Development of a Low Cost SNP Barcode Panel

B Marosy, JM Romm, KN Hetrick, KF Doheny, EW Pugh, Y Tsai

Center for Inherited Disease Research (CIDR) and Genetic Resources Core Facility (GRCF) SNP Center, IGM, JHUSOM, Baltimore, MD

Introduction:

Advances in SNP genotyping technology have made it possible to produce up to 1 million genotypes for a given sample. These assays are costly; therefore, it is imperative to establish a validation process to pre-test and track samples through the lab and analysis process. To address this, CIDR has developed a SNP barcode panel that will uniquely identify individuals by creating a 'genotype barcode' across all SNP platforms for each individual. This barcode could then be tracked through sample processing to ensure quality (i.e., sample transfers from plate to plate and sample well to sample slide or array) and accuracy of each sample being genotyped (i.e., application of data analysis, data release and data acquisition through dbGAP). This panel will also provide pre-testing data to confirm gender and detect Mendelian inconsistencies, in addition to identifying poor performing samples prior to production genotyping on the Illumina® Golden Gate™ assay.

In order to utilize a SNP barcode, a more cost effective SNP assay is required. Illumina has recently made available a new technology, VeraCode™ BeadXpress™, which allows for low-plex SNP genotyping using the Illumina Golden Gate chemistry. The VeraCode system is flexible to offer either a low cost 96 plex or custom plexing for lower plex mixing allowing the user the freedom to customize their SNPs of interest and alternative platforms (i.e. Allele Specific Primer Extension).

Methods:

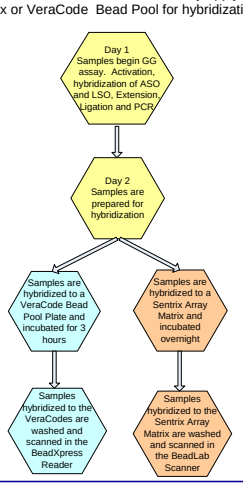
SNP Selection Process

Markers were chosen using the 384 'Ancestry Informative Markers' from an Illumina panel that is included in all their GWA products. We selected a subset of SNPs (Table 1) based on assay characteristics, including: separation between clusters, polymorphic in all three Hapmap populations, distance between SNPs, and allele frequencies that minimized the probability that two samples will have identical genotypes even if the two samples are related (e.g., siblings). The random match probabilities (Evett and Weir 1998; Butler 2005) were calculated in SAS and in PowerStats using the genotypes provided by Illumina on the HapMap samples. Table 2 gives results for unrelated individuals and siblings for 32 of the 384 markers.

Table 2. Random Match Probabilities (no inbreeding) expressed as 1 in…

	CEU	YRI	CHP+JPT
RMP unrelated individuals	457,246,194,673	946,161,806	265,420,527
RMP siblings	1,209,360	36,250	21,793

Figure 1. Workflow of Golden Gate chemistry applying either a Sentrix Array Matrix or VeraCode Bead Pool for hybridization and scanning.



SNP Panel Design Pre-Test

A STRP panel, consisting of 14 autosomal, 3 X and 2 Y loci, has previously been used to pre-test the quality of samples. While the STRP panel is sufficient to detect most Mendelian and/or gender inconsistencies and DNA contamination, it is not a perfect predictor of sample performance and cannot provide a barcode to track sample identity in a SNP assay. In order to pretest the SNP panel design, without inflating false-positive and negative poor performance rates, we utilized previous performance data of 5204 samples from a linkage study. Criteria for failing a test panel included a call rate of 85% for the SNP test panel, and failure to generate a genotype at >40% of loci for the STRP test panel in conjunction with a DNA concentration cut-off of 10ng/uL. This data was then compared to the actual performance (pass or fail) and outcome (released or dropped) of each sample in the SNP linkage study. False positive and negative rates were then calculated based on these comparisons, STRP 0.48%, 3.2% and SNP 0.15% and 2.5%, respectively.

Genotyping Platform

Samples were processed using the Illumina Golden Gate chemistry according to the manufacturers specifications. Sample product was hybridized to either a VeraCode Bead Pool or a Sentrix® Array Matrix (Figure 1). The VeraCode technology uses cylindrical glass beads containing a unique holographic code to identify each bead type allowing for multiplexing of uniquely coded beads. The internal fluidics of the BeadXpress reader apply a combination of fluid flow, gravity and capillary force to align the VeraCodes beads across a groove plate contained within the bottom of each fluid cell. The VeraCode beads are then scanned with a dual laser detection system to identify the holographic code of each bead and measure fluorescence intensity. Data generated from the BeadXpress Reader was analyzed using the Illumina BeadStudio software and call rates and concordance were determined.

