Document Title:	**Forensic Identification of an Individual in Complex DNA Mixtures Utilizing SNP Technology**
Author(s):	**Aaron LeFebvre, Ph.D., Sheri Ayers, M.S., Cynthia Smitherman, Lev Voskoboinik, Ariel Darvasi, Ph.D.**
Document Number:	**252004**
Date Received:	**November 2018**
Award Number:	**2011-DN-BX-K555**

*formerly Orchid Cellmark

Forensic Identification of an Individual in Complex DNA Mixtures Utilizing SNP Technology

Award No.: 2011-DN-BX-K555

Final Technical Report

Aaron LeFebvre PhD, Sheri Ayers MS, Cynthia Smitherman (Cellmark Forensics, Dallas, TX)

Lev Voskoboinik (Forensic Biology Laboratory, Israel Police, Israel Police National HQ, & Hebrew University, Jerusalem, Israel)

Ariel Darvasi PhD (Hebrew University, Jerusalem, Israel)

Abstract

Common forensic scenarios present DNA evidence that comprises a mixture of several contributors. Identifying the presence of an individual in such mixtures has proven difficult. In the current study we evaluate the usefulness of SNP microarrays for such purposes. We found that a set of 3000 SNPs specifically selected for this purpose can accurately identify the presence of an individual in complex DNA mixtures of various compositions, including mixtures of up to ten individuals. For example, in mixtures containing six individuals, those contributing as little as 5% to a complex DNA mixture can be robustly identified even if the starting DNA amount was as little as 5.0ng and had undergone whole-genome amplification (WGA) prior to SNP analysis. Additionally, under these same conditions individuals can be identified with similar statistical power in the presence of common forensic sample type inhibitors, such as hematin, humic acid, or indigo dye and with more moderate statistical power when subjected to conditions resulting in moderate to severe DNA degradation. Finally, this approach was able to identify each of ten individuals in a single DNA mixture with high likelihood when the recommended amount of starting DNA was used as well as distinguish between parent-child pairs, though the latter required a slight alteration in likelihood ratio calculations. In the absence of a focused forensic SNP microarray containing ~3000 SNPs, the use of standard arrays can be similarly used thereby increasing statistical power due to the larger amount of available information.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents

Executive Summary

This executive summary will briefly describe the proposed research for Award No. 2011-DN-BX-K555, "Forensic Identification of an Individual in Complex DNA Mixtures Utilizing SNP Technology", as well as the reasons for delays in funding release and the resulting modifications to the originally proposed research plan.

As originally proposed, our research plan was designed to develop and test a custom single nucleotide polymorphism (SNP) microarray based on Illumina's GoldenGate SNP genotyping technology. This custom SNP microarray was to be composed of several thousand (1000-3000) SNPs all with a minor allele frequency (MAF) in the range of 0.05 – 0. Due to the nature of SNP microarray optimization it was anticipated that development of the custom array was to be completed through multiple iterations in order to get those SNPs that both fulfilled the MAF requirement and performed well together as a complete multiplex on the custom array; validation of performance of the panel would have to be accomplished. The original proposal was to develop this microarray and test basic proof of principle, that a SNP microarray composed of several thousand "rare" SNPs could identify an individual in a complex forensic mixture composed of up to ten individuals, based on the previously published work of Voskoboinik and Darvasi [1].

Within a few months of the notification of our award by the National Institute of Justice (NIJ), Orchid Cellmark was acquired by Laboratory Corporation of America Holdings (LabCorp) in December 2012 and re-branded as Cellmark Forensics shortly thereafter. As a result of this business acquisition, release of our awarded funds was delayed several months due to transition of administrative tasks associated with the grant within Cellmark Forensics

management. As a result of and concurrent with this delay we investigated a more cost effective approach to accomplish our original research goal to maximize the amount of data we could generate within a shorter period of time. Rather than take a customized approach utilizing Illumina's GoldenGate technology we identified an "off-the-shelf" SNP microarray that would allow us to accomplish our goal of testing proof of principle without costly development of a customized chip. To this end, we selected Illumina's HumanCytoSNP-12 Infinium HD microarray for our research. This change in scope eliminated development time and costs since this microarray was pre-validated. Though slightly more expensive per sample the new approach would give us a much larger amount of informative data for our purposes as compared to the custom GoldenGate assay we originally proposed. With this change in scope we also made a change in collaborating institution from Baylor Research Institute (BRI; Dallas, TX) to the DNA Microarray Core Facility at the University of Texas – Southwestern (UTSW; Dallas, TX), both within driving distance of the Cellmark Forensics facility. We had initially proposed buying the custom GoldenGate microarrays and kits as well as processing of samples on the microarrays ourselves while BRI simply ran SNP detection; unfortunately after deciding to change our approach we learned they did not have the capabilities to process the HumanCytoSNP-12 Infinium HD microarray. However, UTSW was able to perform all SNP genotyping setup as well as SNP detection and general SNP analysis for the HumanCytoSNP-12 Infinium HD microarray once given DNA extracts to process. The total cost per sample for sample processing by UTSW using the HumanCytoSNP-12 chip was $235.00/sample compared to $75.67/sample for the originally proposed custom GoldenGate microarray. Whereas initially we proposed 960 samples to be processed with the custom GoldenGate microarray, including custom microarray development costs, we would be able to process 325 samples with the HumanCytoSNP-12 array

with zero development costs. Since the HumanCytoSNP-12 array is pre-validated and all SNPs present on the array are known to perform well together, the initial experiments designed to develop and fine-tune the custom array would no longer be required and therefore 325 samples would be enough to perform the experiments in the original proposal that were to be done following final custom array development and validation.

To begin our studies, we first had to generate complex mixtures of purified genomic DNA from various individuals. To this end, genomic DNA was extracted from Caucasians of the North American population and Ashkenazi Jews and utilized to compile several different complex mixtures of varying proportions of individuals (Tables 1 & 2). Six types of mixtures containing up to ten individuals with predetermined proportions were pooled from DNA samples of single individuals. In addition, these individual samples were used as reference samples for later SNP allelotyping analysis and likelihood ratio calculations.

The recommended DNA input for genotyping with the HumanCytoSNP-12 chip is 200ng, an amount atypical of common forensic samples. To more closely mimic conditions of typical forensic samples, in addition to testing mixtures that had been concentrated to the recommended amount we examined the use of whole genome amplification (WGA) for mixtures with low starting amounts of DNA (25.0ng, 5.0ng, and 1.0ng). Furthermore, we examined the effect of inhibitors commonly found in forensic samples (*i.e.* hematin – blood, humic acid – dirt) by compiling the same mixtures in the presence of various concentrations of these inhibitors.

Once all mixtures were generated, they were subjected to SNP allelotyping at the UTSW DNA Microarray Core Facility using the HumanCytoSNP-12 DNA Analysis BeadChip Kit

(Illumina, USA) on the iScan platform in accordance with the manufacturer's protocol. Once this SNP allelotyping data was generated, unique analysis parameters had to be developed.

The HumanCytoSNP-12 microarray contains approximately 300,000 SNP markers. To select the most appropriate SNPs for our analysis, first we filtered for the markers with a minor allele frequency between 0.02 and 0.15 based on the CEU HapMap population. Of the remaining markers, we next filtered for those that had a cluster separation score above 0.999, based on the GenomeStudio (Illumina®) software. Finally, we ranked the remaining 25,000 markers by their predicted performance at separating between a presence of 5% of a rare allele and the absence of the rare allele. Out of the 20,000 best ranking SNPs, a set of 3000 SNPs was selected to maximize the genetic distance between the loci, according to the genetic map of Matise et al. This selection was performed using the mean of normalized theta values as well as the standard deviation values recorded in GenomeStudio. Finally, the calling algorithm in GenomeStudio is designed for calling the genotype of a single source DNA sample. This software was used for calling the alleles in the individual DNA samples analyzed. However, this calling algorithm missed alleles present in low proportions in our complex mixtures. In order to call the presence of these alleles we used the "B allele frequency" value computed by GenomeStudio. If this frequency was above the threshold we set, then we called the allele as present in the DNA mixture.

Likelihood ratios were calculated between the following two competing hypotheses, where N is the number of contributors as estimated using a maximum likelihood approach:

1. Prosecution hypothesis: The "suspect" and N-1 unknown individuals comprise the mixture.
2. Defense hypothesis: N unknown individuals comprise the mixture.

The likelihood ratio was calculated following the equations listed in Table 4, where a is the frequency of the major allele, N is the number of donors, D is the drop-out rate of a heterozygous allele, D_h is the drop-out rate of a homozygous allele, and I is the drop-in rate. $\overline{D}$, $\overline{D_h}$ and $\overline{I}$ are the complimentary values of D, D_h and I. Since the B allele is rare, the probability of a mixture composed of all B alleles is negligible, and therefore was omitted from the table. The drop-out rates were evaluated from the number of missing minor alleles of each contributor in the mixture, up to a maximum rate of 0.5. The drop-in rate was calculated as the average drop-in rate in all tested mixtures for each starting DNA amount. Derivations of all equations used in this study are given in Materials & Methods. In addition, an explanation of how drop-out rates were estimated is given in Materials & Methods and Figure 1.

To establish the presence of an allele we needed to determine a threshold that the estimated frequency should pass (*i.e.* above noise levels) in order to be considered a true allele; this is analogous to a limit of detection; for a more detailed explanation see Materials & Methods. The lower we set the calling threshold, the lower the number of drop-out of minor alleles that will be observed (Type I error); conversely, the higher the number of drop-in of alleles (Type II error) that will be observed (Figure 2A). Since our goal was to maximize the LR of real donors and minimize the LR of non-donors, we examined the LR of minor donors and non-donors under different calling thresholds (Figure 2B). Since the range of calling threshold values between 0.01 and 0.02 seemed to form a plateau of close to optimal LR results for donors in the more challenging situations of relatively low DNA amounts or when low proportions of the "suspect" are present and for non-donors, a steady decrease in LR is observed as the calling threshold gets higher, the higher end of the optimal calling threshold range (0.02) was selected for further analysis.

Approximately 25,000 of the 300,000 SNPs present on the HumanCytoSNP-12 microarray have a high cluster separation score and a minor allele frequency in the range of 2%-15%, which is optimal for analyzing mixtures of 3-10 contributors [1]. In Figure 3 we present the LR for various conditions as a function of the numbers of SNPs. The LR increases logarithmically with the increase in the number of SNPs. Since the best performing SNPs were included in the analysis first, above 20,000 SNPs the improvement of the LR is a bit more moderate. Furthermore, the results of the current laboratory analysis, as well as our previous theoretical analysis [1] show that about 3000 SNPs are sufficient for the vast majority of forensically relevant mixtures. Therefore, in this study we elected to use 3000 SNPs for all of the following analyses.

As expected, the allele drop-out rate increased as the amount of starting DNA quantity is lowered (Figure 4). In contrast, the drop-in rate didn't show a similarly direct correlation with the starting DNA amounts. We found an average drop-in rate of 11.9%, 1.6% and 6.6% for starting DNA quantities of 1ng, 5ng and 25ng, respectively. These values serve as the drop-in estimates in the LR equations. Despite the increase in drop-out rate with lower starting DNA amounts, the analyses performed on low DNA amounts allowed a robust identification of the suspect when present in the mixture at reasonable proportions. For example, alleles of individuals contributing 10% of the DNA in a mixture of 5.0ng total DNA displayed a drop-out rate of 14.0%. The LR for such contributors ranged between 10^{58} and 10^{126}, thus robustly identifying them.

