Evaluating the Specificity of Genotypic Inference with TrueAllele® Casework Software

Jay A. Caponera MS, F-ABC

New York State Police Forensic Investigation Center, Albany, NY

1. ABSTRACT

Interpretation of low template and complex mixed DNA profiles with the binary inclusion/exclusion approach often reduces or precludes statistical weight from being applied to probative evidence items. Quantitative data modeling of DNA data offers an alternative strategy that can result in more informative profiles. This study uses probabilistic genotyping software to objectively infer individual genotypes from both low template and mixed samples with up to four contributors. An approximate limit of detection with the software was observed using DNA inputs of 15.6pg for single source samples, and maximum separation between known donor and non-donor genotypes was achieved with as little as 62.5pg. Average computer-inferred genotype specificity between donor and non-donor profiles was over 13 log units for two person mixtures, 5 log units for three person mixtures, and 4 log units for four person mixtures. Results from this study show that probabilistic genotyping match statistics were both reproducible and specific to all known donor profiles.

2. INTRODUCTION

The forensic literature has increasingly made recommendations for the use of probabilistic genotyping[1], including most recently a strong encouragement from the DNA Commission of the International Society of Forensic Genetics (ISFG) to adopt likelihood ratio-based approaches that include drop-in and drop-out for solving mixed template samples.[2]

TrueAllele Casework (Cybergenetics) is a fully continuous Bayesian method that uses an iterative Markov chain Monte Carlo (MCMC) method to infer genotypes from evidentiary profiles and compute DNA match statistics, and can easily accommodate drop-in and drop-out.[3]

By preserving more identification information, the computer is also able to add increased specificity to genotypic inference, ultimately resulting in a high degree of separation between known donor and non-donor likelihood ratios. The high genotype specificity observed with this approach can then be translated into simplified DNA match reporting based on likelihood ratio calculations.

3. METHODS

• **Single Source**: Two sets of serially diluted single donor samples were amplified with DNA input concentrations ranging from 3.9 to 500pg. Known donor log(LR) values were then compared against 19 non-donor log(LR) values.

• **Two Person Mixtures**: Two separate sets of two person mixtures were amplified in ratios of 1:1, 1:2, 1:5, 1:9, and 1:19. Known donor log(LR) values were then compared against 18 non-donor log(LR) values.

• **Three Person Mixtures**: Two separate sets of three person mixtures were amplified in ratios of 1:1:1, 1:2:1, 1:5:1, 1:10:1, 1:2:3, 2:2:1. Known donor log(LR) values were then compared against 17 non-donor log(LR) values.

• **Four Person Mixtures**: Two separate sets of four person mixtures were amplified in ratios of 2.5:2:1.5:1, 3.5:3:1.5:1, 5:3:2:1, and 7:4.5:2.5:1. Known donor log(LR) values were then compared against 16 non-donor log(LR) values.

• Buccal swab samples from unrelated donors were extracted on an EZ1® Advanced XL using QIAGEN Investigator chemistry, quantified with Quantifiler® Duo, amplified with Identifiler® Plus (target input of 1ng for all mixtures), and run on a 3130xl using 1μl amplicon input with 10 second injection times.