Table 1. SNP Barcode Panel

Name	Chr	Pos Mb	Linkage4b	Linkage5	Human Linkage-12	Human CNV370	Human 550	Human 650Y	Illmn 1M	Affy 5.0	Affy 6.0	MAF
rs8627	1	8.34	1	1	1	1	1	1	0	0	0	0.48
rs2007350	1	29.52	1	1	1	1	1	1	1	0	1	0.40
rs1739897	1	75.89	1	0	0	0	0	0	1	1	1	0.45
rs800292	1	194.91	0	0	0	1	1	1	0	1	0	0.40
rs2878979	1	206.46	1	1	1	1	1	1	1	0	0	0.22
rs1991315	2	17.70	1	1	1	1	1	1	0	0	0	0.43
rs1868092	2	46.47	1	1	1	1	1	1	0	1	0	0.31
rs741418	2	75.22	1	1	1	1	1	1	1	0	0	0.34
rs972881	2	107.39	1	1	1	1	1	1	1	1	1	0.22
rs1521527	2	165.14	1	0	0	0	0	0	1	1	0	0.30
rs2054615	2	213.19	1	1	1	1	1	1	0	0	0	0.44
rs2229571	2	215.35	0	0	0	0	0	0	1	0	0	0.46
rs749477	3	10.63	1	1	1	1	1	1	0	0	0	0.49
rs737516	3	43.53	1	1	1	0	0	0	1	1	1	0.43
rs1441443	3	74.01	1	1	1	1	1	1	1	0	0	0.38
rs1479371	3	103.50	1	0	0	0	0	0	1	1	1	0.31
rs3817672	3	197.29	0	0	0	1	1	1	1	0	0	0.31
rs225160	4	52.58	1	1	1	1	1	1	1	1	1	0.40
rs2237051	4	111.12	0	0	0	1	1	1	1	0	0	0.44
rs1880863	4	123.10	1	1	1	1	1	1	1	0	0	0.49
rs4957114	5	1.05	1	1	1	1	1	1	1	0	0	0.47
rs10941312	5	34.04	0	0	0	0	1	1	1	0	0	0.41
rs34609	5	60.52	1	1	1	1	1	1	1	0	0	0.39
rs173686	5	82.85	1	1	1	1	1	1	1	1	1	0.42
rs1864003	5	141.76	1	1	1	1	1	1	1	0	0	0.37
rs767022	6	9.08	1	1	1	1	1	1	1	0	0	0.43
rs722269	6	42.30	1	1	1	1	1	1	1	0	0	0.43
rs1541317	6	118.13	1	1	1	1	1	1	1	0	0	0.41
rs200148	6	143.39	1	1	1	1	1	1	1	0	0	0.50
rs4619	7	45.90	0	0	0	1	1	1	1	1	1	0.31
rs1801131	7	92.89	0	0	0	1	1	1	1	0	1	0.39
rs201492	7	101.54	1	1	1	1	1	1	1	0	0	0.37
rs901170	8	25.83	1	1	1	1	1	1	1	1	1	0.46
rs2725362	8	31.12	0	0	0	1	1	1	1	0	0	0.47
rs718251	8	52.88	1	1	1	1	1	1	1	0	0	0.41
rs17337252	8	53.75	0	0	0	0	1	1	1	1	1	0.48
rs1868280	8	141.97	1	1	1	1	1	1	1	1	1	0.42
rs6474795	9	1.39	1	0	0	0	0	0	0	1	1	0.33
rs2151065	9	16.24	1	1	1	1	1	1	1	1	1	0.46
rs2026999	9	102.18	1	1	1	1	1	1	1	0	0	0.45
rs2230808	9	106.60	0	0	0	1	1	1	1	1	1	0.43
rs7860423	9	140.08	1	1	1	1	1	1	1	1	1	0.46
rs994502	10	26.54	1	1	1	1	1	1	1	0	0	0.37
rs909349	12	3.36	1	1	1	1	1	1	1	0	0	0.40
rs2051713	12	89.50	1	1	1	1	1	1	1	0	0	0.43
rs903770	12	115.78	1	1	1	1	1	1	1	1	1	0.47
rs310935	13	61.93	1	1	1	1	1	1	1	0	1	0.35
rs17655	13	102.33	0	0	0	0	0	0	0	0	0	0.39
rs1969888	13	105.65	1	1	1	1	1	1	1	0	0	0.41
rs1760904	14	19.92	0	0	0	1	1	1	1	0	1	0.34
rs311848	14	58.27	1	1	1	1	1	1	1	0	0	0.45
rs760335	14	93.88	1	1	1	1	1	1	1	0	0	0.36
rs2873	15	29.02	1	1	1	1	1	1	1	1	1	0.40
rs352476	15	63.59	1	1	1	1	1	1	1	0	0	0.36
rs461785	16	64.37	1	1	1	1	1	1	1	1	1	0.48
rs717571	17	72.13	1	1	1	1	1	1	1	0	0	0.35
rs168206	18	3.68	1	0	1	1	1	1	1	0	0	0.37
rs917711	18	52.01	1	1	1	1	1	1	1	0	0	0.46
rs718066	19	45.01	1	1	0	1	1	1	1	0	0	0.46
rs354731	20	58.38	1	1	1	1	1	1	1	1	1	0.48
rs915774	21	44.40	1	1	1	1	1	1	1	0	0	0.43
rs756658	22	17.84	1	1	1	1	1	1	1	1	1	0.31
rs739096	22	35.07	1	0	0	0	0	0	0	1	1	0.38
rs1405446	X	12.03	0	0	0	1	1	1	1	0	0	0.32
rs11095017	X	25.84	0	0	0	1	1	1	1	0	0	0.41
rs1454190	X	28.50	0	0	0	0	0	1	1	0	0	0.27
rs10856562	X	86.30	0	0	0	1	1	1	1	0	1	0.50
rs525869	X	90.48	1	1	1	1	1	1	1	1	1	0.50
rs12006784	X	115.17	0	0	0	0	0	1	1	0	0	0.33
rs2040962	X	116.49	1	1	1	1	1	1	1	0	0	0.35
rs12841685	X	119.84	0	0	0	1	1	1	1	0	0	0.42
rs1883906	X	126.52	1	1	1	1	1	1	1	0	0	0.50
rs530501	X	148.38	1	1	1	1	1	1	1	0	0	0.37
rs2058276	Y	2.73	1	1	1	0	1	1	1	0	0	0.33
rs1865680	Y	6.93	1	1	1	0	1	1	1	0	0	0.37
rs2032624	Y	13.54	0	0	0	0	1	1	1	0	1	0.37
rs3848982	Y	20.18	0	0	0	0	1	1	1	0	1	0.38
rs3900	Y	20.19	0	0	0	0	0	0	0	0	1	0.46
rs2032631	Y	20.33	0	0	0	0	0	0	0	0	1	0.37
			57	52	50	62	67	69	71	23	34	

