

TrueAllele[®] Validation on PowerPlex[®] 16 HS Mixture Data

Jennifer Hornyak, MS, Martin Bowkley, MS
and Mark W. Perlin, PhD, MD, PhD

Cybergenetics, Pittsburgh, PA

July 15, 2014

Cybergenetics © 2014



Corresponding Author:

Mark W. Perlin, PhD, MD, PhD
Chief Scientific Officer
Cybergenetics
160 North Craig Street, Suite 210
Pittsburgh, PA 15213 USA
(412) 683-3004
(412) 683-3005 FAX
perlin@cybgen.com

Acquiring data

The data in the form of .fsa files were received via email from the Cuyahoga County Medical Examiner's Office in June 2014. The mixture proportions for each sample and known references were also noted in the file names and emails, respectively.

The evidence items were processed using an ABI 3130 sequencer and the PowerPlex® 16 HS STR panel.

Items

A total of 17 laboratory synthesized mixtures comprising 2 and 3 contributors were received. There were 7 mixture items for the two contributor set with mixture ratios of 1:10, 1:5, 1:3, 1:1, 3:1, 5:1, and 10:1. A total of 10 mixture items made up the three contributor set with mixture ratios of 0.5:0.5:1, 0.5:2:0.5, 0.5:2:1, 1:0.5:2, 1:1:1, 1:1:2, 1:2:1, 2:0.5:0.5, 2:1:0.5, and 2:1:1. These mixtures were synthesized from 3 known reference samples (Table 1).

Cybergenetics processing

The .fsa files sent to Cybergenetics were processed through the TrueAllele Casework Analyze module. The quality-checked peaks were uploaded to a database.

The data was downloaded from the database and requests were created assuming either 2 or 3 contributors, depending on the data set. These requests were processed at a burn-in/read-out of 100K/100K. All results were run in duplicate, and additional replicates were run as needed.

Collecting results

Data was collected, and the contributor genotype matching to each known reference was chosen based on match statistic, KL, and mixture weight.

Match statistics were calculated using a co-ancestry coefficient of 1% and the Promega Corporation African American, Caucasian, and Hispanic populations. The lowest match among populations was used.

A total of 44 genotype comparisons were made from the 17 items from the different mixture sets.

Sensitivity

Sensitivity statistics were calculated using the average match statistic between the two replicate computer runs. The count, minimum, average, median, standard deviation, and maximum match statistic values for each contributor group were calculated (Table 2a). In addition, the match statistics were binned by log(LR) value and plotted in a

frequency distribution (Figure 1). The number of false exclusions was also recorded (Table 2b). The average match information for the two contributor set was around ten trillion, with no false exclusions. The average match information for the three contributor set was around a million, with 4 false exclusions.

Specificity

Specificity statistics were calculated by comparing the inferred genotype for each calculated match from the first replicate against 10,000 randomly generated profiles from a population. The statistics and counts for the non-matching log(LR) values were recorded (Table 3a). The Promega Corporation African American, Caucasian, and Hispanic populations were used for a total of 30,000 comparisons for each evidence genotype. Frequency histograms were also produced (Figure 2). The total number of comparisons for the two contributor group was 420,000, while there were total of 900,000 comparisons for the three contributor group. The average non-match information for two contributors was around 1/septillion and around 1/quintillion for three contributors. There were few false positives (Table 3b).

Reproducibility

A comparison was made between the log(LR) values obtained between replicate runs on the same data (Figure 3). The within-group standard deviations were calculated and recorded (Table 4). For two contributors, the runs differed by a factor of 1.6 (within group std dev = 0.2149). For three contributors, the runs differed by a factor of 2.1 (within group std dev = 0.3211).

Table 1: Design. Information regarding the construction of the mixture samples for the two and three contributor groups. Each mixture sample was created using different individuals (Contributor 1, 2, or 3) at different mixing proportions (Ratio).

