# TrueAllele® Validation
# on Minifiler™ Mixture Data

Jennifer Hornyak, MS, William P. Allan, MS
and Mark W. Perlin, PhD, MD, PhD


Cybergenetics, Pittsburgh, PA

July 15, 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 sFTP from Sorenson Forensics in July 2014.  Validation documentation was also received which outlined the mixture study design as well as the known reference samples.

The evidence items were processed using an ABI 3130xl sequencer and the Minifiler™ STR panel.

### Items

A total of 42 laboratory synthesized mixtures comprising 2 contributors at both 5 and 10 second injection times were analyzed.  There were three mixture groups (MMTK, MRRP, and RHSW), where the first two letters correspond with the first reference contributing and the second two letters correspond with the second reference.  There were 15 mixture items injected at 5 seconds and 27 mixture items injected at 10 seconds.  For each mixture group, the contributor ratios were 1:10, 1:3, 1:1, 3:1, and 10:1.  The mixtures were synthesized from 6 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 2 unknown contributors.  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 United States FBI African American, Caucasian, and Hispanic populations.  The lowest match among populations was used.

A total of 84 genotype comparisons were made from the 42 items from the different mixture groups and injection times.

### 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

2).  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 reviewed.  The average match information for the two contributors at both 5 and 10 second injection times was around a hundred million, with no 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 United States FBI African American, Caucasian, and Hispanic populations were used for a total of 30,000 comparisons for each evidence genotype.  Frequency histrograms were also produced (Figure 2).  The total number of comparisons for the 5 second injection time was 900,000, while there were total of 1,620,000 comparisons for the 10 second injection time.  The average non-match information for the 5 second injection was around 1/trillion and around 1/quintillion for 10 seconds.  The overall false positive rate was 0.0315% (Table 3b).  The false positive rate for those false matches greater than 2 ban was 0.00198%.

### *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 both the 5 second and 10 second injection times (two contributors), the runs differed by a factor of 1.4 (5 sec $\sigma_w$ = 0.132; 10s $\sigma_w$ = 0.150).

**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 or 2) at different mixing proportions (Ratio).

| Sample | Ratio | ncon | Contributor 1 | Contributor 2 |
|---|---|---|---|---|
| MMTK_10to1 | 10:1 | 2 | MM | TK |
| MMTK_3to1 | 3:1 | 2 | MM | TK |
| MMTK_1to1 | 1:1 | 2 | MM | TK |
| MMTK_1to3 | 1:3 | 2 | MM | TK |
| MMTK_1to10 | 1:10 | 2 | MM | TK |
| MRRP_10to1 | 10:1 | 2 | MR | RP |
| MRRP_3to1 | 3:1 | 2 | MR | RP |
| MRRP_1to1 | 1:1 | 2 | MR | RP |
| MRRP_1to3 | 1:3 | 2 | MR | RP |
| MRRP_1to10 | 1:10 | 2 | MR | RP |
| RHSW_10to1 | 10:1 | 2 | RH | SW |
| RHSW_3to1 | 3:1 | 2 | RH | SW |
| RHSW_1to1 | 1:1 | 2 | RH | SW |
| RHSW_1to3 | 1:3 | 2 | RH | SW |
| RHSW_1to10 | 1:10 | 2 | RH | SW |

**Table 2: Sensitivity.** Statistics were calculated for both 5 and 10 second injections for two contributor mixtures. The table shows the number of comparisons as well as the log(LR) minimum, mean, median, standard deviation, and maximum values. No false exclusions were observed.

| ncon | 2 | |
| --- | --- | --- |
| **injection** | **5 sec** | **10 sec** |
| N= | 30 | 54 |
| min | 0.055 | 0.860 |
| mean | 8.143 | 8.105 |
| median | 9.432 | 9.374 |
| std dev | 2.763 | 2.686 |
| max | 11.421 | 11.420 |