Individuals contributing to the mixtures as well as 267 individuals also genotyped with the HumanCytoSNP-12 microarray in the HapMap project[8] from CEU, CHB, JPT and YRI populations were compared to the mixtures. The results of the comparison are summarized in

Figure 5A. All individuals contributing 15% or more to a mixture had a LR far above 1 (the minimum was 1.65 $*10^{9}$), even when the starting DNA quantity was only 1.0ng (Figure 5B). Likewise, when the starting DNA amount was 5.0ng or more, all individuals contributing at least 5% to the mixture resulted in a LR far above 1 (with the minimum of 1.43 $*10^{28}$). Out of the non-contributing CEU individuals analyzed (about 5500 comparisons in all mixtures and for all tested individuals) none displayed a LR above 1. When comparing individuals from CHB, JPT and YRI populations, in which allele frequencies are substantially different, less than 0.1% had a LR above 1 (Figure 5B) and none reached the minimum values presented above (the highest LR for a non-contributor was 2406).

We further tested the robustness of this approach by testing a more complex DNA mixture composed of ten individuals, each comprising 10% of the mixture (Figure 6). When the starting DNA amount was high (200ng), the LR of all 10 contributors was high, ranging from 10^{17} to 10^{61}. When lower starting amounts of DNA subjected to WGA were utilized a majority of contributors had a LR > 10^{20}, 7 contributors when 5.0ng was the starting amount and 5 contributors when 1.0ng was the starting amount, with some higher than 10^{80}. For the most part this was due to extensive dropout of alleles from a few specific contributors, possibly due to degradation or competition for reagents (*i.e.* stochastic effects) during the WGA process. Regardless, this establishes that when recommended procedures are utilized, this approach can robustly identify individuals in complex mixtures containing up to ten individuals. A focused forensic array containing only the SNPs we have selected here may negate the need for WGA and therefore give results more comparable to those generated here with 200ng of starting DNA but for far lower amounts of starting DNA (*i.e.* 1.0ng or less).

In forensic DNA samples, it is not uncommon for certain inhibitory compounds to be present due to the nature of the sample type from which the DNA was collected. For example, hematin in blood can be present in purified DNA from this sample type. Likewise, humic acid in soil and indigo dye from denim material can be present in forensic DNA samples collected from samples that have had contact with soil or that were deposited on denim material (*i.e.* jeans). Additionally, it is not uncommon for forensic DNA to be highly degraded, resulting in poor STR amplification, and difficulty with interpretation due to the limited data that can be generated with these types of samples. Recently, STR amplification kits such as Identifiler® Plus (Applied Biosystems) and PowerPlex® 16 HS (Promega) have been developed in order to overcome the presence of both inhibitory compounds as well as result in improved amplification of degraded DNA. However, as previously discussed, for our purposes, current commercially available STR amplification kits are sufficient for interpretation of complex mixtures with more than three contributors or for cases where the contribution of the suspect is relatively low.

We therefore tested our approach with complex mixtures using various amounts of DNA either in the presence of various forensic sample type inhibitors (hematin, humic acid, and indigo dye) or with treatment resulting in degradation, all resulting in some level of inhibition of STR amplification. With the recommended starting DNA amount of 200ng (Figure 7A), when compared to the same mixtures that had been untreated all three inhibitors as well as DNA degradation had virtually no effect on the ability to robustly identify individuals in these mixtures with a LR of 10^{100} or higher when an individual contributed as little as 5% to the mixture and a LR of 10^{20} or higher when the contribution was only 2%. Lower starting DNA amounts gave slightly more variable results and were more dramatically affected by conditions of inhibition or degradation (Figures 7B-C), suggesting that these conditions affect WGA more

than downstream SNP genotyping. The effect on the LR of each of the conditions tested, from least to most dramatic, were indigo dye > hematin >> humic acid > degradation. This in turn was dependent on the starting DNA amount. For example, when 5.0ng was the starting DNA amount, hematin resulted in a LR > 1 while humic acid resulted in a LR < 1for individuals who contributed 2% to the mixture. When 1.0ng was the starting DNA amount, all treatments except indigo dye resulted in a LR < 1even for individuals who contributed as much as 20% to the mixture. Indigo dye had the weakest effect, with high LR values for contributors of as little as 5% when as little as 1.0ng was tested. Degradation by far had the most severe effect, with a LR ~ 1for contributors of up to 20% when 1.0ng was the starting DNA amount. However, when 5.0ng was the starting DNA amount, a LR of 10^{25} was achieved for individuals who contributed 20% to the mixture and greater than 10^{100} for those who contributed 50% to the mixture. Similar to the results of the 10 person mixture, a focused forensic array may negate the need for WGA and therefore lessen or even eliminate the effects of inhibition or degradation observed in this study as well as result in LRs comparable to those generated here with 200ng of starting DNA for far lower amounts of starting DNA (*i.e.* 1.0ng).

Additionally, we tested our approaches ability to distinguish between a true contributor and a relative of the contributor. The possibility that a suspect's relative is contributing to a mixture introduces two problems:

1. The suspect may receive a high LR although he is not a real contributor to the mixture, due to the shared genetic material with a relative that is a contributor (corresponds to error type I).
2. A real contributor might claim he isn't part of the mixture, but was identified as a contributor because of the presence of his relative in the mixture (corresponds to error type II).

First examining Problem 1 (the suspect's relative is the true contributor to the mixture) we computed the likelihood ratio between our two classic hypotheses:

1. Prosecution hypothesis: The "suspect" and N-1 unknown individuals comprise the mixture.
2. Defense hypothesis: N unknown individuals comprise the mixture.

Comparing Parent - Child pairs in different mixtures, relatives of real contributors receive a high LR and may be falsely included. We found that the log(LR) of a relative is linearly correlated with the log(LR) of the real contributor. In order to minimize this problem, the hypothesis for the likelihood ratio calculation needed to be modified:

1. Prosecution hypothesis: The "suspect" and N-1 unknown individuals comprise the mixture.
3. Defense hypothesis: The "suspect's" relative and N-1unknown individuals comprise the mixture.

In the case of Parent-Child pairs the LR equations are modified. Since the drop-out rate of the hypothetical contributing relative cannot be calculated directly, the drop-out rate was selected by numerically maximizing the likelihood of hypothesis **3** above. The new hypotheses pair (**1 vs 3**) lowers the LR of all contributors by 50% or more but has a more dramatic effect for relatives, bringing the LR in most cases close to or below 1. Nevertheless, in a few cases the LR of a non-contributing relative remained high even with the new hypothesis, the highest result being $10^{8.42}$.

We can overcome this problem by reducing the maximum allowed drop-out rate from 0.5 to 0.3. The result is that all relatives received a LR below 1. The downside of this solution is that low level contributors that do have a drop-out rate between 0.3-0.5 would receive a lower LR, and may be incorrectly excluded. When compared to a maximum drop-out rate of 0.3 to

accommodate this problem with related pairs, the result was lower LRs for some contributors, including one that previously could not be excluded (maximum drop-out rate of 0.5) that was now clearly excluded (maximum drop-out rate of 0.3).

Next examining Problem 2, a suspect may claim that he is wrongly included because his relative is in the mixture, we again need to change the hypothesis pair from 1 vs. 2 to 1 vs. 3. As mentioned above the change in the hypothesis reduces the LR by at least 50%. Moreover, contributors that previously displayed a LR < 50 may now receive a negative value. When analyzing this data as a function of the starting DNA amount, a high LR for the inclusion of the suspect (as compared to his relative) was achieved for contributors of at least 5%, 10% and 20% in mixtures starting with 25.0ng, 5.0ng and 1.0ng total DNA, respectively. By comparison, with our previous conditions the minimal contribution required for a high LR was 2%, 5% and 15% in mixtures starting with 25.0ng, 5.0ng and 1.0ng total DNA respectively.

To summarize, we can ensure a non-contributing "suspect" is not wrongly included even if his relative is in a mixture. To do so, we need to change the defense hypothesis of the LR from hypothesis 2 to hypothesis 3, as indicated above, as well as lower the maximum allowed drop-out rate. Both of these changes lower the calculated LR of contributors meaning a higher proportion of a contributor in the mixture is required to robustly identify his presence. The change in the required proportion to ensure inclusion is not very large, so minor contributors can still be identified (depending on the total DNA amount used).

The presence of complex DNA mixtures is common in forensic cases. When analyzed with standard STRs, results are mostly inconclusive or not admissible in court. We have previously suggested an alternative method using a few thousand SNPs to accurately determine

the presence or absence of an individual in a complex DNA mixture found at the scene of crime [1]. In the current study, using standard genotyping technologies and off-the-shelf SNP microarrays we have shown the practical applicability of this method. Our results show in general that this technology is accurate and meets the strict requirements of a forensic setting. Additionally, we have shown that this technology has a relatively high sensitivity that allows the detection of an individual that contributes as little as 5% of the total DNA in a forensic mixture.

Furthermore our likelihood ratio (LR) analysis provided very strong support for the presence of an individual in a mixture even when the relative contribution to the mixture was small and the number of contributors was high. Non-contributors rarely exhibited a LR>1. In this study we described mixtures of up to ten contributing individuals. However, the parameter that primarily affects performance is the proportion of DNA contributed by the suspect. The effect of the number of contributors does not affect accuracy of the typing methods but does in fact have a statistical effect as previously described [1].

Even in mixtures with only 1.0ng of total starting DNA, a robust inclusion of contributors was observed as long as they contributed 15% (only 150pg) or more to the DNA mixture. With higher DNA starting amounts, individuals contributing only 5% to the mixture were detected. This is of particular importance in some forensic cases. For example, often in sexual assault cases the amount of total DNA may be high while the individual of interest represents a very minor contribution to the mixture. We found that when 25ng of DNA is used as the starting amount it is possible to detect individuals with high certainty even if their contribution to the mixture was as little as 2%. Furthermore, when the recommended starting DNA amount was utilized, we were able to robustly identify all ten individuals in a single mixture where each individual contributed only 10% to the mixture. With lower starting amounts of DNA the result

was more variable, only indentifying a slight majority of the contributors, but may in part have been due to the requirement of WGA to achieve the recommended starting DNA amount. A focused forensic array containing only the SNPs we have selected here may negate the need for WGA and therefore result in more robust identification with smaller starting DNA amounts. Finally, with slight alterations to our likelihood ratio calculations we were able to ensure a non-contributing "suspect" is not wrongly included even if his relative is in a mixture.