Results:

Comparison of Platforms using the SNP Barcode Panel

To validate and compare the VeraCode assay to the standard GoldenGate processing, 2 sets of samples, 93 production samples and 96 CIDR controls, were processed across the GoldenGate chemistry and hybridized to both a VeraCode Bead Pool and a Sentrix Array Matrix. Genotypes generated were calculated to be 99.988% and 100% concordant, respectively.

Validation of Barcode Test Panel Predictive Performance

To determine how well the barcode test panel would predict performance in a higher-plex Golden Gate SNP assay, a set of 248 samples with known performance (good, bad, gender discrepancy and contaminated) were genotyped for the SNP barcode test panel (Table 3).

Table 3. Validation of SNP Test Panel With known Performance Samples

Sample Type	# of Samples Tested	Outcome	False Pos/Neg Rates
Good Performing CIDR Controls	182	0 Failed Test Panel	False Pos Rate = 1.11%
Good Performing Production Samples	58	3 Failed Test Panel	
Poor Performing Production Samples	30	0 Passed Test Panel	False Neg Rate = 0%
CIDR Controls – Gender	182	All Genders were correctly identified	
Known Gender Discrepant Production Samples	5	All 5 were caught by test panel	
Known Contaminated Production Samples	3	All 3 Failed Test Panel	

An additional set of 95 production samples, which failed the CIDR STRP test panel, were also genotyped on the SNP barcode test panel for comparison (Table 4).

Table 4. Comparison of Failed STRP Samples to SNP Barcode Test Panel

	Failed SNP TP	Passed SNP TP	Total
Failed STRP Loci	43	1	44
Passed STRP Loci	3	48	51
Total	48	49	95

A set of 95 production samples were processed in parallel on the SNP barcode test panel, the Illumina Human1M BeadChip and the Affymetrix Genome-wide Human SNP Array 6.0. Concordance between the 1M Chip and SNP Barcode was calculated at 99.919%. Heterozygous-Homozygous concordance determined between the Affymetrix 6.0 SNP array and the SNP barcode test panel was 99.925%. Gender was confirmed in all samples except three. The three gender discrepancies were also reproduced in the 1M chip data and the Affymetrix 6.0 SNP array data. Mendelian inconsistancies were identified in 5 families and were confirmed in the 1M and 6.0 SNP data. An unexpected duplicate was also identified in the sample set and confirmed in the 1M data.

Discussion:

CIDR has developed a SNP barcode panel that will uniquely identify individuals by creating a 'genotype barcode' for each individual and allow for sample tracking through all aspects of high throughput genotyping. The VeraCode platform was validated for genotyping the SNP Barcode panel as it provides a low cost alternative for either a low-plex or standard 96 plex assay. Multiple samples of varying performance have been tested using the panel to confirm the predictive performance the panel may provide as well as confirmation of pedigree and gender. CIDR is beginning to implement the SNP barcode panel in production. Additional evaluation will include the ability to predict performance in GWA studies.