• All samples were solved in duplicate with TrueAllele Casework using MCMC cycle times ranging from 25K/25K (burn in/read out) to 50K/50K with a theta value of .01. All likelihood ratio match statistics are shown in log(10) form.
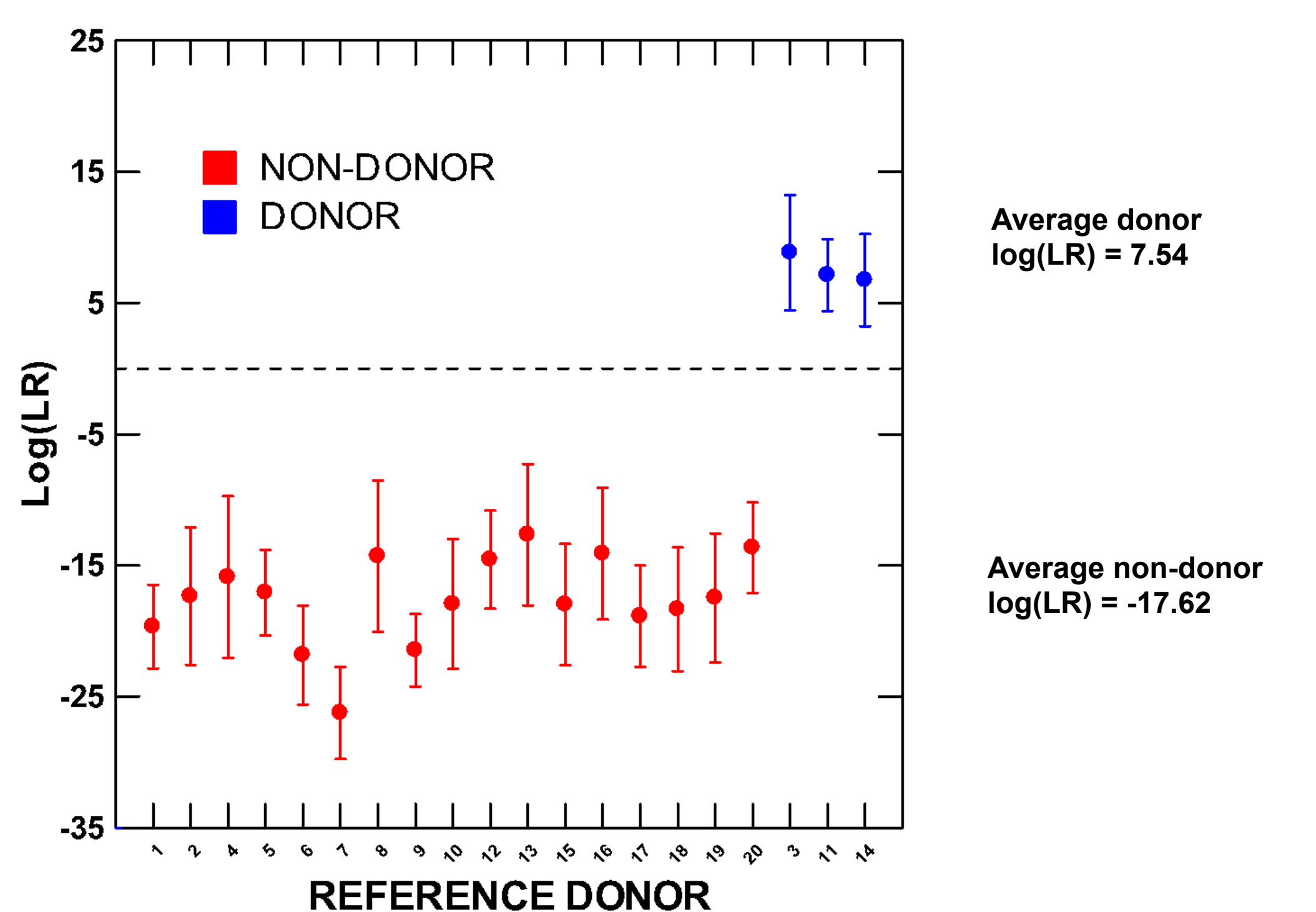
4. RESULTS