We also testing our approaches ability to identify individuals in complex mixtures in the presence of inhibitors, such as hematin and humic acid, as well as when DNA has been highly degraded. Using the recommended amount of starting DNA (200ng) we could robustly identify individuals who comprised as little as 2% of the mixture. With lower starting amounts of DNA, the effects of inhibitors and degradation were more dramatic, with only individuals contributing much higher percentages (20% or greater for 5.0ng, 50% or greater for 1.0ng) to the mixtures able to be identified with high likelihood ratios. Since the negative effects of inhibition and degradation were more dramatic with lower starting amounts of DNA, this suggests that these conditions affect WGA more than downstream SNP genotyping. An alternative WGA kit designed for formalin fixed – paraffin embedded (FFPE) tissue may perform better in this aspect and may be tested in future studies. Again, the development of a focused forensic microarray with only the SNPs required for our analysis may negate the need for WGA due to the proportionally lower requirement for starting DNA; 2ng for our selected 3000 SNPs as compared to 200ng for 300,000 SNPs. This may lessen or even eliminate the effects of inhibition or degradation observed in this study and could result in LRs comparable to those generated here with 200ng of starting DNA without the need for such amounts of DNA atypical of forensic samples.

Based on our results we believe that the current method can be implemented in forensic laboratories relatively quickly. We used a commercially available DNA chip that contains about 300,000 SNPs. However in the analyses we considered only 3000 of them. The development of a focused forensic DNA chip with a few thousand SNPs is likely to prove valuable in terms of reducing per sample costs. Additionally, this would reduce the amount of DNA required for typing and would improve typing accuracy. As observed in our studies, the main hurdle to overcome was the requirement for large amounts of starting DNA. The development of a focused forensic SNP microarray containing only the SNPs from the HumanCytoSNP-12 microarray we utilized for our analysis would eliminate this hurdle by greatly reducing the required amount of starting DNA. In addition, such a forensically focused SNP microarray may include SNPs of forensic importance such as appearance or ancestry [9-13]. On the other hand, until such a microarray is commercially available, the immediate use of commercially available chips as shown here has the advantage of additional statistical power due to the large number of SNPs available for analysis on these microarrays.

Introduction

Detecting the presence of individuals in complex DNA mixtures is an ongoing challenge in the current realm of forensic DNA analysis. Several statistical approaches for analyzing complex DNA mixtures have been proposed when using standard short tandem repeat (STR) kits [2-4]. However, the small number of STRs in those kits (up to 16) is insufficient for interpretation of complex mixtures with more than three contributors or for cases where the contribution of the suspect is relatively low. An alternative approach to analyzing complex DNA mixtures has been suggested that consists of using high density single nucleotide polymorphism (SNP) microarrays [5], although it has recently been criticized as inadequate for forensic use [6].

We have previously proposed yet another approach for analyzing complex mixtures that involves the genotyping of a few thousand SNPs, each with a relatively low minor allele frequency (MAF) [1]. We have examined the type and number of SNP markers required for robust analysis of complex DNA mixtures and presented the associated statistical calculations. Briefly, we have demonstrated that by genotyping 3000 SNPs with MAFs around 0.05-0.1 we can identify the presence of a person in a mixture of up to ten contributors with high statistical significance. In the current study, we present the implementation of this method by genotyping artificially constructed mixtures of DNA from several individuals using a standard SNP microarray, including mixtures composed of up to ten individuals. Additionally, we tested the ability of our approach to ensure a non-contributing "suspect" is not wrongly included even if his relative is in a mixture. We also examine the usefulness of whole genome amplification (WGA) of complex mixtures before genotyping as a means to cope with low DNA quantities, a common scenario in forensics. Furthermore, we tested the ability of our approach to identify individuals in mixtures when common inhibitors found in forensic type samples were present (*i.e.* hematin, humic acid) or when DNA had been highly degraded.

Currently accepted procedures are not able to easily distinguish between individual in DNA mixtures when more than three individuals are present. Using the HumanCytoSNP-12 microarray (Illumina®) we have developed an alternative solution to this problem and established proof of principle of the statistical model previously developed by Voskoboinik and Darvasi [1]. Based on our results we believe that the current method can be implemented in forensic laboratories relatively quickly while future research will attempt to develop a focused forensic microarray that may both lower costs and improve the performance of this method further.

Materials & Methods

Subjects, DNA mixtures and WGA

DNA of 12 unrelated individuals was used for mixture construction. Six of them originated from Ashkenazi Jews and six originated from the general Caucasian population in the United States. Genomic DNA was extracted using either of two standard extraction technologies in accordance with the manufacturer's protocol: manually with the Nucleon Blood and Cell Culture DNA Extraction Kit (Hologic Gen-Probe Incorporated, San Diego, USA) or in automated fashion on the BioRobot EZ1 using the EZ1 DNA Investigator Kit (Qiagen, USA).

Extracted DNA was quantified using either the Quantifiler® Human or Quantifiler® Duo real-time PCR kits on a 7500 Real-Time PCR System (Applied Biosystems, USA) in accordance with the manufacturer's protocol. Each sample was quantified in duplicate and normalized to a final DNA concentration of 50ng/µl.

For initial studies, six types of mixtures containing between 3 and 6 individuals with predetermined proportions were pooled from DNA samples of single individuals (Table 1). The desired proportions were obtained by including the corresponding amount of DNA from each individual, based on the quantification described above. Each type of mixture was prepared twice, once from each of the individual sets (Ashkenazi Jews and US Caucasians).

To test the ability to identify individuals in mixtures containing inhibitors commonly found in forensic type samples, mixtures (50ng/µl or diluted in preparation for whole genome amplification) were prepared in the presence of hematin (50µM), humic acid (50µM), and indigo dye (200µM) based on their ability to inhibit amplification with Identifiler.

To test the ability to identify individuals in mixtures that were degraded, mixtures 8 & 9 (50ng/µl or diluted in preparation for whole genome amplification) were incubated at 94ºC for 15 minutes prior to SNP genotyping. Degradation levels were confirmed by amplification with Identifiler. This treatment resulted complete dropout at larger amplicons and lower peak intensities at smaller amplicons.

For whole genome amplification (WGA), DNA mixtures containing 1.0ng, 5.0ng or 25ng of total genomic DNA were amplified with the REPLI-g Mini Kit (Qiagen, USA) in accordance with the manufacturer's protocol. Mixtures 8 and 9 (Tables 2 & 3) as well as all mixtures in Table 1 were diluted with water to final concentrations of either 1.0ng/µl or 0.2ng/µl. Following whole genome amplification, samples were quantified using Quant-iT™ PicoGreen® (Invitrogen, USA) in accordance with the manufacturer's protocol diluted to the recommended

SNP Allelotyping

All samples and mixtures were genotyped twice using the HumanCytoSNP-12 DNA Analysis BeadChip (Illumina, USA) in accordance with the manufacturer's protocol. For samples that did not undergo WGA, a total of 200ng of unamplified genomic DNA was used in each reaction as recommended. For samples that did undergo WGA, mixtures post-WGA were genotyped twice with 200ng DNA where the DNA concentrations allowed this. However, in several samples the DNA concentration after WGA was not high enough to allow for 200ng of total DNA in SNP genotyping. For such samples, the maximum possible DNA quantity was used with a minimum of 57ng of total DNA.

The calling algorithms of microarray analysis programs such as GenomeStudio (Illumina®) are designed for calling the genotype of a single source DNA sample. This software

was used for calling the alleles in the individual DNA samples analyzed. However, this calling algorithm will miss any alleles present in low proportions. Therefore, in order to call the presence of the more rare alleles in the DNA mixtures tested we used the “B allele frequency” value computed by GenomeStudio. If this frequency was above the threshold we have set, then we called the allele as present in the DNA mixture. For example, the raw intensities of florescence of the two dyes associated with the two alleles are translated by GenomeStudio into a single value called "B Allele Frequency" that is assigned values between 0 and 1. If we attribute a calling threshold of 0.02 this means that if the "B Allele Frequency" is below 0.02, the "B" allele is not called, and the mixture is considered "AA". Likewise, if the "B Allele Frequency" is above 0.98, the "A" allele is not called and the mixture is considered "BB". Any "B Allele Frequency" value in the range of 0.02-0.98 is considered as mixture allelotype "AB". With low level samples, lowering the calling threshold allows more alleles to be correctly call “BB” rather than incorrectly called “AB” or “AA”.

SNP Selection

The HumanCytoSNP-12 microarray contains approximately 300,000 SNP markers. To select the most appropriate SNPs for our analyses, we first filtered for the markers with a minor allele frequency between 0.02 and 0.15 based on the CEU HapMap population. From this subset of SNPs we further selected those that have a cluster separation score above 0.999, based on the GenomeStudio software. This resulted in 25,000 SNPs, which were ranked by their predicted performance accuracy. This selection was performed using the mean of normalized *theta* values formula: (0.9*(MeanAA– 2* StdAA)+ 0.1*(MeanAB – 2* StdAB)) – (MeanAA + 3* StdAA), as well as the standard deviation values recorded in GenomeStudio, according to the following

,where MeanAA is the mean of normalized *theta* values for the major allele homozygote cluster, StdAA is the standard deviation of normalized *theta* values for the major allele homozygote cluster, MeanAB is the mean of normalized *theta* values for the heterozygote cluster and StdAB is the standard deviation of normalized *theta* values for the heterozygote cluster. Out of the 20,000 best ranking SNPs, a set of 3000 SNPs was selected to maximize the genetic distance between the loci, according to the genetic map of Matise et al [14] . The resulting set was relatively uniformly distributed across the genome with a minimal distance of 0.93cM between any two adjacent loci and an average of 1.2cM. Linkage disequilibrium at such distances is essentially null.

Likelihood Ratio Calculations

Likelihood ratios were calculated for the following alternative hypotheses (where N is the number of contributors):

1. Prosecution hypothesis: The "suspect" and N-1 unknown individuals are the contributors to the mixture.
2. Defense hypothesis: N unknown individuals are the contributors to the mixture.

To calculate the LR, a value for the number of contributors (N) is required. We estimated N using a maximum likelihood estimation (MLE) approach, considering all values in the range of 1 to 15 with a step of 0.1, using a numerical maximization of the likelihood function. The non-integer values for N represent a mixture with one or more individuals contributing a low amount of DNA and having drop-out alleles. The likelihood of an AA allelotype at a given SNP equals the probability that all the alleles of all contributors are A and no drop-in of allele B occurs (Equation 1):

$$a^{2N} \cdot \overline{c} \qquad \text{Equation 1}$$

,where a is the frequency of the major allele, N is the number of contributors and $\overline{c}$ is the probability of no drop-in.

The likelihood of a BB allelotype is symmetrically $b^{2N} \cdot \overline{c}$. Since the SNPs are selected so that the allele frequency of B is low (0.1 on average), the probability of a BB allelotype, even in the simplest complex mixture of three, equals approximately $b^{2N} = 0.1^6$, or 1 in a million, thus it can be neglected.

The likelihood of the remaining allelotype, AB, is approximately the compliment of Equation 1:

$$1 - (a^{2N} + b^{2N}) \cdot \overline{c} \approx 1 - a^{2N} \cdot \overline{c} \qquad \text{Equation 2}$$

We find the N that maximizes the product of likelihoods across all L loci using indicators:

$$N = \underset{N}{argmax} \prod_{i=1}^{L} \left[({a_i}^{2N} \cdot \overline{c}) I_{\{G_i = AA\}} + (1 - {a_i}^{2N} \cdot \overline{c}) I_{\{G_i = AB\}} \right] \qquad \text{Equation 3}$$

,where G_i is the mixture allelotype at loci i. Once N has been established, calculation of the LR is possible.