**Table 3: Specificity.**  Statistics were calculated for two contributors both 5 and 10 second injection times across all three United States FBI 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 | | | | | |
|---|---|---|---|---|---|---|
| injection | 5 sec | | | 10 sec | | |
| ethnicity | BLK | CAU | HIS | BLK | CAU | HIS |
| N = | 300000 | 300000 | 300000 | 540000 | 540000 | 540000 |
| minimum | -16.0000 | -16.0000 | -16.0000 | -16.0000 | -16.0000 | -16.0000 |
| mean | -13.1116 | -12.4888 | -12.6432 | -12.8082 | -12.1632 | -12.3496 |
| median | -13.2555 | -12.9459 | -13.1390 | -13.1702 | -12.7540 | -12.8926 |
| maximum | 3.4328 | 3.0907 | 3.4953 | 4.3563 | 3.4447 | 4.6143 |
| std dev | 2.8080 | 3.0728 | 3.0465 | 3.0139 | 3.2726 | 3.2437 |
| positive | 19 | 48 | 55 | 107 | 279 | 287 |
| mu | -16.0000 | -16.0000 | -16.0000 | -16.0000 | -16.0000 | -16.0000 |
| sigma | 4.0284 | 4.6660 | 4.5331 | 4.3898 | 5.0429 | 4.8833 |

(b) False inclusions

| ncon | 2 | | | | | |
|---|---|---|---|---|---|---|
| injection | 5 sec | | | 10 sec | | |
| ethnicity | BLK | CAU | HIS | BLK | CAU | HIS |
| 0 | 15 | 35 | 36 | 77 | 187 | 201 |
| 1 | 3 | 10 | 17 | 21 | 70 | 73 |
| 2 | 0 | 2 | 1 | 5 | 12 | 7 |
| 3 | 1 | 1 | 1 | 2 | 10 | 4 |
| 4 | 0 | 0 | 0 | 2 | 0 | 2 |
| *Total* | *19* | *48* | *55* | *107* | *279* | *287* |

**Table 4: Reproducibility.** The mean ($\mu$), standard deviation ($\sigma$) and within-group standard deviation ($\sigma_w$) measure of reproducibility are shown for two contributors for both the 5 second and 10 second injection time.

| ncon | 2 | |
|---|---|---|
| injection | 5 sec | 10 sec |
| $\mu$ | 8.143 | 8.105 |
| $\sigma$ | 2.742 | 2.678 |
| $\sigma_w$ | 0.132 | 0.150 |

**Figure 1: Sensitivity.** Histograms show the log(LR) genotype match distribution for (a) two contributor, 5 second injection and (b) two contributor, 10 second injection.