Dot plot showing single donor specificity as a function of DNA input (pg). Reference samples include one known donor and 19 non-donors from each sensitivity set. Mean values from all replicated single unknown requests are pooled (n = 32). Error bars represent one standard deviation; dashed line is set at zero.
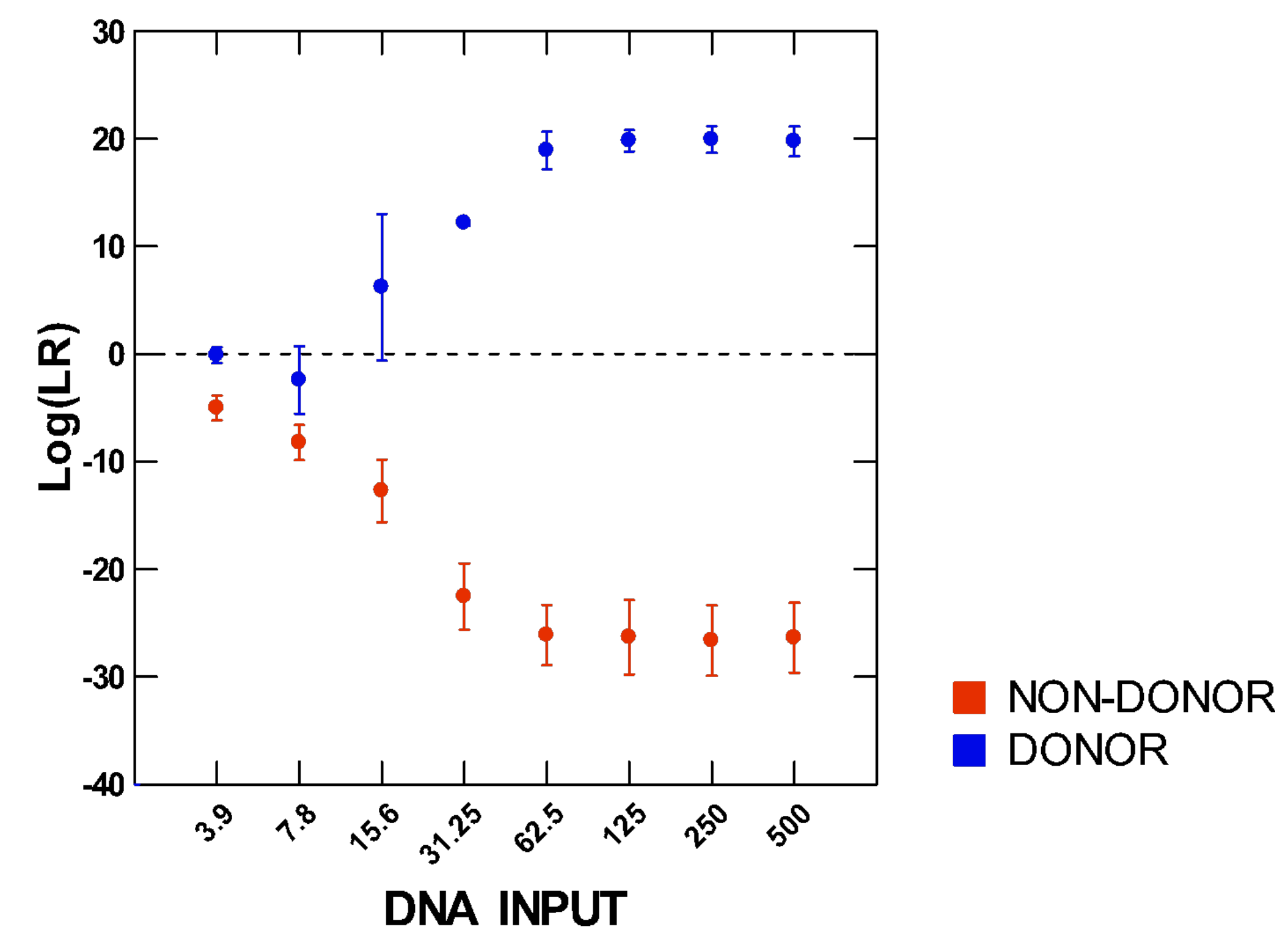
Dot plot showing two person mixture specificity by reference sample. Mean values from one mixture set are shown below; reference donors 4 and 18 (far right) were used in the creation of all mixture ratios from this mixture set. Error bars represent one standard deviation; dashed line is set at zero.

