Study of DNA Variant Screening Strategies: A Joint Research Group Pilot Study

K. Sol-Church [1 †], D. S. Grove[2], G. J. Wiebe[3], D. Adam[4], R. Welch[5], C. H. Lytle[6], P. S. Adams[7]

[1]AI duPont Hospital For Children, Nemours Children's Clinic (NCC) Wilmington, DE, United States, [2]Penn State University, University Park, PA, United States, [3]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, [4]University of British Columbia, Vancouver, BC, Canada, [5]NCI Core Genotyping Facility, SAIC-Frederick, Gaithersburg, MD, United States, [6]Dartmouth Medical School, Hanover, NH, United States, [7]Trudeau Institute, Saranac Lake, NY, United States.

† to whom correspondence should be directed

ABSTRACT

In recent years, the genotyping core laboratory has been transformed by the increasing number of technologies and platforms suitable for STR and SNP genotyping. This breadth of investigational tools, while facilitating the analysis of DNA variants, has created a method selection dilemma since information is lacking regarding the specific advantages and limitations of even the most routinely used mutation detection techniques.

To address this issue, members of the DNA Sequencing (DSRG), Fragment Analysis (FARG), and Nucleic Acid (NARG) research groups of the ABRF, designed a Genotyping Pilot Study that was validated by RG members. The technologies surveyed in the members' laboratories (N=20) included micro-satellite/STR-PCR analysis (N=14), direct DNA sequencing (N=15), real-time PCR (TaqMan N=7, MGB N=3), SNaPshot (N=4), pyrosequencing (N=2), Sequenom (N=1) and heteroduplex analysis (N=1). To compare the performance and data accuracy of each of the available genotyping methods, identical sets of test samples were genotyped at 10 microsatellite loci and 4 selected SNP loci. A limited number of sub-optimal samples were also included to test each technology's performance. This poster reports on the genotyping of 12 blinded DNA samples and presents performance data obtained from each of the tested technologies and platforms. Evaluation of the various technologies will include an assessment of the quality/concordance of the genotyping calls reported by each participant, and will provide comparative performance evaluation and cost.

RATIONALE OF THE PILOT STUDY

DNA variants exist in the genomes of higher plants and mammals that are responsible for genetic polymorphism. Typical examples include single nucleotide polymorphisms (SNPs), and microsatellite - short tandem repeats (STRs). There is a variety of methods used to genotype SNPs and STRs (DNA sequencing, pyrosequencing, real time PCR, single base extension methods, and heteroduplex analysis). Unfortunately, little information is available regarding the specific advantages and limitations of these mutation detection techniques. For this reason, The Joint RG designed a pilot study whose goals are to compare operational parameters for a sampling of genotyping techniques and platforms available within the membership of the DSRG, FARG and NARG. Towards this end, a set of samples containing known STRs and SNPs were genotyped using STR-PCR analysis, direct DNA sequencing, real-time PCR, SNaPshot, pyrosequencing, Sequenom-Mass Array and WAVE.

STUDY DESIGN

Study Genomic Sample Preparation: Genomic DNA were obtained from Coriell established Cell Lines and amplified by whole genome amplification (WGA) using a kit from GE Healthcare. Sample 7 was subjected to 5 additional rounds of WGA to create a suboptimal sample (ID 8). A set of 4 artificial DNA was generated by mixing varying amounts (10%, 25%, 50% and 75%) of 2 of the unrelated WGA DNAs.

Genotyping Templates Preparation: Ready-made PCR for SNapShot, STR-PCR and sequencing were prepared in bulk and shipped to participants. Pyrosequencing, Real time TaqMan and MGB ECLIPSE (MGB) assays were distributed to participants along with WGA gDNA. Sequenom, and WAVE assays were developed by the participating members.