(a) 2 contributors, 5 seconds
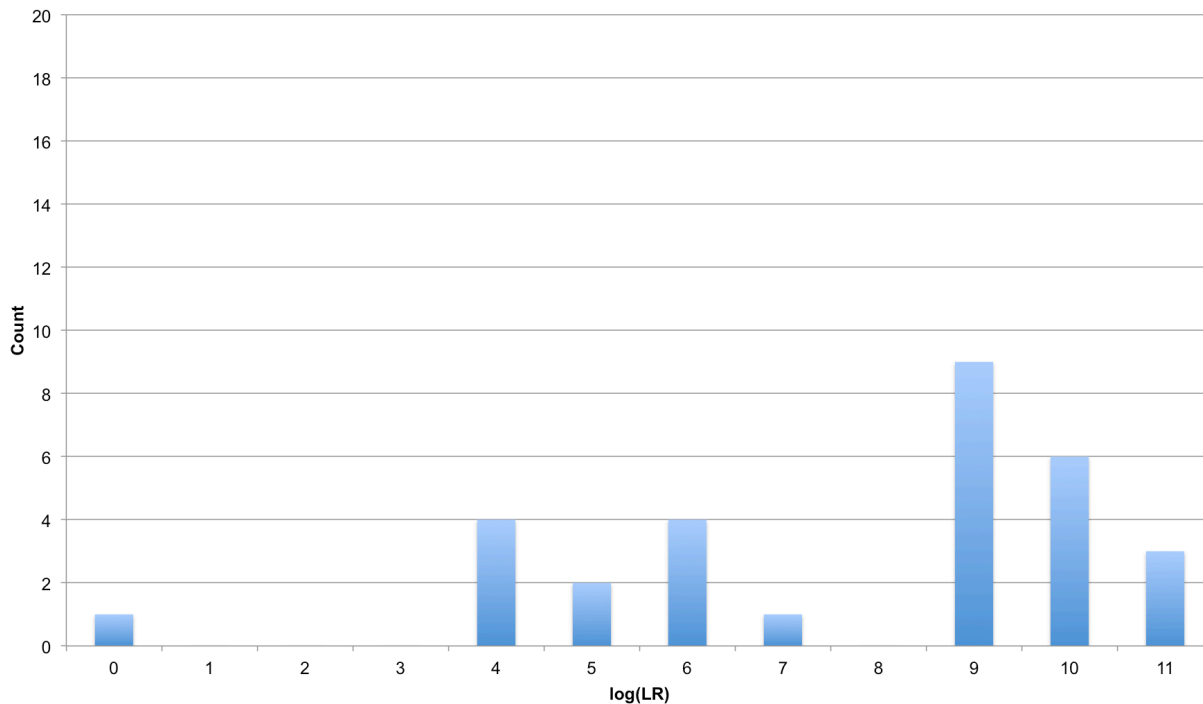

