.1177/1524839913485962

20. West, K. M., Hopkins, S. E., Hopper, K. J., Mohatt, G. V. & Boyer, B. B. Found in translation: Decoding local understandings of genetics and heredity in a Yup'ik Eskimo community. *Public Underst. Sci. Bristol Engl.* **22,** 80–90 (2013).
21. Welcome to the Center for Genomics and Healthcare Equality | CGHE. at <http://depts.washington.edu/cghe/>
22. Havasupai Tribe and the lawsuit settlement aftermath | Genetics. at <http://genetics.ncai.org/case-study/havasupai-Tribe.cfm>
23. Fohner, A. *et al.* Pharmacogenetics in American Indian populations: analysis of CYP2D6, CYP3A4, CYP3A5, and CYP2C9 in the Confederated Salish and Kootenai Tribes. *Pharmacogenet. Genomics* **23,** 403–414 (2013).
24. User Documentation - by Product and Application | Fluidigm. at <http://www.fluidigm.com/product-documents.html>
25. SNP Genotyping Analysis Using TaqMan® Assays. at <http://www.lifetechnologies.com/us/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html>
26. Smith, M. J. *et al.* Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. *Mol. Ecol. Resour.* **11 Suppl 1,** 268–277 (2011).
27. Deo, A. K., Prasad, B., Balogh, L., Lai, Y. & Unadkat, J. D. Interindividual Variability in Hepatic Expression of the Multidrug Resistance-Associated Protein 2 (MRP2/ABCC2): Quantification by Liquid Chromatography/Tandem Mass Spectrometry. *Drug Metab. Dispos.* **40,** 852–855 (2012).
28. Life Science Technologies: Exome Sequencing: Toward an Interpretable Genome. at <http://www.sciencemag.org/site/products/lst_20131011.xhtml>
29. Affymetrix Arrays (GeneChips) - Genomic and RNA Profiling Core - Baylor College of Medicine, Houston, Texas. at <https://www.bcm.edu/research/advanced-technology-core-labs/lab-listing/genomic-profiling/?pmid=2902>
30. Fluidigm Posters. at <http://www.fluidigm.com/posters.html>
31. Guzzi, P. H. *et al.* DMET-Analyzer: automatic analysis of Affymetrix DMET Data. *BMC Bioinformatics* **13,** 258 (2012).
32. Bustamante, C. D., De La Vega, F. M. & Burchard, E. G. Genomics for the world. *Nature* **475,** 163–165 (2011).