Sample	Ratio	ncon	Contributor 1	Contributor 2	Contributor 3
1.0-10.0	1:10	2	IC37	IC41	
1.0-5.0	1:5	2	IC37	IC41	
1.0-3.0	1:3	2	IC37	IC41	
1.0-1.0	1:1	2	IC37	IC41	
3.0-1.0	3:1	2	IC37	IC41	
5.0-1.0	5:1	2	IC37	IC41	
10.0-1.0	10:1	2	IC37	IC41	
0.5-0.5-1.0	0.5:0.5:1	3	IC16	IC37	IC41
0.5-2.0-0.5	0.5:2:0.5	3	IC16	IC37	IC41
0.5-2.0-1.0	0.5:2:1	3	IC16	IC37	IC41
1.0-0.5-2.0	1:0.5:2	3	IC16	IC37	IC41
1.0-1.0-1.0	1:1:1	3	IC16	IC37	IC41
1.0-1.0-2.0	1:1:2	3	IC16	IC37	IC41
1.0-2.0-1.0	1:2:1	3	IC16	IC37	IC41
2.0-0.5-0.5	2:0.5:0.5	3	IC16	IC37	IC41
2.0-1.0-0.5	2:1:0.5	3	IC16	IC37	IC41
2.0-1.0-1.0	2:1:1	3	IC16	IC37	IC41

Table 2: Sensitivity. Statistics were calculated for 2 and 3 contributors. Table (a) shows the number of comparisons as well as the log(LR) minimum, mean, median, standard deviation, and maximum values. Table (b) shows the number of false exclusions occurring in each log(LR) bin such that “0” indicates the interval [0,1).

(a) Summary statistics

ncon	2	3
N=	14	30
min	5.921	-10.676
mean	13.363	6.421
median	14.807	6.525
std dev	5.045	6.812
max	18.415	17.778

(b) False exclusions

ncon	2	3
-1	0	0
-2	0	1
-3	0	1
-4	0	0
-5	0	0
-6	0	0
-7	0	0
-8	0	0
-9	0	1
-10	0	0