Data Reporting and Analysis: A web based survey was posted on SurveyMonkey.com, to collect information on the genotyping methods used by the participant. The genotyping data was uploaded by participant to an ftp site and was made accessible for analysis by Joint RG members. In most cases raw data were also uploaded to the ftp site for critical analysis.

RESULTS

Figure 1- Study Design and Genotypes

• In addition to pure DNA, "poor quality" DNA (sample 8) was also included to test participants' skills & instruments' capabilities

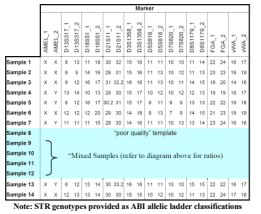
• A subset of "mixed" DNAs (samples 9-12) was generated, to test the capabilities of each method for predicting the correct genotype

Table 1 - SNP Genotyping

SNP	AR	MTHFR	TNF	TYMS
Sample 1	A/G	T	G	C
Sample 2	G	C/T	G	C/T
Sample 3	G	C	G	C/T
Sample 4	G	C/T	G	C
Sample 5	G	C/T	A/G	C/T
Sample 6	G	C/T	G	T
Sample 7	G	C/T	A/G	C/T
Sample 8	G	C/T	A/G	C/T
Sample 9	10A:90G	90C:5T	95A:5G	95C:5T
Sample 10	25A:75 G	87.5C:12.5T	87.5A:12.5G	87.5C:12.5T
Sample 11	50A:50G	75C:25T	75A:25G	75C:25T
Sample 12	75A:25G	62.5C:37.5T	62.5A:37.5G	62.5C:37.5T

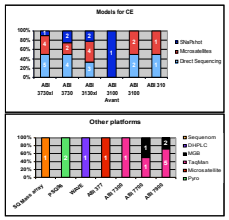
Note: For mixed samples the % alleles present in each SNP combination is as indicated

Table 2 - STR Analysis



Note: STR genotypes provided as ABI allelic ladder classifications

Figure 2 - Overview of genotyping platforms



•Capillary Electrophoresis (CE) are the genotyping platforms of choice for this group of participants.

•A trend for platform with higher throughput capabilities is apparent for both direct sequencing and microsatellite analysis.

•Future studies including the greater scientific community will enable testing of methods and instrumentations under-represented in this pilot study.

Table 3 - Participant's Scores

STR	
Participant	Score %
DC	97
TVM	94.3
BCL	93.4
2088	90
99PMGF	90
UMC7484	90
MKCDADC	89.3
RPCI	85
DK1	84.5
BK2007	82.3
CGF001	82
CUPAS007	80
TGR005	76

TaqMan	
Participant	Score %
BCL	83.33
VT001	83.33
MB001	81.25
7952	75
CGF001	72.92
3080	66.67

MGB ECLIPSE (MGB)	
Participant	Score
BCL	66.67
CGF001	60.42
3080	45.83

Pyrosequencing	
Participant	Score
BCL	92.8
CSS1148	89.3

Direct Sequencing	
Participant	Score %
2088	97.92
BK2007	97.92
3080	93.75
MKCDADC	88.54
01307	87.50
DC	87.50
RPCI	87.50
VT001	83.33
99PMGF	81.25
TGR005	81.25
7952	79.17
MB001	79.17
CUPAS007	54.17
DSN001	54.17
UMC7484	52.08

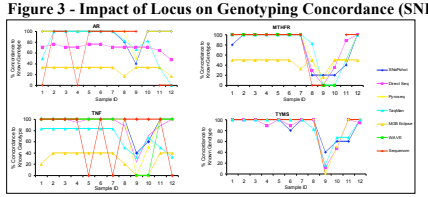
SNaPShot	
Participant	Score %
2088	91.67
99PMGF	91.67
VT001	89.58
CUPAS007	79.17

WAVE	
Participant	Score
VT001	61.3

Sequenom	
Participant	Score
TVM	80

This table displays for each of the genotyping methods tested during this pilot study the participant's unique identifier as well as the % correct genotyping calls that were reported. For STR analysis: % score combines the overall score for genotyping 14 samples with 10 STR markers. For SNP analysis, 2 to 4 locus were genotyped in 12 samples.