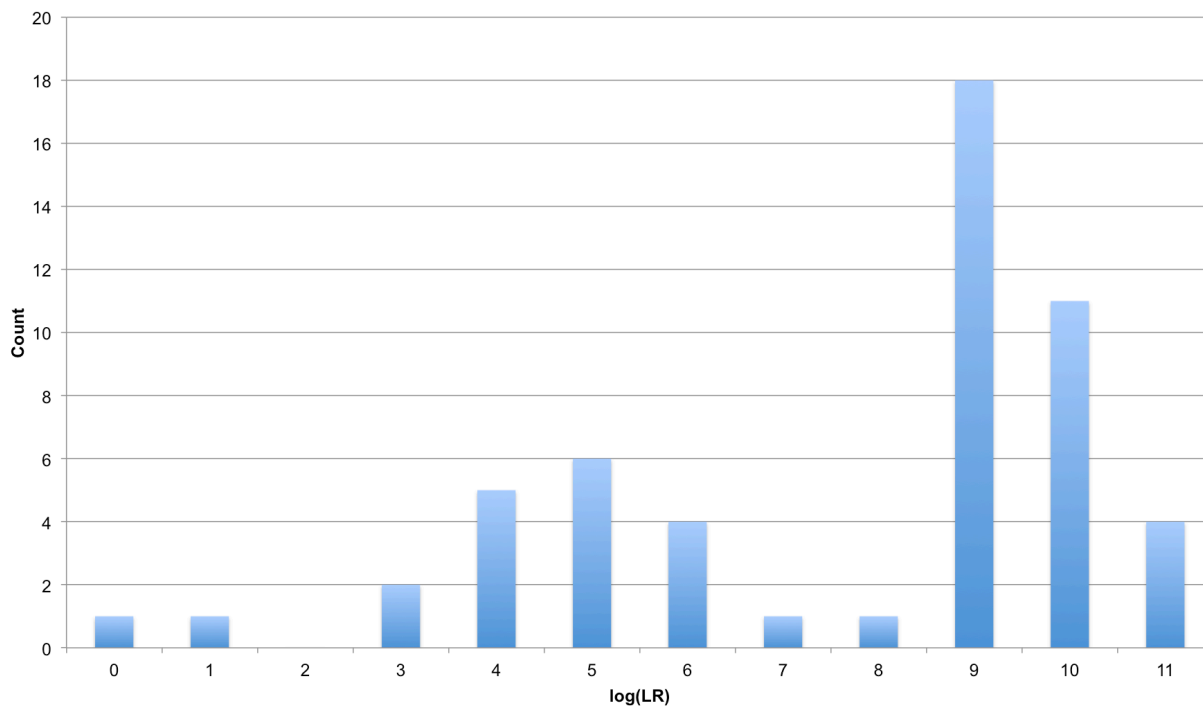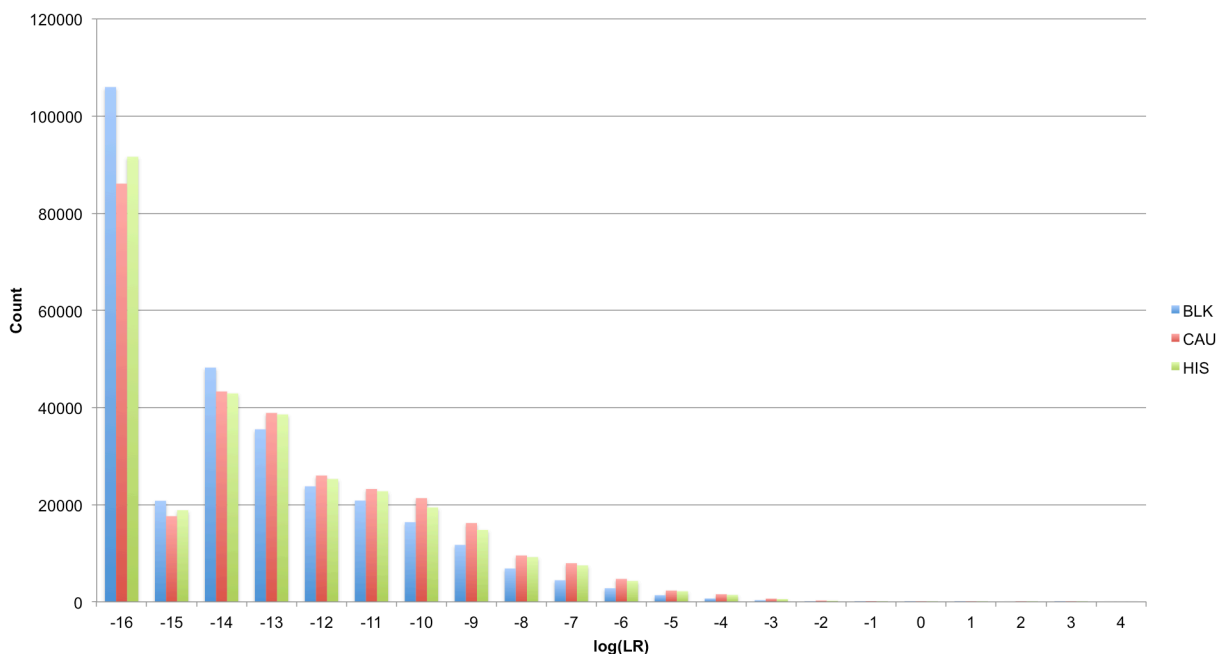
(b) 2 contributors, 10 seconds