Average donor
log (LR) = 12.72

Average non-donor
log (LR) = -22.82
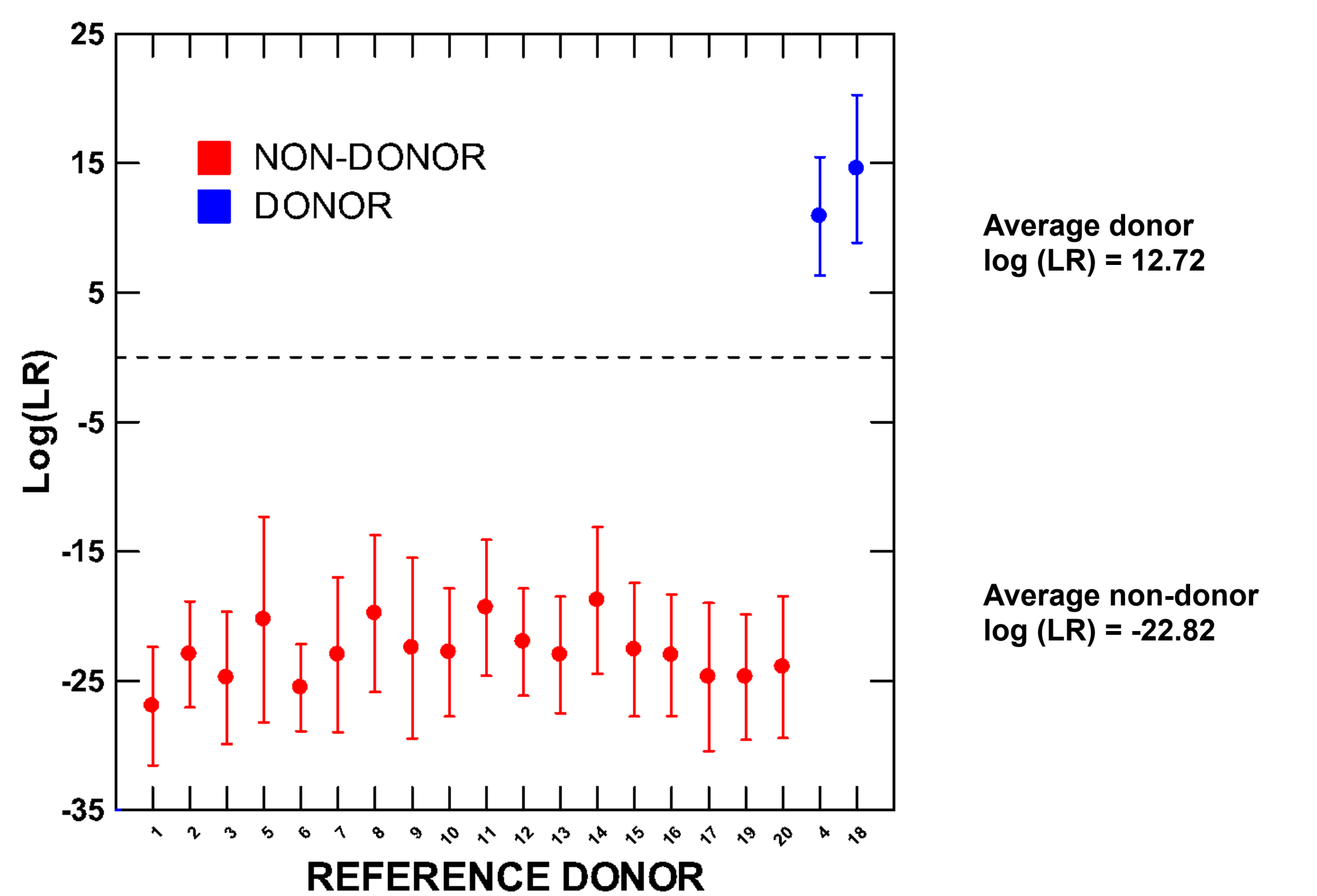
Dot plot showing three person mixture specificity by reference sample. Mean values from one mixture set are shown below; reference donors 3, 11, and 14 (far right) were used in the creation of all mixture ratios from this mixture set. Error bars represent one standard deviation; dashed line is set at zero.

Average donor
log(LR) = 7.54

Average non-donor
log(LR) = -17.62

5. RESULTS CONTINUED

Dot plot showing four person mixture specificity by reference sample. Mean values from one mixture set are shown below; reference donors M5, M6, M7, and M8 (far right) were used in the creation of all mixture ratios from this mixture set. Error bars represent one standard deviation; dashed line is set at zero.