The likelihood of the defense hypothesis when the mixture allelotype is AA equals the probability that all N individuals are AA homozygotes, and no drop-in of allele B occurred. This results in Equation 1 above with the only difference that N is now constant. Similarly, the likelihood of the defense hypothesis for mixture allelotype AB is calculated by Equation 2. Since the equations for selecting the maximum likelihood estimate of the number of contributors (N) are the same equations as for calculating the likelihood of the defense hypothesis, this selected N

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

optimizes the likelihood of the defense hypothesis and thus is the most conservative possible estimate.

The equations for the likelihood of the prosecution hypothesis depend on the genotype of the suspect. If the mixture allelotype is AA and the suspect is AA then the likelihood of observing the allelotype is equal to the probability that all the contributors besides the suspect have only A alleles and no drop-in of allele B occurs, $a^{2(N-1)} \cdot \overline{c}$. If we allow for drop-out of the suspect alleles at a rate of *c*, then the number of the alleles that the suspect account for is not 2 but the number of un-dropped alleles, $2\overline{D}$. Thus the likelihood is calculated as:

$$a^{2(N-\overline{D})} \cdot \overline{c} \qquad \text{Equation 4}$$

If the mixture allelotype is AA and the suspect is AB then the likelihood of observing the allelotype is equal to the probability that all the contributors besides the suspect have only A alleles and no drop-in of allele B occurred and the B allele of the suspect has dropped out. This probability is calculated as:

$$a^{2(N-\overline{D})} \cdot D \cdot \overline{c} \qquad \text{Equation 5}$$

The likelihood for mixture allelotype AA when the suspect is BB can be calculated similarly to Equation 5:

$$a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c} \qquad \text{Equation 6}$$

,where D_h is the dropout rate of a homozygote allele of the suspect. Since the proportion of the homozygote B allele in the mixture is twice as high, the rate of its dropout would be lower than *D*.

The likelihoods of an AB allelotype in the mixture given a suspect genotype of AA, AB and BB are the complementary equations to Equations 3, 4 and 5 respectively. Hence,

$$1 - a^{2(N-\overline{D})} \cdot \overline{c} \qquad \text{Equation 7}$$

$$1 - a^{2(N-\overline{D})} \cdot D \cdot \overline{c}. \qquad \text{Equation 8}$$

$$1 - a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c} \qquad \text{Equation 9}$$

The formulas for calculating the likelihood ratio between the prosecution hypothesis (Equations 3-8) and the defense hypothesis (Equations 1-2) for each combination of mixture allelotype and suspect genotype are given in Table 4.

The drop-out rates D, D_h were evaluated for each suspect and mixture by estimating the number of missing minor alleles of the suspect in the mixture, up to a maximum rate of 0.5. The simplest estimation of the heterozygote drop-out rate (D) is the ratio between the number of suspects' heterozygous rare alleles missing from the mixture (M) and the total number of suspects' heterozygote rare alleles (AB). This estimation however underestimates the true value since some rare alleles of the suspect are shared by other individuals in the mixtures and are masked from being dropped-out. To compensate for this masking, the simple ratio M/AB is multiplied by the ratio between the number of "AA" loci in the suspect and the number of "AA" loci in the mixture. In this ratio, the number of "AA" loci in the suspect are not only real "AA" loci from the suspect's reference profile but includes the "AB" and "BB" loci that would have dropped-out in the mixture, if no masking occurred:

$$D = \frac{M}{AB} \cdot \frac{AA+AB \cdot D+BB \cdot D_h}{A_{mix}} \qquad \text{Equation 10}$$

,where M is the number of heterozygote B alleles of the suspect missing from the mixture (loci in which the suspect is AB while the mixture is AA, indicating a drop-out), AB and BB are the number of heterozygote and homozygote loci in the suspect, A_{mix} is the number of loci that display only the A allele in the mixture, and D and D_h are the drop-out rates of suspects' heterozygote and homozygote alleles respectively.

D_h can be estimated similarly to D but the estimation would be less accurate since the number of homozygote BB genotypes is small (on average 35 loci out of 3000) while a heterozygote genotype AB is common (average of 520 loci out of 3000). Thus we preferred to find an empiric correlation between the drop-out rate of an allele of proportion X and an allele of proportion 2X, corresponding to a homozygote allele (Figure 1) and use this correlation to estimate the D_h based on D alone. Thus D_h was estimated as

$$D_h = 1.0319 \cdot \mathrm{D}^2 - 0.1015 \cdot \mathrm{D} + 0.0035 \qquad \text{Equation 11}$$

D cannot be calculated immediately by Equation 10 since D and D_h which are part of the formula, are unknown. However we can approximate D by repeating the calculation of Equations 10 and 11 several times. For the first time we can place 0 at D and D_h. Then we repeat the calculation, each time inputting the D and D_h results from the previous iteration into Equations 10 and 11. With each round of calculation, D and D_h are asymptotically approaching fixed values and in practice after four or five rounds the estimations of D and D_h do not change beyond the fourth decimal point. Once we obtained these estimations of the drop-out rates we proceeded to the calculation of LR.

The drop-in rate (c) was calculated as the average drop-in rate in all tested mixtures for a given starting DNA amount. That is, we consider the drop-in rate to be a constant for a specific method that can be measured as part of its validation.

Results

Setting the Calling Threshold

As presented in Materials & Methods, to establish the presence of an allele we need to determine a threshold that the estimated frequency should pass (*i.e.* above noise levels). The lower we set the calling threshold, the lower the number of drop-out of minor alleles that will be observed (Type I error); conversely, the higher the number of drop-in of alleles (Type II error) that will be observed (Figure 2A). Since our goal is to maximize the LR of real donors and minimize the LR of non-donors, we examined the LR of minor donors and non-donors under different calling thresholds (Figure 2B).

The range of calling threshold values between 0.01 and 0.02 seem to form a plateau of close to optimal LR results for donors in the more challenging situations of relatively low DNA amounts or when low proportions of the "suspect" are present. For non-donors, a steady decrease in LR is observed as the calling threshold gets higher. Thus the higher end of the optimal calling threshold range at a value of 0.02 was selected for further analysis.

Number of Markers

The HumanCytoSNP-12 microarray contains about 300,000 SNPs. Approximately 25,000 of them have a high cluster separation score and a minor allele frequency in the range of 2%-15%, which is optimal for analyzing mixtures of 3-10 contributors [1]. In Figure 3 we present the LR for various conditions as a function of the numbers of SNPs. The LR increases

logarithmically with the increase in the number of SNPs. Since the best performing SNPs were included in the analysis first, above 20,000 SNPs the improvement of the LR is a bit more moderate. Furthermore, the results of the current laboratory analysis, as well as our previous theoretical analysis [1] show that about 3000 SNPs are sufficient for the vast majority of forensically relevant mixtures. Therefore, in this study we use 3000 SNPs for all of the following analyses. The purpose of this study is to evaluate the merit of a focused array for forensic purposes with about 3000 SNPs. In the absence of such technology the use of currently existent arrays with large numbers of SNPs should include all available SNPs in the analysis.

WGA to Address Low DNA Amounts

The recommended DNA input for genotyping with the HumanCytoSNP-12 chip is 200ng. In forensic samples, such a large quantity of DNA is not common. We therefore examined the use of WGA for mixtures with low starting amounts of DNA. The starting DNA quantities examined were 1.0ng, 5.0ng and 25ng of DNA. As expected, the allele drop-out rates increase as the amount of starting DNA quantity is lowered (Figure 4). In contrast, the drop-in rate didn't show a similarly direct correlation with the starting DNA amounts. We found an average drop-in rate of 11.9%, 1.6% and 6.6% for starting DNA quantities of 1ng, 5ng and 25ng, respectively. These values serve as the drop-in estimates in the LR equations. Despite the increase in drop-out rate, the analyses performed on low DNA amounts allowed a robust identification of the suspect when present in the mixture at reasonable proportions. For example, alleles of individuals contributing 10% of the DNA in a mixture of 5.0ng total DNA displayed a drop-out rate of 14.0%. The LR for such contributors ranged between 10^{58} and 10^{126}, thus robustly identifying them utilizing this approach.

Performance and statistics

Individuals contributing to the mixtures in Table 1 as well as 267 individuals also genotyped with the HumanCytoSNP-12 microarray in the HapMap project[8] from CEU, CHB, JPT and YRI populations were compared to the mixtures. The results of the comparison are summarized in Figure 5A.

All individuals contributing 15% or more to a mixture had a LR far above 1 (the minimum was $1.65 * 10^{9}$), even when the starting DNA quantity was only 1.0ng (Figure 5B). Likewise, when the starting DNA amount was 5.0ng or more, all individuals contributing at least 5% to the mixture resulted in a LR far above 1 (with the minimum of $1.43 * 10^{28}$). Out of the non-contributing CEU individuals analyzed (about 5500 comparisons in all mixtures and for all tested individuals) none displayed a LR above 1. When comparing individuals from CHB, JPT and YRI populations, in which allele frequencies are substantially different, less than 0.1% had a LR above 1 (Figure 5B) and none reached the minimum values presented above (the highest LR for a non-contributor was 2406).

We further tested the robustness of this approach by testing a more complex DNA mixture composed of 10 individuals, each comprising 10% of the mixture (Figure 6). When the starting DNA amount was high (200ng), the LR of all 10 contributors was high, ranging from 10^{17} to 10^{61}. When lower starting amounts of DNA subjected to WGA were utilized a majority of contributors had a LR > 10^{20}, 7 contributors when 5.0ng was the starting amount and 5 contributors when 1.0ng was the starting amount, with some higher than 10^{80}. For the most part this was due to extensive dropout of alleles from contributors 2, 3, and 4, possibly due to degradation or competition for reagents (*i.e.* stochastic effects) during the WGA process.

Regardless, this establishes that when recommended procedures are utilized, this approach can robustly identify individuals in complex mixtures containing up to 10 individuals. A focused forensic array containing only the SNPs we have selected here may negate the need for WGA and therefore give results more comparable to those generated here with 200ng of starting DNA but for far lower amounts of starting DNA (*i.e.* 1.0ng or less).

Inhibition/Degradation Studies

In forensic DNA samples, it is not uncommon for certain inhibitory compounds to be present due to the nature of the sample type from which the DNA was collected. For example, hematin in blood can be present in purified DNA from this sample type. Likewise, humic acid in soil and indigo dye from denim material can be present in forensic DNA samples collected from samples that have had contact with soil or that were deposited on denim material (*i.e.* jeans). . Additionally, it is not uncommon for forensic DNA to be highly degraded, resulting in poor STR amplification, and difficulty with interpretation due to the limited data that can be generated with these types of samples. Recently, STR amplification kits such as Identifiler® Plus (Applied Biosystems) and PowerPlex® 16 HS (Promega) have been developed in order to overcome the presence of both inhibitory compounds as well as result in improved amplification of degraded DNA. However, as previously discussed, for our purposes, current commercially available STR amplification kits are sufficient for interpretation of complex mixtures with more than three contributors or for cases where the contribution of the suspect is relatively low.