Figure 3 - Impact of Locus on Genotyping Concordance (SNP markers)

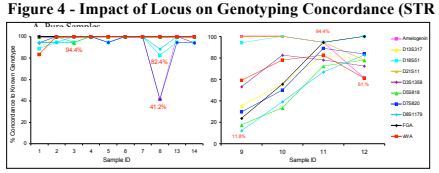


•Sequences surrounding the SNP impacts call accuracy (e.g. Direct sequencing, Sequenom and WAVE are impaired by the presence of short repeats close to the AR SNP).

•Most methods are challenged by samples containing SNPs mixtures (samples 9-12), or samples that show signs of degradation (sample 8).

•MGB ECLIPSE (MGB) data indicates the need for assay optimization and does not reflect method capabilities.

Figure 4 - Impact of Locus on Genotyping Concordance (STR markers)

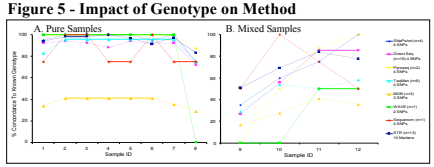


•Good concordance across all loci in pure samples

•Sample 8 locus-specific allelic imbalance or loss of allele is most visible at D7S820 and D18S51 and account for the lower concordance rate at these loci.

•Minor alleles in "tweaked" sample 9 are poorly detected regardless of the locus interrogated.

Figure 5 - Impact of Genotype on Method



•All genotyping methods perform well with optimized assays in pure samples. *Note that MGB and WAVE assays will require further optimization before they are included in the next study.*

•Detection of minor alleles in mixed samples is dependent on the genotyping method used as well as participant's skills. *As reported, Sequenom and STR-PCR generated the best scores to detect low level contaminations in sample 9.*

Figure 6-Cost Analysis

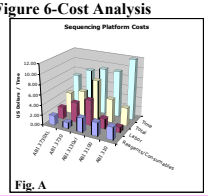


Fig. A

•Reported Cost and Time need to be assessed with caution do to discounts, and variation of laboratory methods.

•Note the labor costs difference between the ABI 3100 and the ABI3130XL (Figure A). Two very similar platforms yet the 3130XL labor costs are more than double the 3100. This may not represent the true situation.

•Averaging the platforms can give a general idea of what to expect in Cost and Time.

•Figure B shows averaged data for the different methods assessed but there is still insufficient data to make any real conclusions.

•Participants reporting; N=1 Sequenome, N=5 Taqman, N=1 WAVE, N=12 Sequencing, N=6 STR, N=4 SNaPshot.

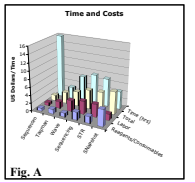


Fig. A

DISCUSSION

The goal of the Joint RG was to design a study to examine available genotyping methodologies in regard to accuracy, cost, time and suitability for high, medium or low throughput. Members of the three groups conferred by phone and set up:

- Procedures to prepare and distribute samples and protocols.
- Surveys to determine how each participant performed their methods.
- Methods of analysis to determine the success of each technology tested.

The results have provided a foundation for setting up a study for laboratories at large. Several issues which will be taken into consideration for the next level of testing will be:

- Suitability of some samples for testing.
- Suitability of some methods for low-throughput genotyping.
- Under-representation of some methods.

ACKNOWLEDGEMENTS

There are many people who made this pilot study possible. For their unique contributions to this project, the Joint RG would like to thank Mike Klein (Utah), Deborah Stabley (NCC-Wilmington, DE), Khashif Haque (NCI-CGF, Gaithersburg, MD), Ashley Price (Penn State). We are grateful to all the members of the DSRG, FARG and NARG for their data contributions and help with the design of the study.