Average donor
log(LR) = 5.41

Average non-donor
log(LR) = -11.39
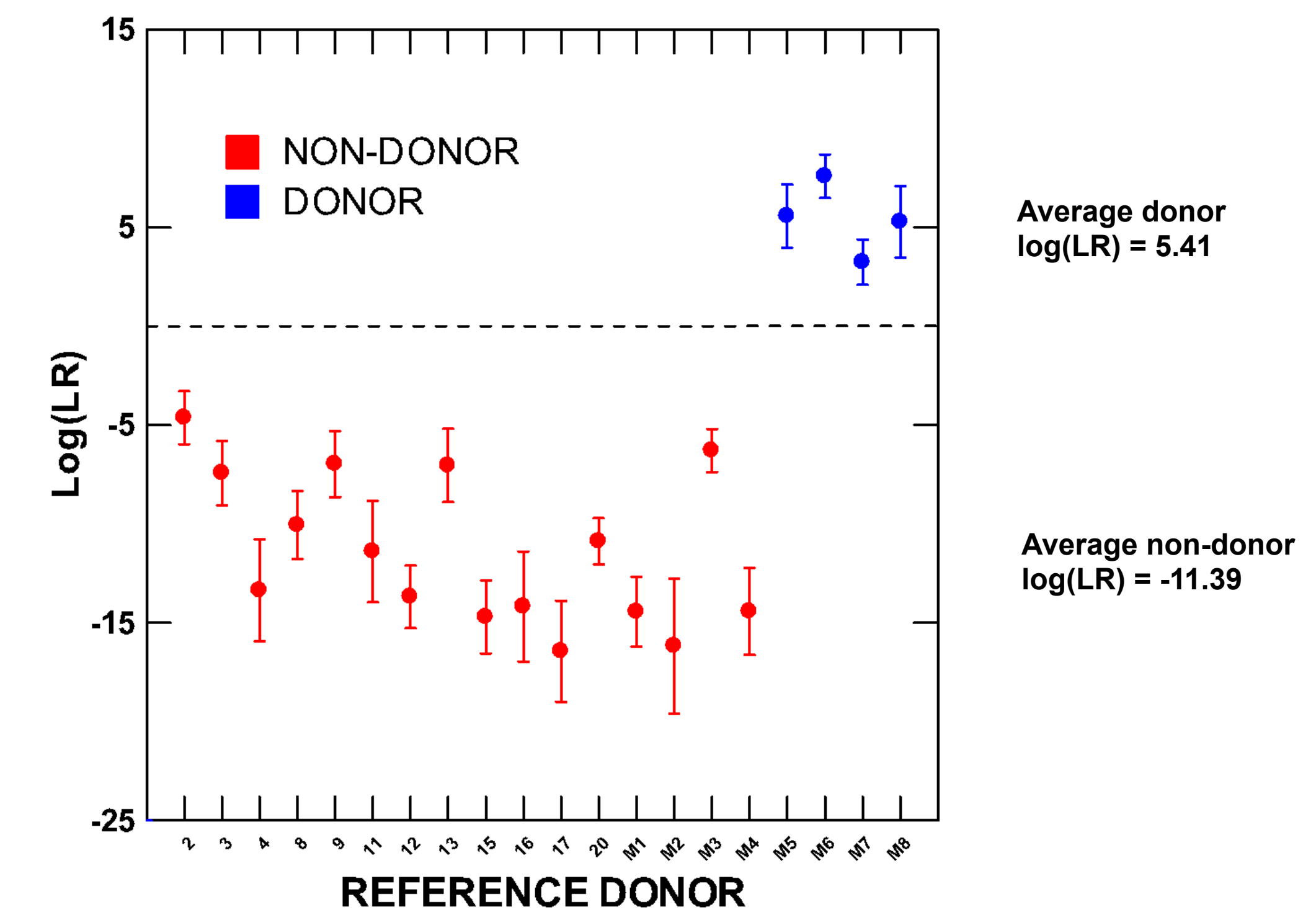
Dot plot showing specificity as a function of contributor number. Mean values from all two, three, and four person mixture sets are shown (n = 64). Error