We therefore tested our approach with complex mixtures using various amounts of DNA either in the presence of various forensic sample type inhibitors (hematin, humic acid, and indigo dye) are with treatment resulting in degradation (Table 2), all resulting in some level of

inhibition of STR amplification. With the recommended starting DNA amount of 200ng (Figure 7A), when compared to the same mixtures that had been untreated all three inhibitors as well as DNA degradation had virtually no effect on the ability to robustly identify individuals in these mixtures with a LR of 10^{100} or higher when an individual contributed as little as 5% to the mixture and a LR of 10^{20} or higher when the contribution was only 2%. Lower starting DNA amounts gave slightly more variable results and were more dramatically affected by conditions of inhibition or degradation (Figure 7B-C), suggesting that these conditions affect WGA more than downstream SNP genotyping. The effect on the LR of each of the conditions tested, from least to most dramatic, were indigo dye > hematin >> humic acid > degradation. This in turn was dependent on the starting DNA amount. For example, when 5.0ng was the starting DNA amount, hematin resulted in a LR > 1 while humic acid resulted in a LR < 1for individuals who contributed 2% to the mixture. When 1.0ng was the starting DNA amount, all treatments except indigo dye resulted in a LR < 1even for individuals who contributed as much as 20% to the mixture. Indigo dye had the weakest effect, with high LR values for contributors of as little as 5% when as little as 1.0ng was tested. Degradation by far had the most severe effect, with a LR ~ 1for contributors of up to 20% when 1.0ng was the starting DNA amount. However, when 5.0ng was the starting DNA amount, a LR of 10^{25} was achieved for individuals who contributed 20% to the mixture and greater than 10^{100} for those who contributed 50% to the mixture. Similar to the results of the 10 person mixture, a focused forensic array may negate the need for WGA and therefore lessen or even eliminate the effects of inhibition or degradation observed in this study as well as result in LRs comparable to those generated here with 200ng of starting DNA for far lower amounts of starting DNA (*i.e.* 1.0ng).

Related Pair (Parent – Child) Studies

The possibility that a suspect's relative is contributing to a mixture introduces two problems:

3. The suspect may receive a high LR although he is not a real contributor to the mixture, due to the shared genetic material with a relative that is a contributor (corresponds to error type I).
4. A real contributor might claim he isn't part of the mixture, but was identified as a contributor because of the presence of his relative in the mixture (corresponds to error type II).

First examining Problem 1 (the suspect's relative is the true contributor to the mixture) we computed the likelihood ratio between our two classic hypotheses:

3. Prosecution hypothesis: The "suspect" and N-1 unknown individuals comprise the mixture.
4. Defense hypothesis: N unknown individuals comprise the mixture.

Comparing Parent - Child pairs in different mixtures (Table 3), relatives of real contributors receive a high LR and may be falsely included. Figure 7A shows that the log(LR) of a relative is linearly correlated with the log(LR) of the real contributor, with a coefficient of about 0.2-0.25. In order to minimize this problem, the hypothesis for the likelihood ratio calculation needed to be modified:

2. Prosecution hypothesis: The "suspect" and N-1 unknown individuals comprise the mixture.
4. Defense hypothesis: The "suspect's" relative and N-1unknown individuals comprise the mixture.

In the case of Parent-Child pairs the LR equations are modified as shown in Table 5. Since the drop-out rate of the hypothetical contributing relative cannot be calculated directly, the

drop-out rate was selected by numerically maximizing the likelihood of hypothesis **3** above. The new hypotheses pair (**1 vs 3**) lowers the LR of all contributors by 50% or more (Figure 9A, red crosses) but has a more dramatic effect for relatives (Figure 9A, blue circles), bringing the LR in most cases close to or below 1. Nevertheless, in a few cases the LR of a non-contributing relative remained high even with the new hypothesis, the highest result being $10^{8.42}$.

We can overcome this problem by reducing the maximum allowed drop-out rate from 0.5 to 0.3. The result is that all relatives received a LR below 1 (Figure 9B, blue circles). The downside of this solution is that low level contributors that do have a drop-out rate between 0.3-0.5 would receive a lower LR, and may be incorrectly excluded. For example, Figure 5B displays the LR of low contributors when the maximum drop-out rate is 0.5. When compared to a maximum drop-out rate of 0.3 to accommodate this problem with related pairs (Figure 10), the result was lower LRs for some contributors, including one that previously could not be excluded (maximum drop-out rate of 0.5) that was now clearly excluded (maximum drop-out rate of 0.3).

Next examining Problem 2, a suspect may claim that he is wrongly included because his relative is in the mixture, we again need to change the hypothesis pair from 1 vs. 2 to 1 vs. 3. As mentioned above the change in the hypothesis reduces the LR by at least 50%. Moreover, contributors that previously displayed a LR < 50 may now receive a negative value (Figure 9). When analyzing this data as a function of the starting DNA amount, a high LR for the inclusion of the suspect (as compared to his relative) was achieved for contributors of at least 5%, 10% and 20% in mixtures starting with 25.0ng, 5.0ng and 1.0ng total DNA respectively (Figure 10). By comparison, with our previous conditions the minimal contribution required for a high LR was 2%, 5% and 15% in mixtures starting with 25.0ng, 5.0ng and 1.0ng total DNA respectively (Figure 5B).

To summarize, we can ensure a non-contributing "suspect" is not wrongly included even if his relative is in a mixture. To do so, we need to change the defense hypothesis of the LR from hypothesis 2 to hypothesis 3, as indicated above, as well as lower the maximum allowed drop-out rate. Both of these changes lower the calculated LR of contributors meaning a higher proportion of a contributor in the mixture is required to robustly identify his presence. The change in the required proportion to ensure inclusion is not very large, so minor contributors can still be identified (depending on the total DNA amount used).

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Tables & Figures

Table 1. Mixture composition for General Studies, Displayed as Ratio of Contributor per Mixture (by Number of Contributors)*

		Number of Contributors					
		1	**2**	**3**	**4**	**5**	**6**
Mixtures	**1**	0.50	0.30	0.20			
	2	0.30	0.25	0.20	0.15	0.10	
	3	0.30	0.20	0.20	0.15	0.10	0.05
	4	0.35	0.35	0.15	0.05	0.05	0.05
	5	0.65	0.15	0.15	0.02	0.02	0.01
	6	0.63	0.10	0.10	0.10	0.05	0.02

*Numbers indicate the proportion that each individual contributed to each mixture.

Table 2. Mixture Composition for Inhibition, Degradation, and Related Pairs Studies, Displayed as Ratio of Contributor per Mixture (by Number of Contributors).

		Number of Contributors									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mixtures	**7**	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	8	0.50	0.30	0.20							
	9	0.66	0.10	0.10	0.05	0.05	0.02	0.02			

*Numbers indicate the proportion that each individual contributed to each mixture

Table 3. Mixtures for Inhibition, Degradation, and Related Pairs Studies, Displayed as Ratio of Contributor per Mixture (by Contributor).

		Contributor												
		1	**2**	**3**	**4**	**5 (P1)**	**6**	**7**	**8**	**9**	**10**	**11 (P2)**	**12 (C1)**	**13 (C2)**
Mixtures	**7**	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10			
	8			0.30		0.50						0.20		
	9		0.02	0.10		0.66			0.02	0.05			0.10	0.05

*Numbers indicate the proportion that each individual contributed to each mixture

*** P1 = Parent 1, P2 = Parent 2, C1 = Child 1, C2 = Child 2*

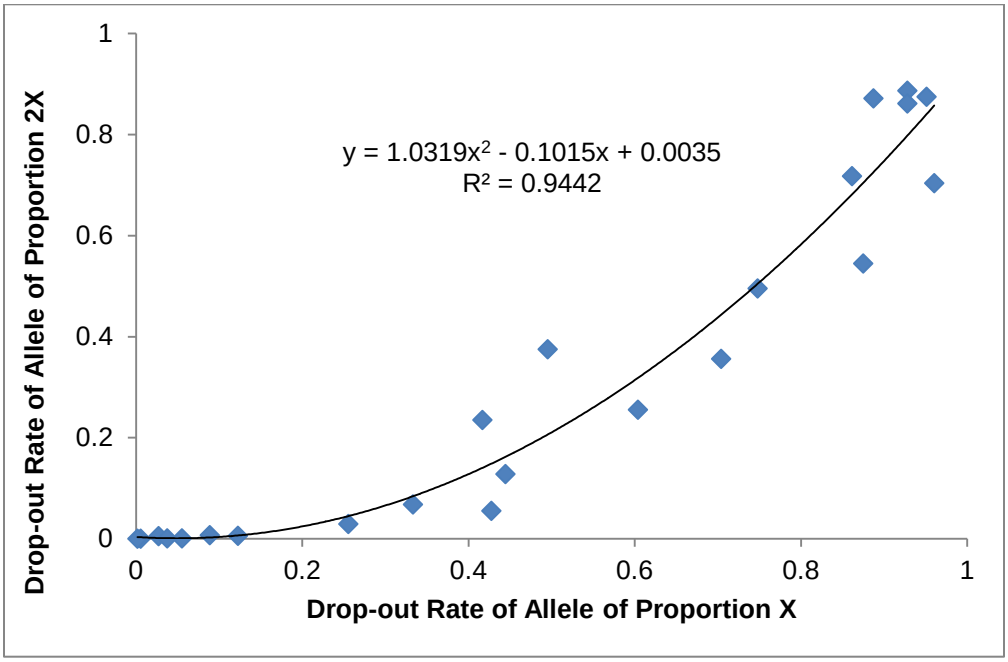
Table 4. Likelihood ratio (LR) formulas.

Mixture type / Suspect genotype	AA	AB
AA	$\frac{a^{2(N-\overline{D})} \cdot \overline{c}}{a^{2N} \cdot \overline{c}} = \frac{1}{a^{2\overline{D}}}$	$\frac{1 - a^{2(N-\overline{D})} \cdot \overline{c}}{1 - a^{2N} \cdot \overline{c}}$
AB	$\frac{a^{2(N-\overline{D})} \cdot D \cdot \overline{c}}{a^{2N} \cdot \overline{c}} = \frac{D}{a^{2\overline{D}}}$	$\frac{1 - a^{2(N-\overline{D})} \cdot D \cdot \overline{c}}{1 - a^{2N} \cdot \overline{c}}$
BB	$\frac{a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c}}{a^{2N} \cdot \overline{c}} = \frac{D_h}{a^{2\overline{D}}}$	$\frac{1 - a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c}}{1 - a^{2N} \cdot \overline{c}}$

Table 5. Likelihood ratio (LR) formulas for Related Pair Study.