-11	0	1
<i>Total</i>	<i>0</i>	<i>4</i>

Table 3: Specificity. Statistics were calculated for 2 and 3 contributors across all three Promega Corporation ethnic populations. Table (a) shows the number of comparisons along with the log(LR), minimum, mean, median, maximum, standard deviation, mu, and sigma values. Table (b) shows the number of false inclusions occurring in each log(LR) bin such that “0” indicates the interval [0,1).

(a) Summary statistics

ncon	2			3		
	BLK	CAU	HIS	BLK	CAU	HIS
ethnicity						
N =	140000	140000	140000	300000	300000	300000
minimum	-30.0000	-30.0000	-30.0000	-30.0000	-30.0000	-30.0000
mean	-24.2726	-23.8234	-24.0119	-18.9710	-17.1114	-17.5831
median	-25.0459	-24.4592	-24.6789	-19.0058	-16.9681	-17.4613
maximum	-0.0671	1.3220	-1.7138	2.1160	2.7145	3.6840
std dev	4.6729	4.7913	4.7755	5.5325	5.8872	5.6997
mu	-30.0000	-30.0000	-30.0000	-19.5510	-17.4503	-17.9220
sigma	7.3918	7.8171	7.6591	6.0831	6.2471	6.0576

(b) False inclusions

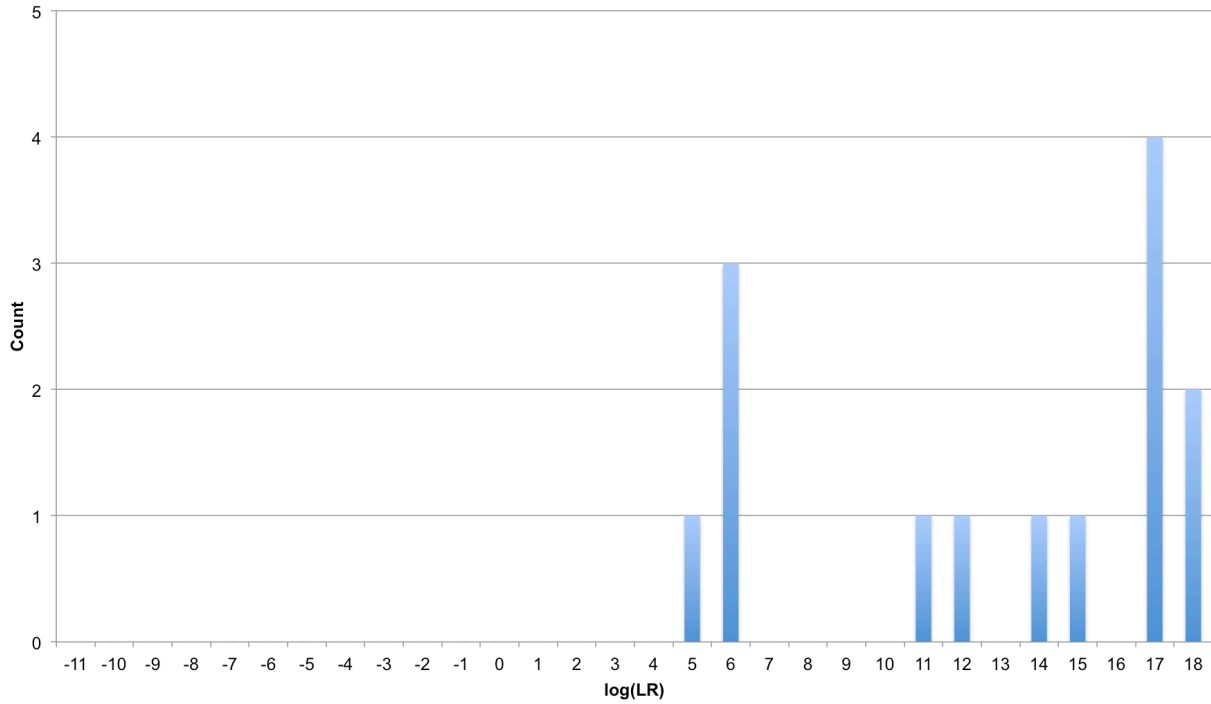
ncon	2			3		
	BLK	CAU	HIS	BLK	CAU	HIS
0	0	0	0	6	28	17
1	0	1	0	2	7	9
2	0	0	0	1	1	3
3	0	0	0	0	0	3
<i>Total</i>	<i>0</i>	<i>1</i>	<i>0</i>	<i>9</i>	<i>36</i>	<i>32</i>