bars represent one standard deviation; dashed line is set at zero.
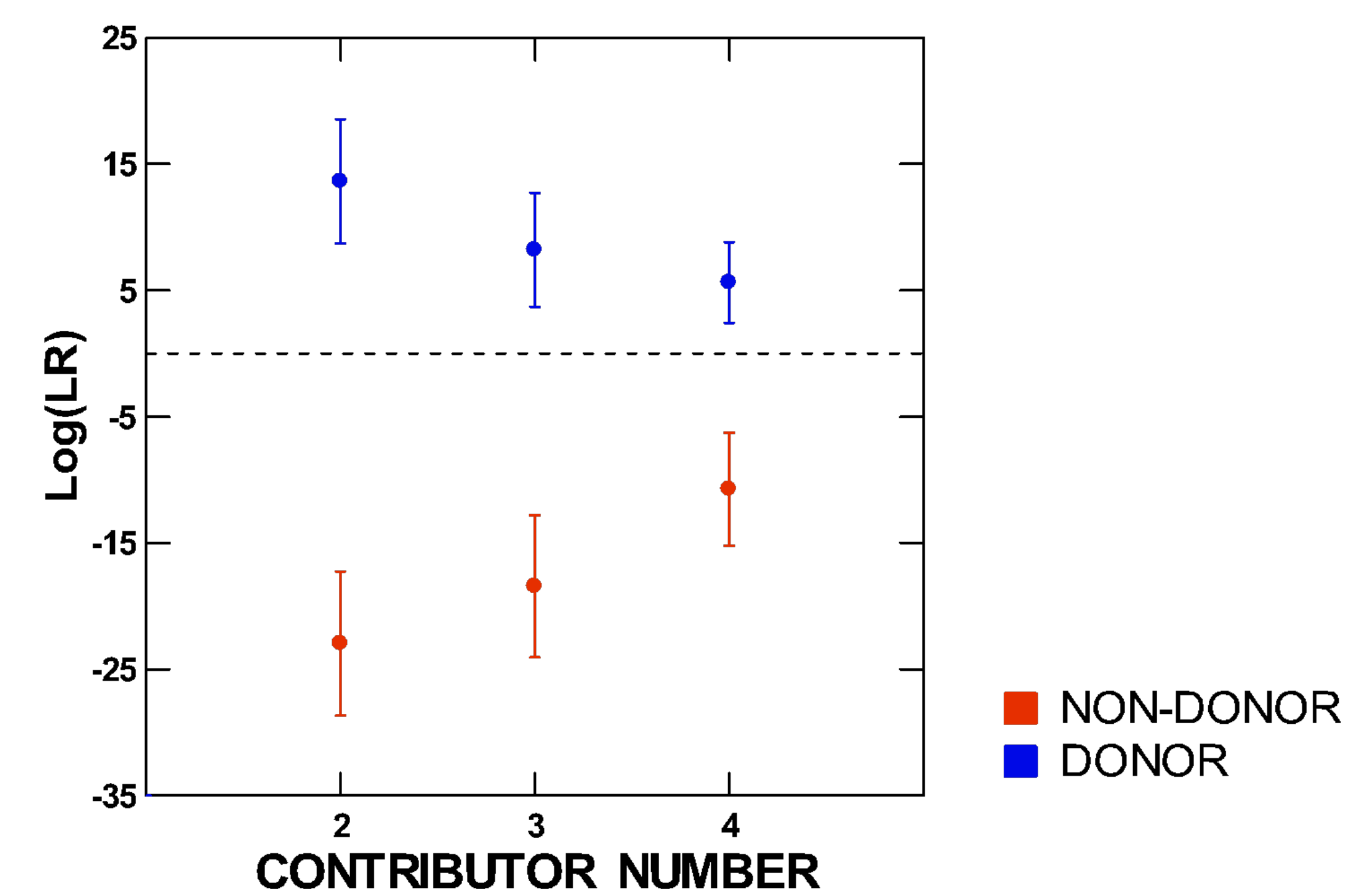
Histogram showing specificity as a function of contributor number. Data shown include all inferred genotypes from two, three and four person mixtures run