**Figure 2: Specificity.** Histograms show the log(LR) genotype match distribution for (a) two contributors, 5 seconds and (b) two contributors, 10 seconds relative to ten thousand randomly generated profiles. Each ethnic population is depicted in a different color.

(a) 2 contributors, 5 seconds
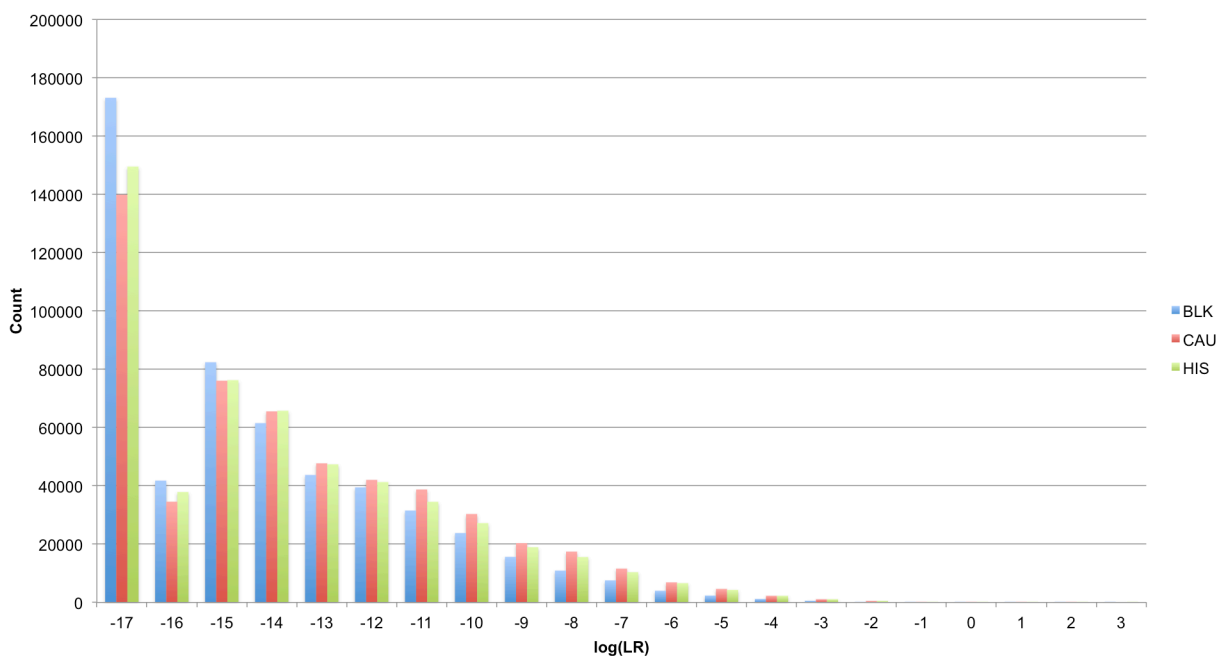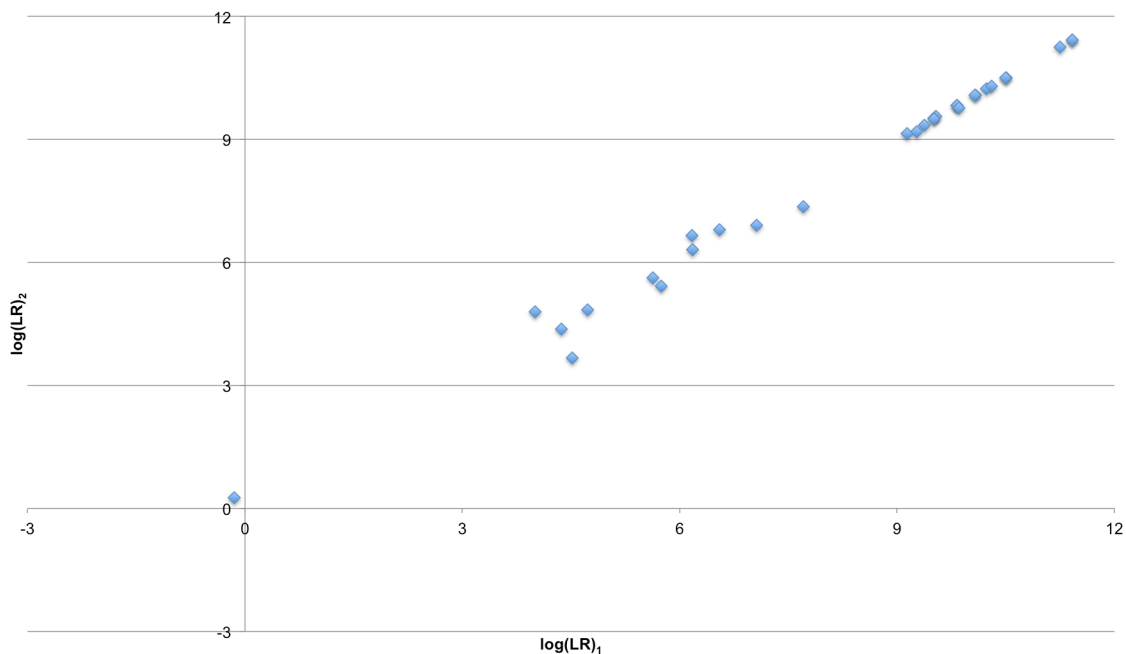


(b) 2 contributors, 10 seconds

**Figure 3: Reproducibility.** The scatterplots show log(LR) genotype match values for duplicate computer runs on the same evidence for the (a) two contributors, 5 second injection and (b) two contributors, 10 second injection. Each point depicts the two match values on the first (x) and second (y) run.

(a) 2 contributors, 5 seconds



(b) 2 contributors, 10 seconds