Table 4: Reproducibility. The mean (μ), standard deviation (σ) and within-group standard deviation (σ_w) measure of reproducibility are shown for both two and three contributors.

ncon	2	3
μ	13.363	6.421
σ	4.956	6.762
σ_w	0.215	0.321

Figure 1: Sensitivity. Histograms show the log(LR) genotype match distribution for (a) 2 contributor mixtures and (b) 3 contributors mixtures.

(a) 2 contributors



(b) 3 contributors

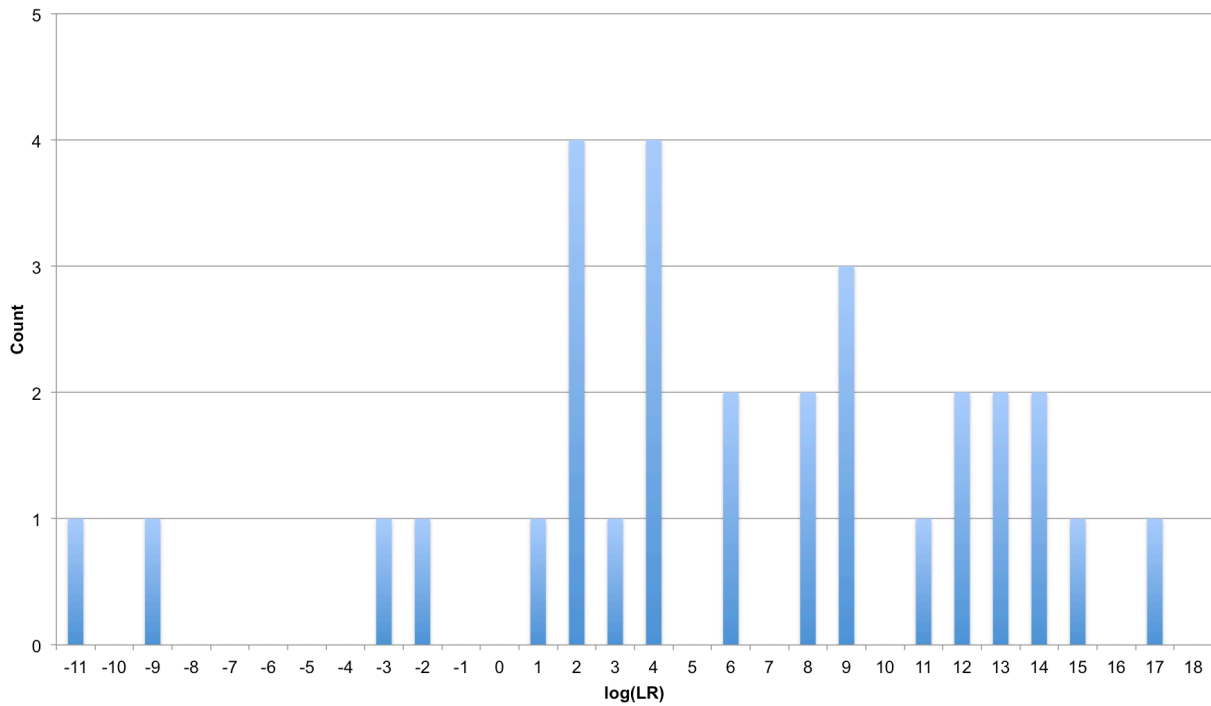
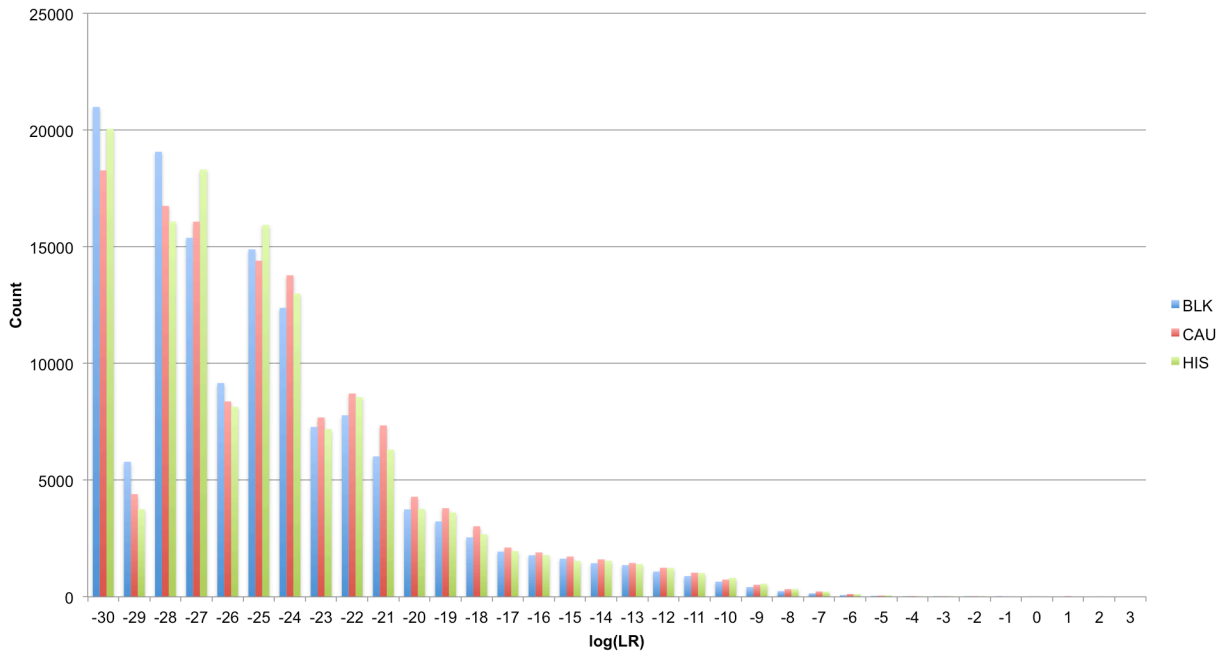


Figure 2: Specificity. Histograms show the log(LR) genotype match distribution for (a) 2 contributor mixtures and (b) 3 contributor mixtures, relative to ten thousand randomly generated profiles. Each ethnic population is depicted in a different color.