against 3,000 randomly generated profiles (564,000 pairwise comparisons). The false positive error rate observed was 0.00024 with a maximum non-donor log(LR) of 1.945.
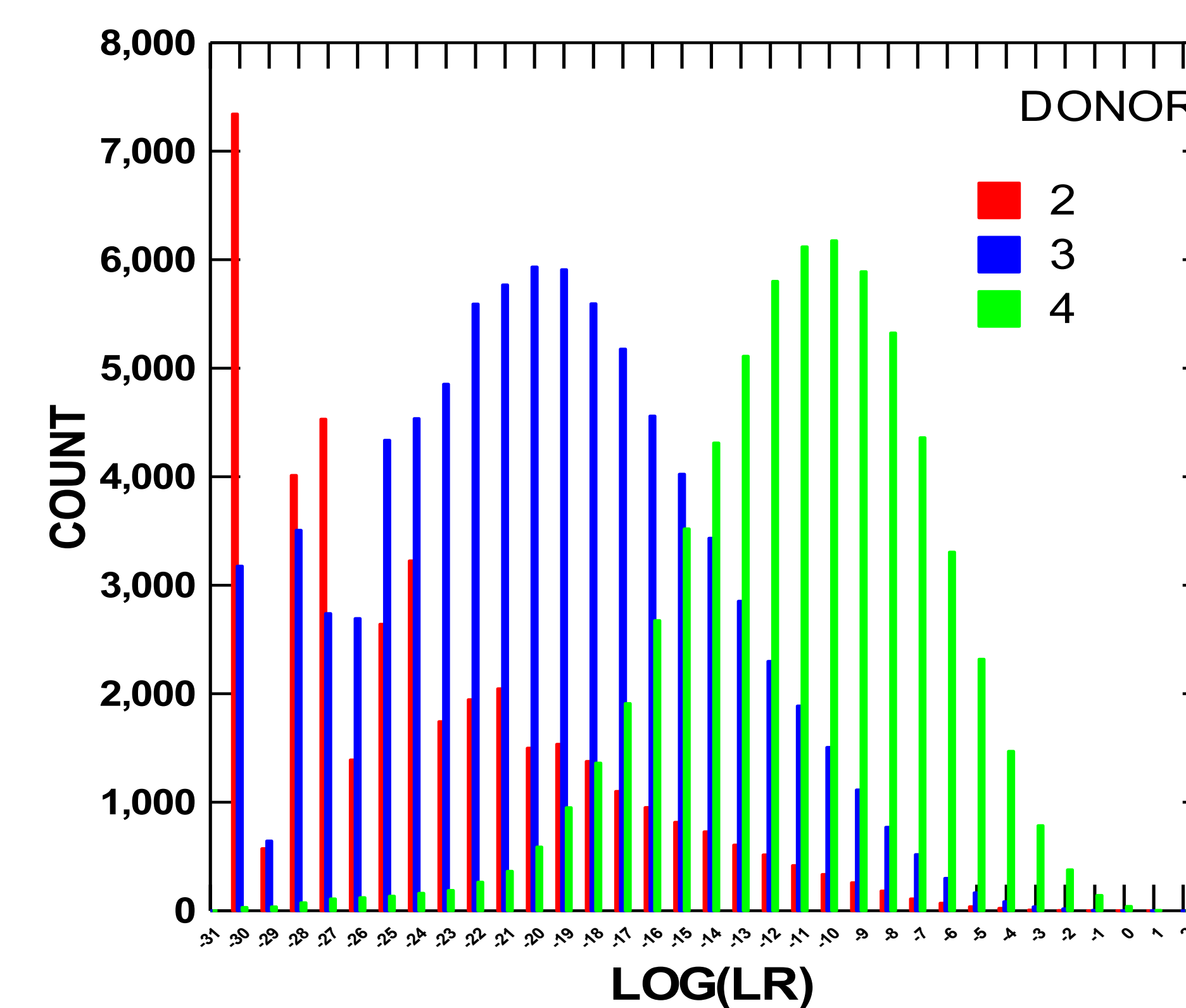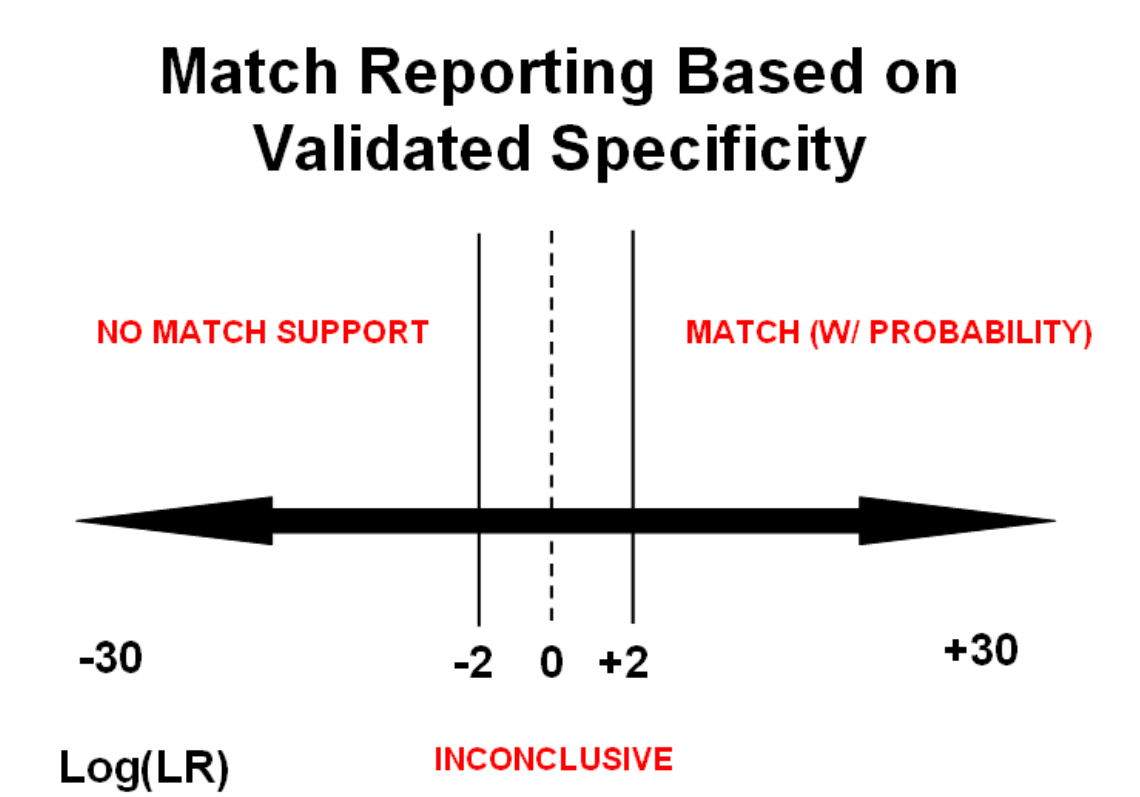
6. DISCUSSION