Mixture type / Suspect genotype	AA	AB
AA	$\frac{a^{2(N-\overline{D})} \cdot \overline{c}}{a^{2(N-\overline{D})} \cdot \overline{c} \cdot (1 - \overline{D} \cdot b)}$	$\frac{1 - a^{2(N-\overline{D})} \cdot \overline{c}}{1 - a^{2(N-\overline{D})} \cdot \overline{c} \cdot (1 - \overline{D} \cdot b)}$
AB	$\frac{a^{2(N-\overline{D})} \cdot D \cdot \overline{c}}{a^{2(N-\overline{D})} \cdot \overline{c} \cdot \frac{1}{2}(D + a + D_h \cdot b)}$	$\frac{1 - a^{2(N-\overline{D})} \cdot D \cdot \overline{c}}{1 - a^{2(N-\overline{D})} \cdot \overline{c} \cdot (1 - \overline{D} \cdot b)}$
BB	$\frac{a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c}}{a^{2(N-\overline{D})} \cdot \overline{c} \cdot (D \cdot a + D_h \cdot b)}$	$\frac{1 - a^{2(N-\overline{D})} \cdot D_h \cdot \overline{c}}{1 - a^{2(N-\overline{D})} \cdot \overline{c} \cdot (D \cdot a + D_h \cdot b)}$

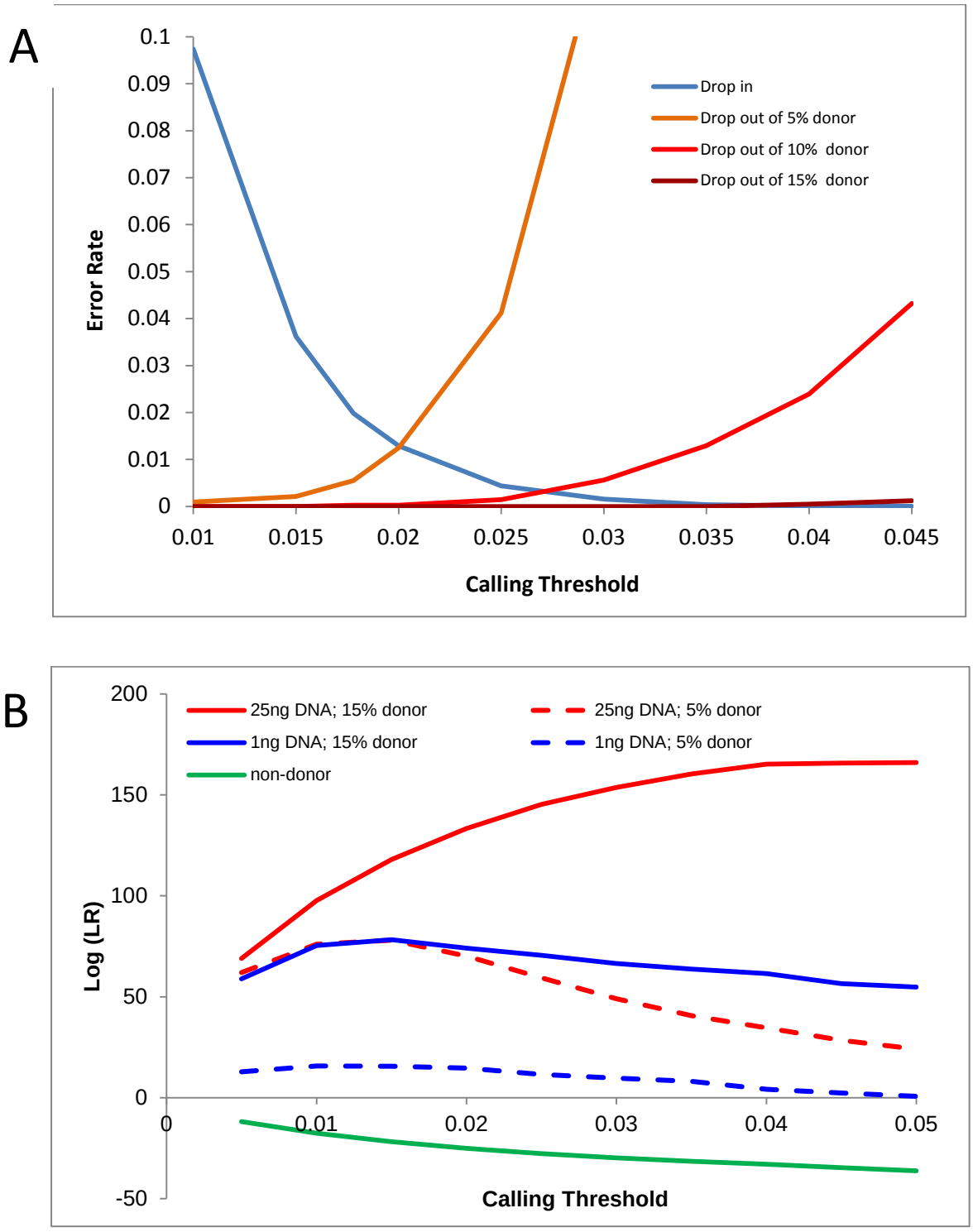
Figure 1. Correlation between the drop-out rate of an allele of a certain proportion X in the mixture and of an allele of twice that proportion in the mixture.



Each point represents the average drop-out rates of alleles present in a pair of proportions such as 0.01-0.02, 0.05-0.1, etc. in all mixtures with the same starting DNA amount (1.0ng, 5.0ng, 25ng or 200ng).

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

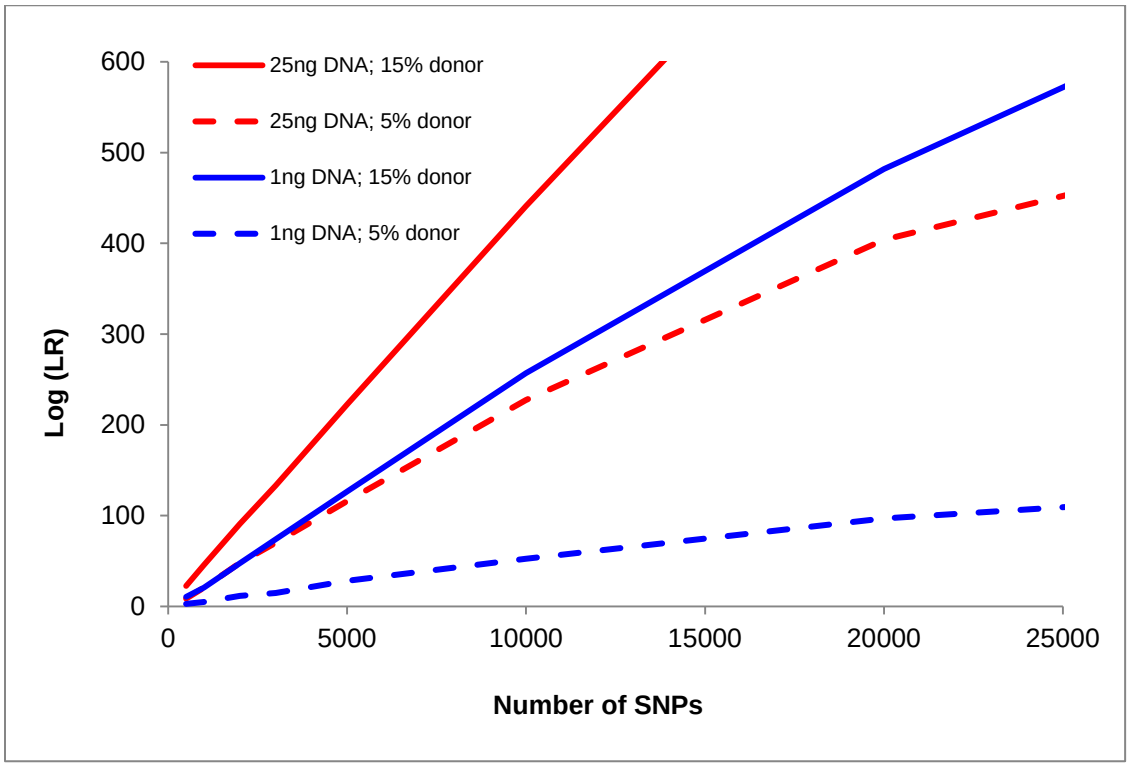
Figure 2. Establishment of a Calling Threshold



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

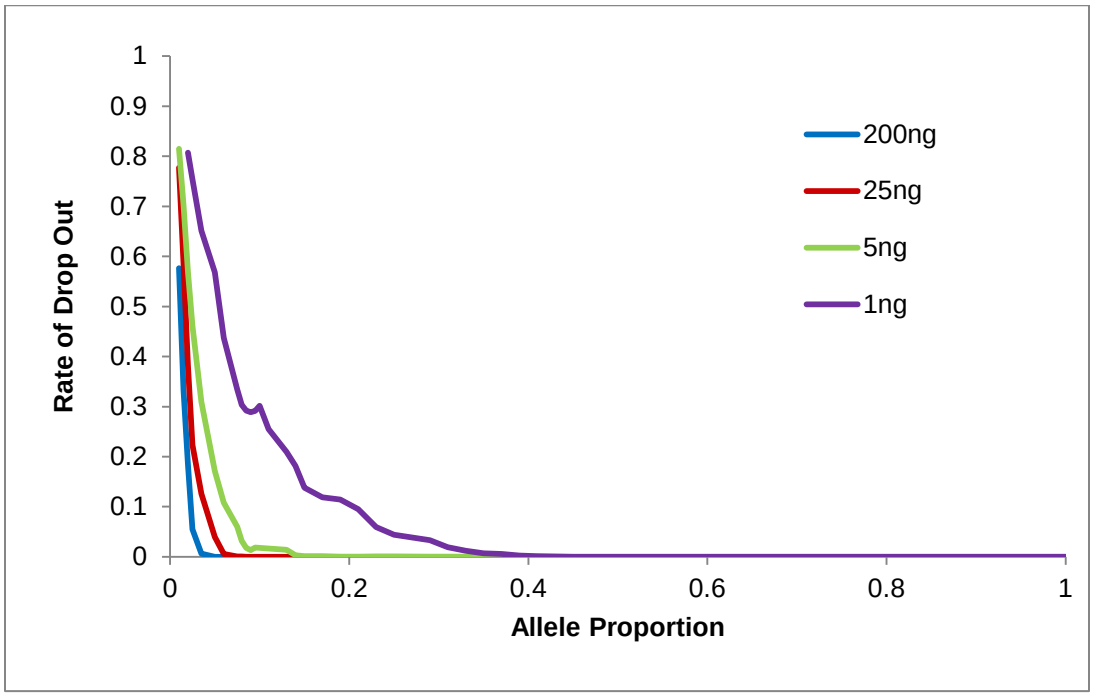
A. Drop-in (blue line) and drop-out error rates (orange, red and purple lines) as a function of calling threshold. For the drop-out rate analysis, three levels of contribution to the mixture were considered: a suspect contributing 5.0%, 10% or 15% of the DNA in the mixture (orange, red or purple lines, respectively). Note that due to heterozygosity this corresponds to a contribution of 2.5%, 5.0% and 7.5% of the analyzed allele. **B.** Average log(LR) as a function of the calling threshold. We present 25ng or 1.0ng (red or blue lines respectively) of initial DNA amount (all underwent WGA) for a donor contribution of 5.0% or 15% (dashed or solid lines, respectively). The case of a non-donor is also presented (green line).