(a) 2 contributors



(b) 3 contributors

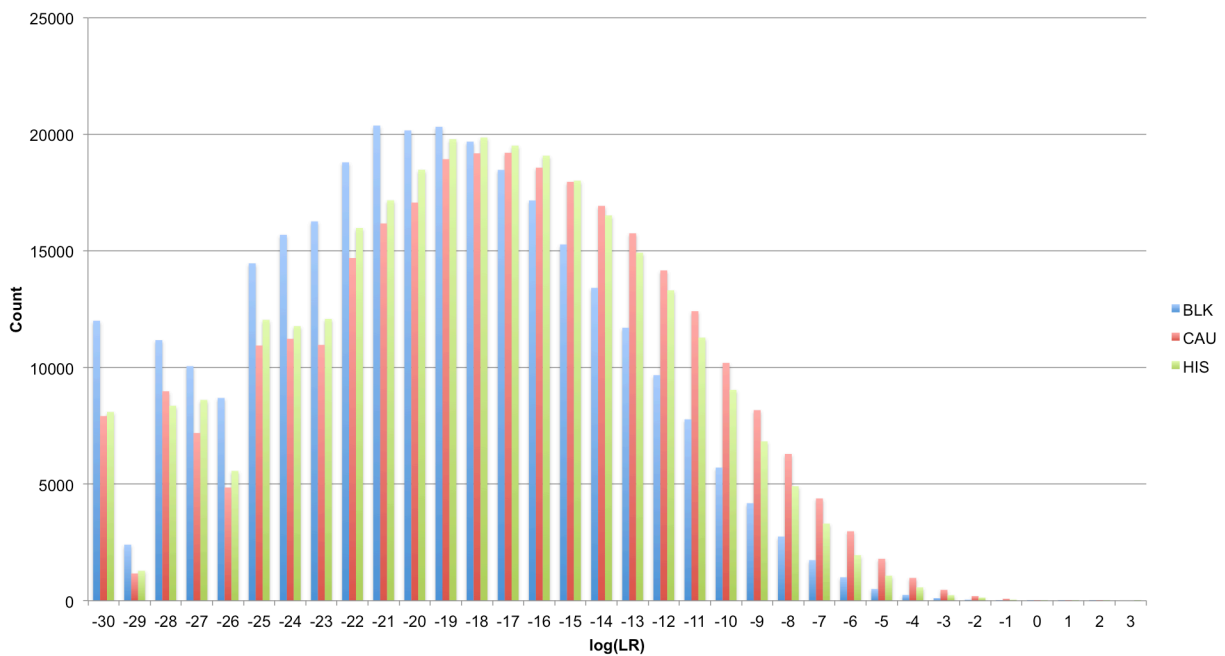
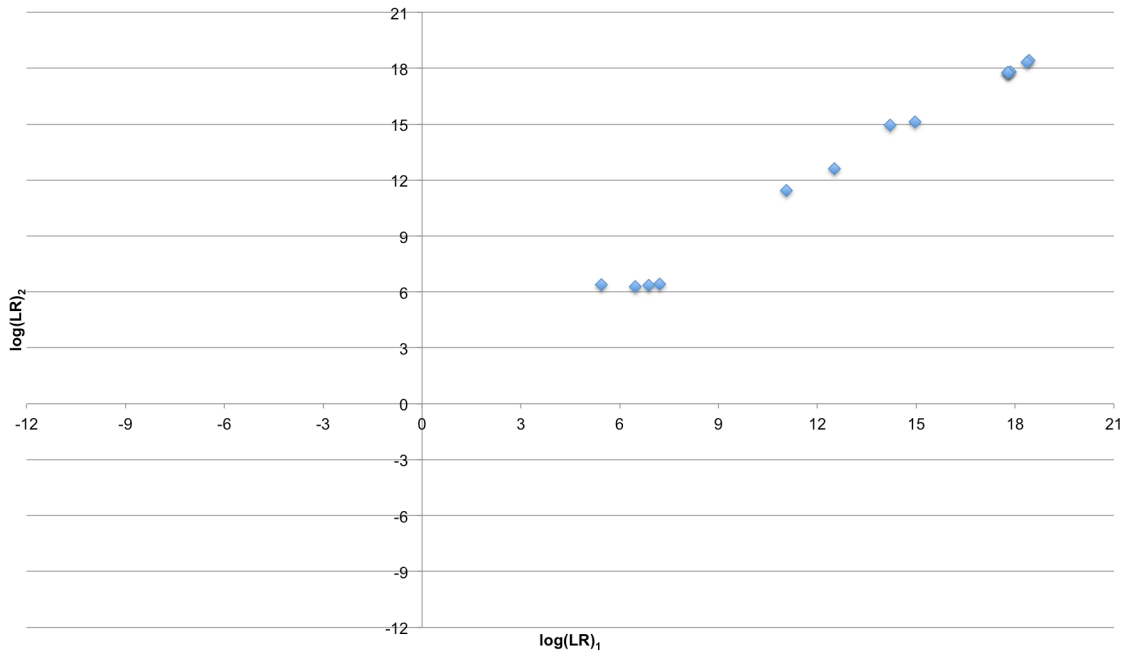


Figure 3: Reproducibility. The scatterplots show log(LR) genotype match values for duplicate computer runs on the same evidence for (a) 2 contributor mixtures and (b) 3 contributor mixtures. Each point depicts the two match values on the first (x) and second (y) run.

(a) 2 contributors



(b) 3 contributors

