Factors Associated with Success of Custom SNP Assays

R. Ingersoll[1], I.A. McMullen[1], G. Ingersoll[2], E.W. Pugh[1], A.F. Scott[1].

1) Institute of Genetic Medicine, Johns Hopkins Univ, Baltimore, MD; 2) Genetic Software Innovations, Cicero, NY

Abstract

•The Illumina BeadLab™ SNP Genotyping Platform is a high-throughput genotyping system that allows the assessment of up to 1,536 Single Nucleotide Polymorphisms (SNPs) simultaneously on 96 samples. The loci used in these assays may come from many sources, the most common of which is the repository of SNPs in the NCBI database, dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/).

•18,624 assays over 20 months.

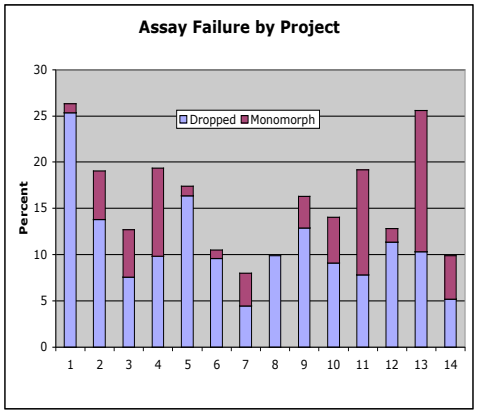
•Project size: 192 to 2,304 samples.

•Custom assay success rate: 75% to 96%.

•Goal: determine factors contributing to assay success.

•Factors tested:
- •Illumina's Design Score
- •Reported average heterozygosity from dbSNP,
- •Validation code from dbSNP.

• Logistic regression was performed to examine the success of SNP loci.

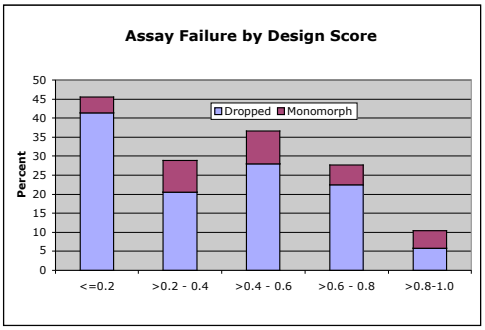


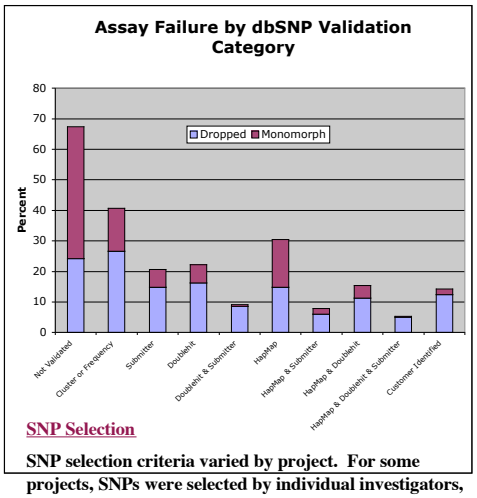
Assay Failure

We considered two different measurements of assay success:

1) The assay failed one or more quality control measures and was not released to the investigator

2) The assay worked, but the SNP was monomorphic (defined here as minor allele frequency <0.5%)





SNP Selection

SNP selection criteria varied by project. For some projects, SNPs were selected by individual investigators, for others SNP Center personnel made the selections. In all cases, design scores supplied by Illumina were taken into consideration. Validation status from dbSNP and reported heterozygosity from dbSNP and/or HapMap were often also used as criteria.

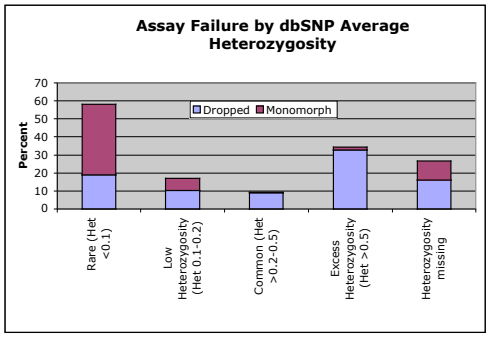
Design Scores are supplied by Illumina and range in value from 0 to 1. Illumina states that SNPs with scores above 0.6 are most likely to work. For this analysis, all SNPs were resubmitted to Illumina using the design algorithm as of May 2005.

Validation Codes can be derived from dbSNP. The criteria for classification include:

- Not validated - No validation information, or MAFof 0 in reported population.
- Frequency - Presence of allele frequency data.
- Cluster - Submitted by multiple labs.
- Double-hit - Each allele has been seen independently in two or more samples.
- Submitter - Submitter validated by two different methods.
- Hapmap - Genotyped as part of the HapMap project.

A single SNP can fall into more than one category, which improves its validation score. Validation scores were derived from build 124 of dbSNP.

Average Heterozygosity is calculated by dbSNP based on reports received with each SNP submission. It does not take variation in ethnicity into consideration, but does consider the size of the population from which the submitted value was generated. The values used for this analysis were also derived from build 124 of dbSNP.



Results

We used logistic regression, adjusting for all three factors simultaneously, considering an assay to have failed if it was not released or if it was monomorphic.

•All three factors are independent predictors of success.

Discussion

Based on these results, it is recommended that multiple factors be considered when selecting SNPs for use on this platform.

Our results suggest that where possible,

•Use a threshhold of 0.8 for design score.
•Use Double hit and/or Submitter validation.
•Avoid non-validated SNPs.
•Customer Validated SNPs worked well (PGA & Celera).
•Avoid SNPs with missing heterozygosity values, or heterozygosity values above 0.5.

Failed vs. Binned Design Score, Binned Validation Code & Binned Reported Heterozygosity

	Odds Ratio	95% CI
Design Scores:		
≤0.2	Reference	Reference
>0.2 - 0.4	2.21	1.32, 3.71
>0.4 - 0.6	1.42	0.95, 2.14
>0.6 - 0.8	1.62	1.12, 2.34
>0.8-1.0	5.34	3.70, 7.73
Validation Level:		
Not Validated	Reference	Reference
Cluster or Frequency	2.36	1.77, 3.16
Submitter	4.47	2.74, 7.29
Doublehit	5.1	4.03, 6.44
Doublehit & Submitter	8.21	5.52, 12.22
HapMap	4.11	3.04, 5.55
HapMap & Submitter	11.29	7.99, 15.95
HapMap & Doublehit	5.95	4.52, 7.84
HapMap & Doublehit & Submitter	12.9	9.67, 17.21
Customer Identified	12.22	9.52, 15.68
Heterozygosity:		
Rare (Het <0.1)	Reference	Reference
Low Heterozygosity (Het 0.1-0.2)	5.79	4.71, 7.12
Common (Het >0.2-0.5)	9.78	8.40, 11.40
Excess Heterozygosity (Het >0.5)	4.49	2.56, 7.88
Heterozygosity missing	4.09	3.29, 5.08