Figure 3. Likelihood Ratio as a Function of the Number of SNPs



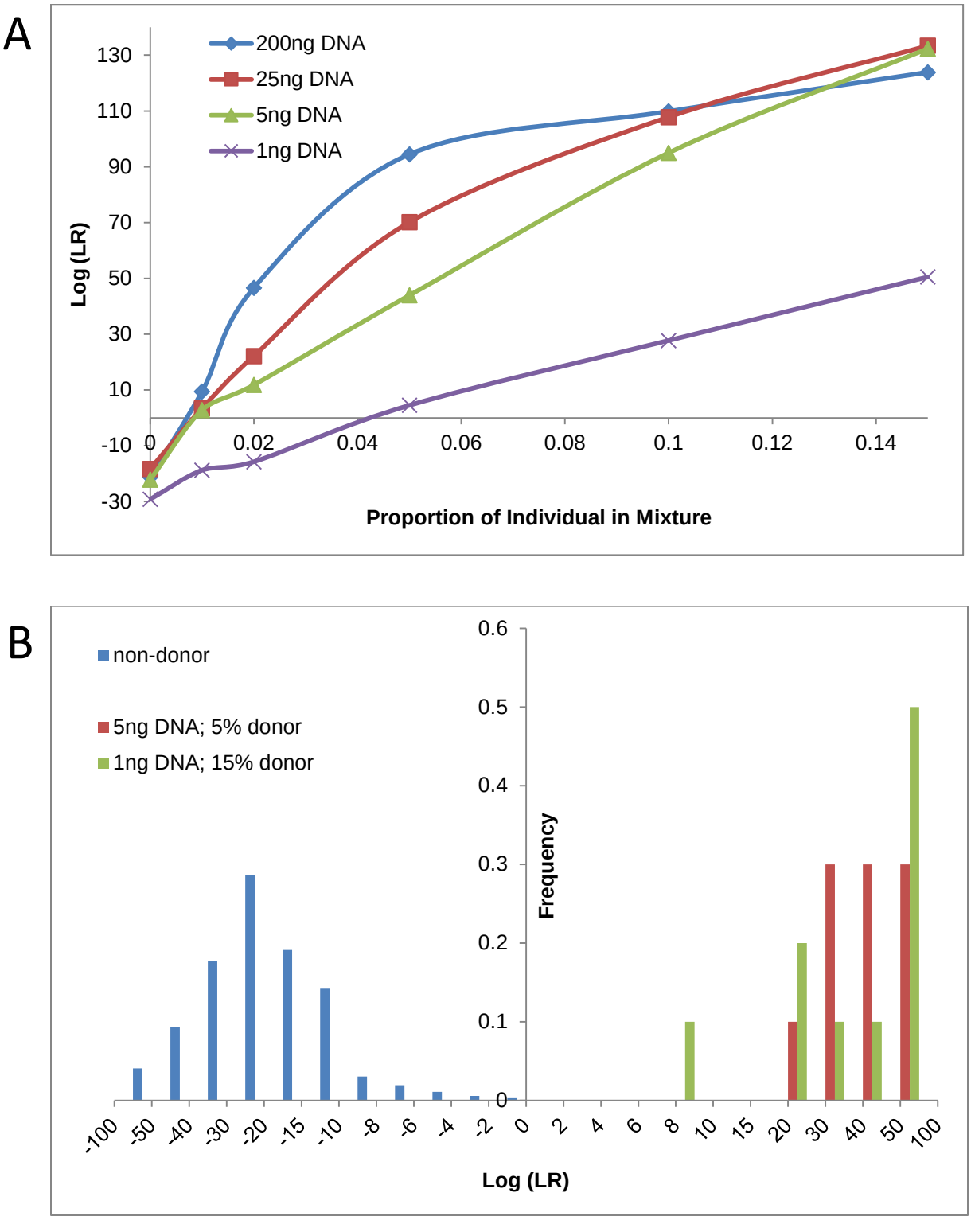
Average log(LR) of contributors of 5.0% or 15% (dashed or solid lines, respectively) out of a mixture with an initial DNA amount of 1.0ng or 25ng (blue or red lines respectively). All mixtures underwent WGA. The calling threshold is set to 0.02. The SNPs are ranked along the X-axis according to their performance score (see SNP selection in Materials and Methods).

Figure 4. Drop-out Rate Distribution at Various Starting DNA Quantities



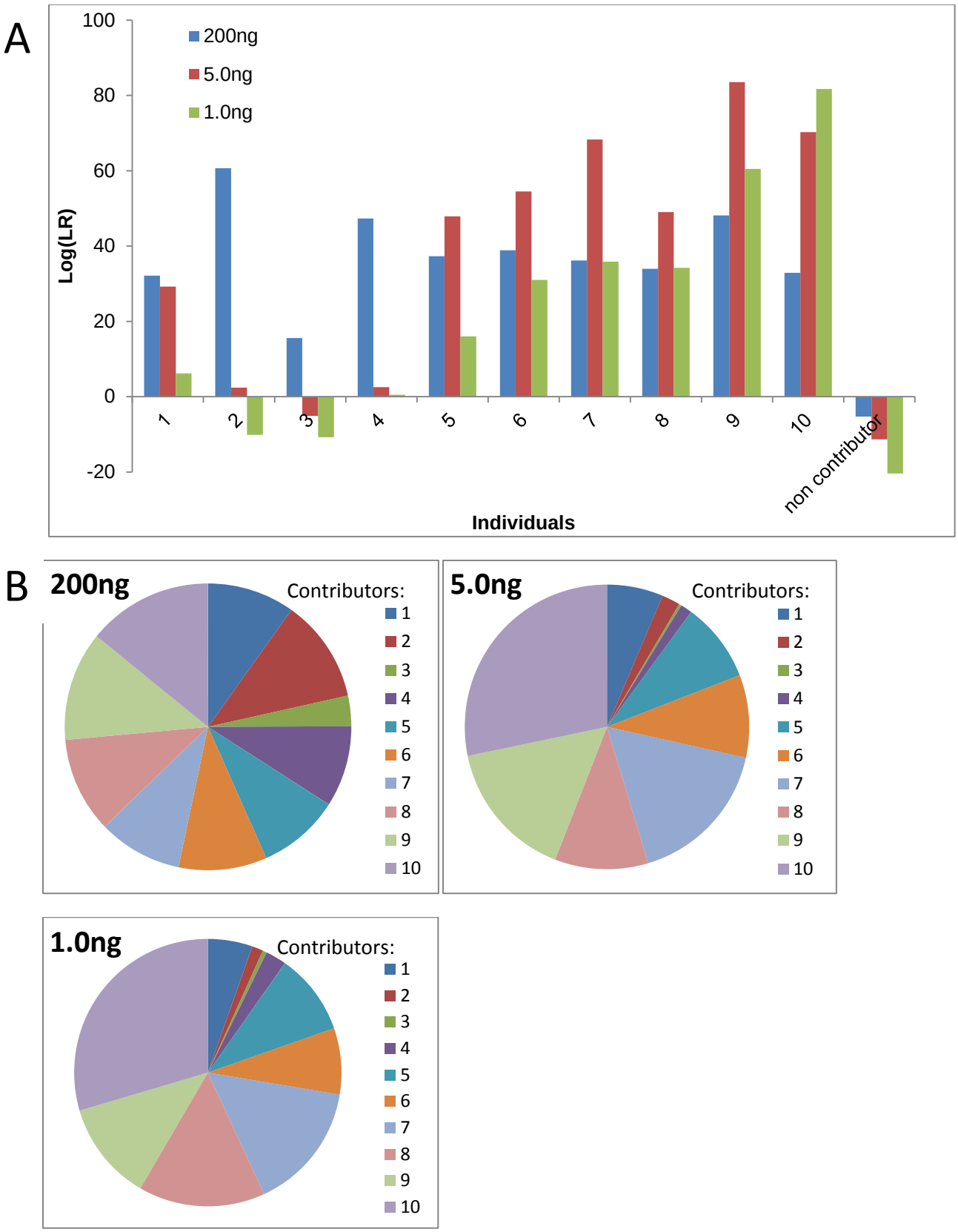
The moving average rate of allele drop-out as a function of the allele proportion in the mixture. Different lines represent different initial DNA amounts, 200ng, 25ng, 5.0ng and 1.0ng. Mixtures containing DNA amounts of 25ng or less underwent WGA. Allele proportion is based both on the known genotypes and the proportions of the contributors. For example, an allele proportion of 0.4 could be a result of a mixture of 5 contributors with proportions 0.4: 0.2: 0.2: 0.1: 0.1 and corresponding genotypes AB:AB:AB:AA:AA making the DNA copies containing the B allele 40% of the total number of copies of this locus.

Figure 5. The Effect of Donor Proportion on LR



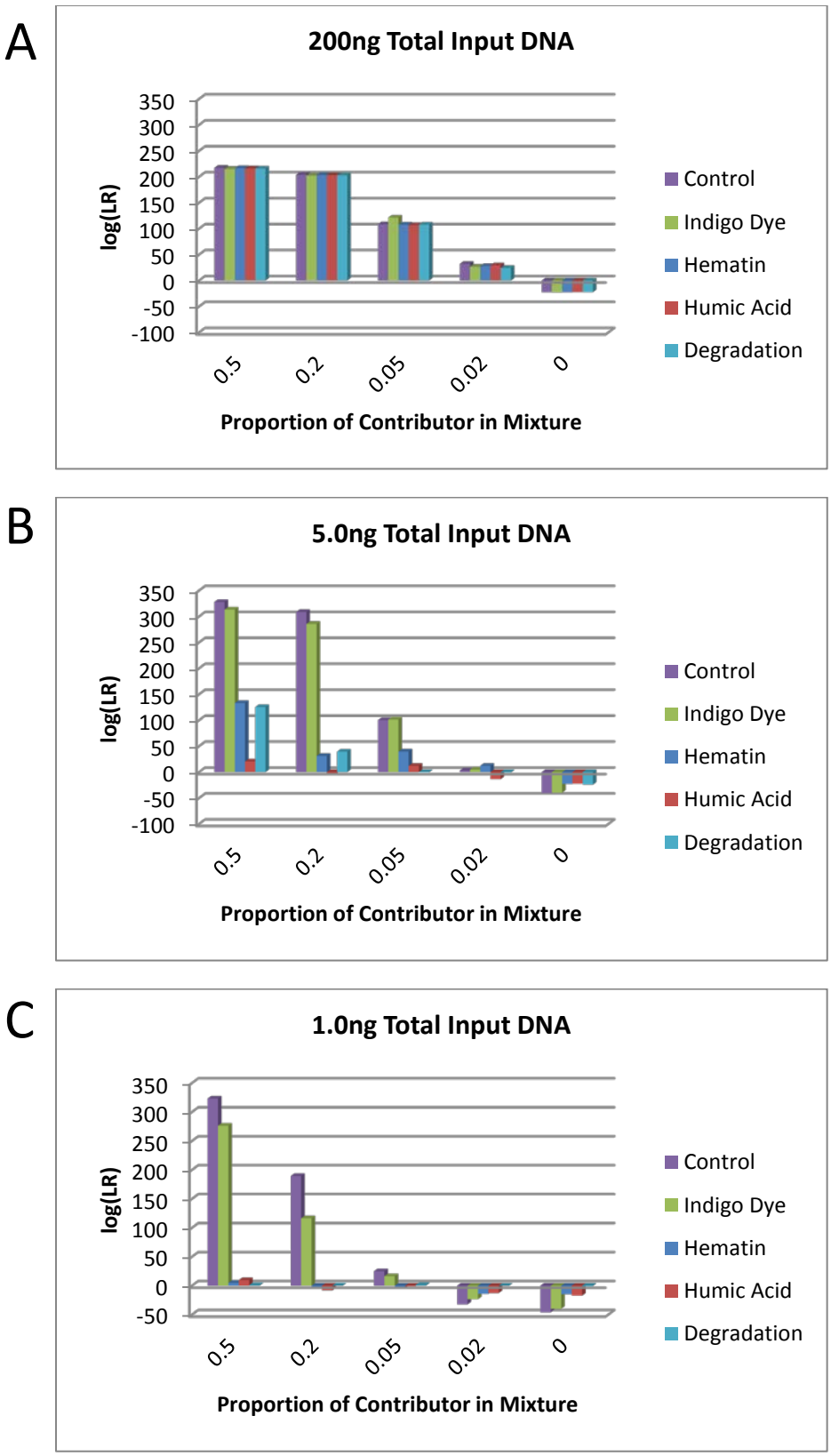
A. Average log(LR) as a function of donor proportion in the mixture. Different lines represent different initial DNA amounts, 200ng, 25ng, 5.0ng and 1.0ng. Mixtures containing DNA amounts of 25ng or less underwent WGA. **B.** Distribution of log(LR) of non-donors from all four considered populations ,CEU, CHB, JPT and YRI (Blue bars), 5.0% donors to mixtures of 5.0ng initial DNA amount (red bars) and 15% donors to mixtures of 1.0ng initial DNA amount (green bars). Mixtures containing DNA amounts of 5.0ng and 1.0ng underwent WGA.