Uncertainty exists in virtually all fields of science. In forensic STR analysis, this uncertainty may take the form of partially recovered genotypes, complex mixture profiles, or an inability to accurately provide weight of evidence.   All currently used threshold-based methods attempt to address uncertainty by either discarding or altering observed DNA data, resulting in a loss of valuable genetic information with potential costs to public safety. By modeling all observed peak height variation with MCMC, computer-based genotype inference can overcome stochastic effects and produce more scientifically rigorous match results. The validation data shown here demonstrate how likelihood ratio calculations based on quantitative peak height information may be used to measure the extent of separation between individual known donor and non-donor genotypes. Results indicate that TrueAllele Casework is highly specific and can reproducibly discriminate between matching and non-matching reference profiles.

7. CONCLUSIONS

• The fully continuous approach to probabilistic genotyping can preserve more information than current inclusion/exclusion methods, resulting in highly specific genotype inference.

• Single source data suggest a donor limit of detection of approximately 15pg input DNA, although clear separation between donor and non-donor log(LR) values may be obtained below that amount.

• A decrease in specificity is evident with increased contributor numbers and mixture complexity. However, an average separation of over 16 log units between donor and non-donor LR was still observed across the four person mixture data for unrelated individuals.

• The high genotype specificity obtained in validation comparisons allows an objective, standardized approach to DNA match reporting based on log(LR) values as shown in the schematic below.

Match Reporting Based on Validated Specificity

NO MATCH SUPPORT   MATCH (W/ PROBABILITY)

-30   -2 0 +2   +30

Log(LR)   INCONCLUSIVE

8. REFERENCES

1. Gill P and Buckleton J (2010) A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Sci. Int. Genet.* 64: 221-227.

2. Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, Prieto L, Prinz M, Schneider H, Schneider PM, and Weir BS (2012) DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Sci. Int. Genet.* 6: 679-688.

3. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, and Duceman BW (2011) Validating TrueAllele® DNA Mixture Interpretation. *J. Forensic Sci.* 56(6): 1430-1447.