Figure 6. Analysis of a Complex DNA Mixture Containing 10 Individuals



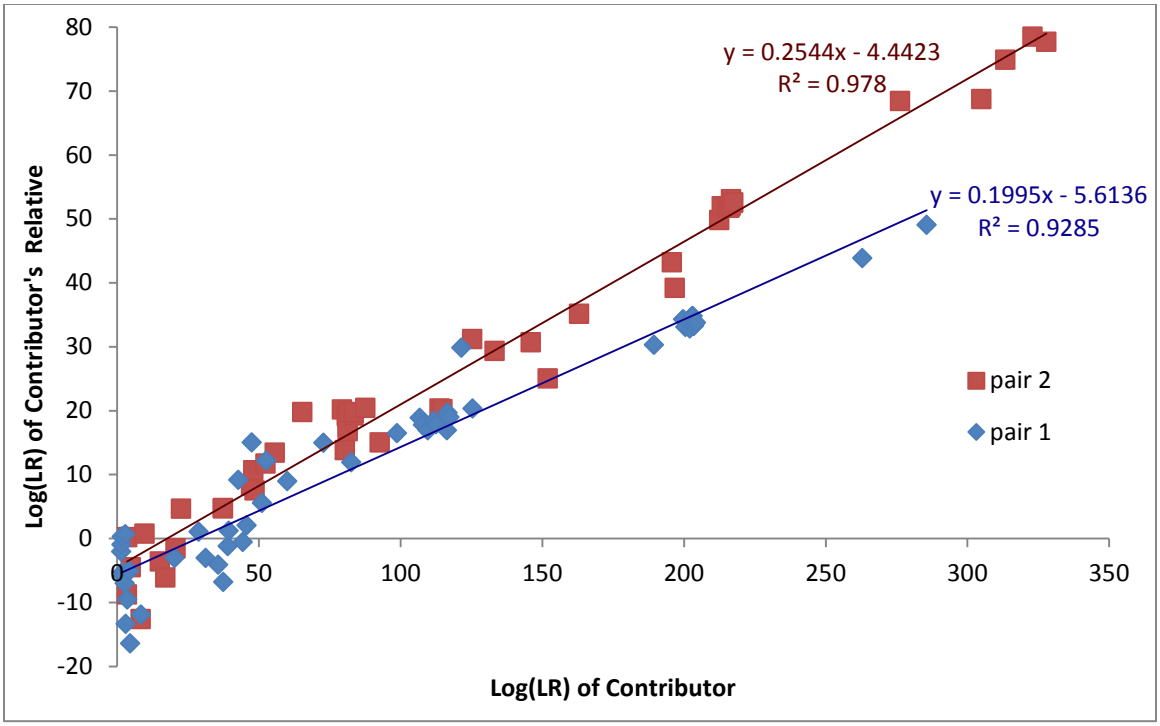
A. Log(LR) for each of 10 individuals in a complex mixture, where each individual contributed 10% of the total amount of starting DNA (200ng – blue, 5.0ng – red, 1.0ng – green) as well as a non-contributor. Mixtures containing DNA amounts of 5.0 and 1.0ng underwent WGA and all inhibitor or degradation treatments were performed prior to WGA. **B.** Pie charts displaying the intensity of alleles of each individual in each mixture at the different starting DNA amounts. For example, when 200ng was the starting DNA amount, each of the 10 contributors displayed relatively equal intensities of allelic data.

Figure 7. Effect of Common Forensic Sample Type Inhibitors & DNA Degradation



Log(LR) as a function of donor proportion in the mixture (0: Non-Contributor). Different plots represent different initial DNA amounts; A) 200ng, B) 5.0ng and C) 1.0ng total input DNA, where each column represents either presence of a common forensic sample type inhibitor (Indigo Dye, Hematin, Humic Acid), DNA degradation, or absence of any treatment (Control). Mixtures containing DNA amounts of 5.0 and 1.0ng underwent WGA and all inhibitor or degradation treatments were performed prior to WGA.

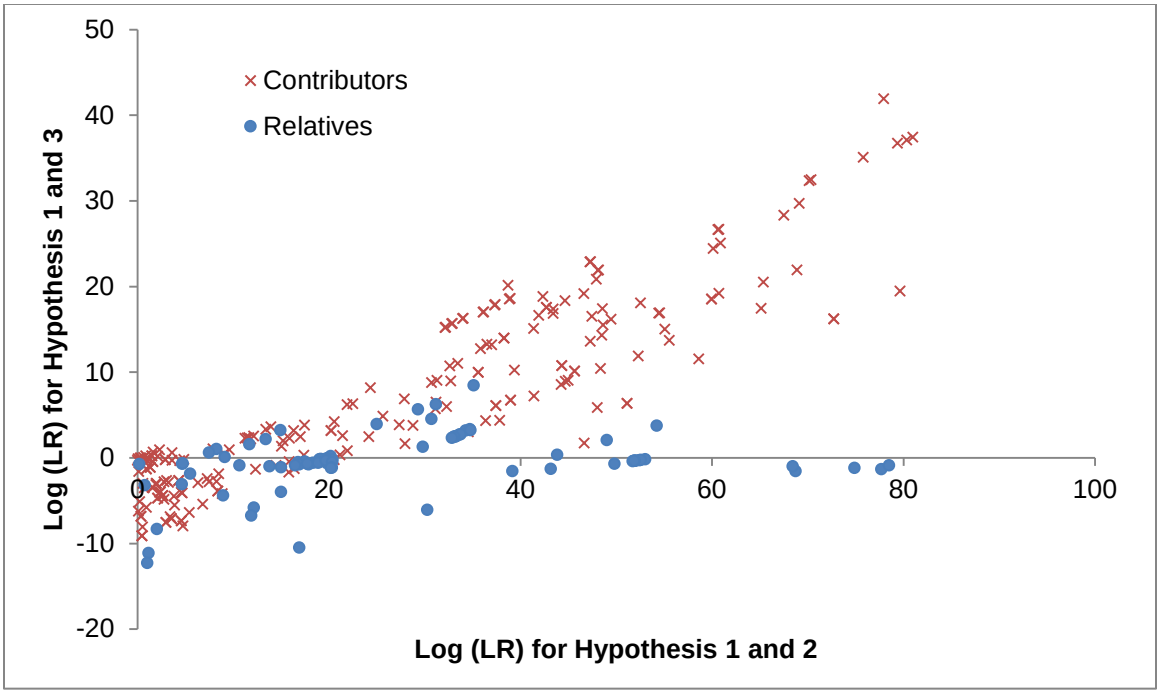
Figure 8. Likelihood Ratios of Contributors Compared to Contributors Relative



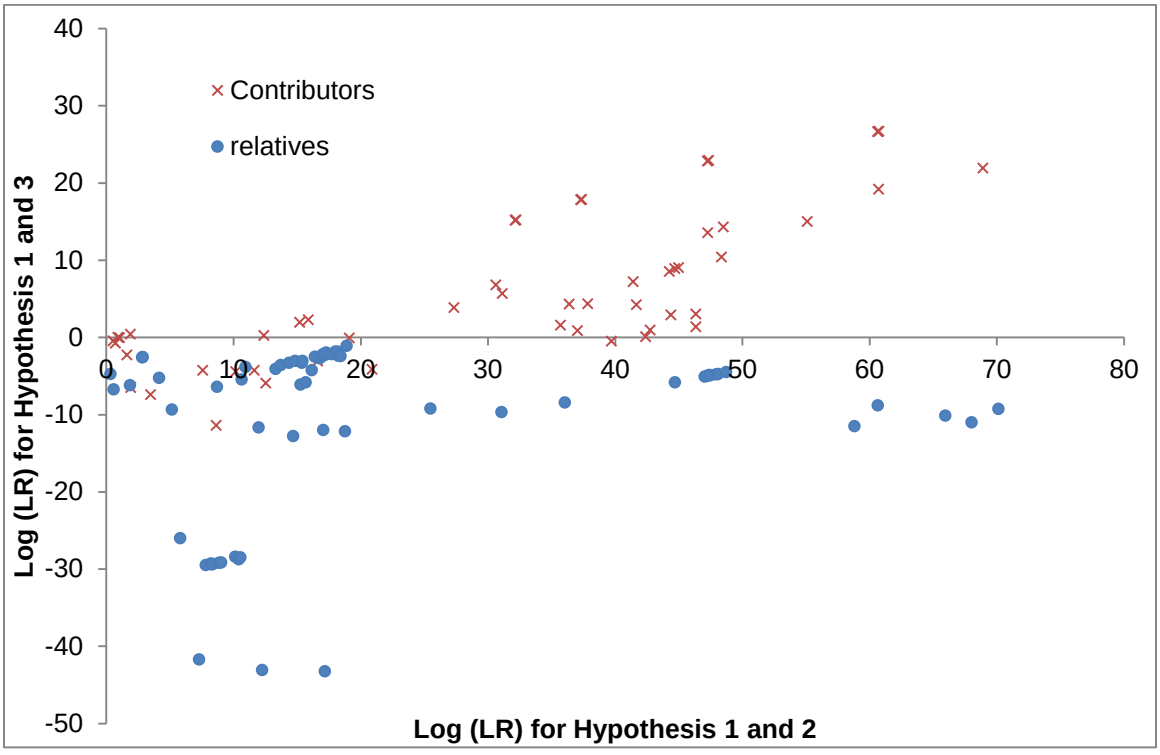
Likelihood ratio calculations for two different related pairs (parent – child) utilizing the original hypotheses (1 versus 2). The x-axis displays the LR of the real contributor and the y-axis displays the LR of the contributor's relative in the same mixture.

Figure 9. Change in Likelihood Ratio Between Hypothesis Pairs

A

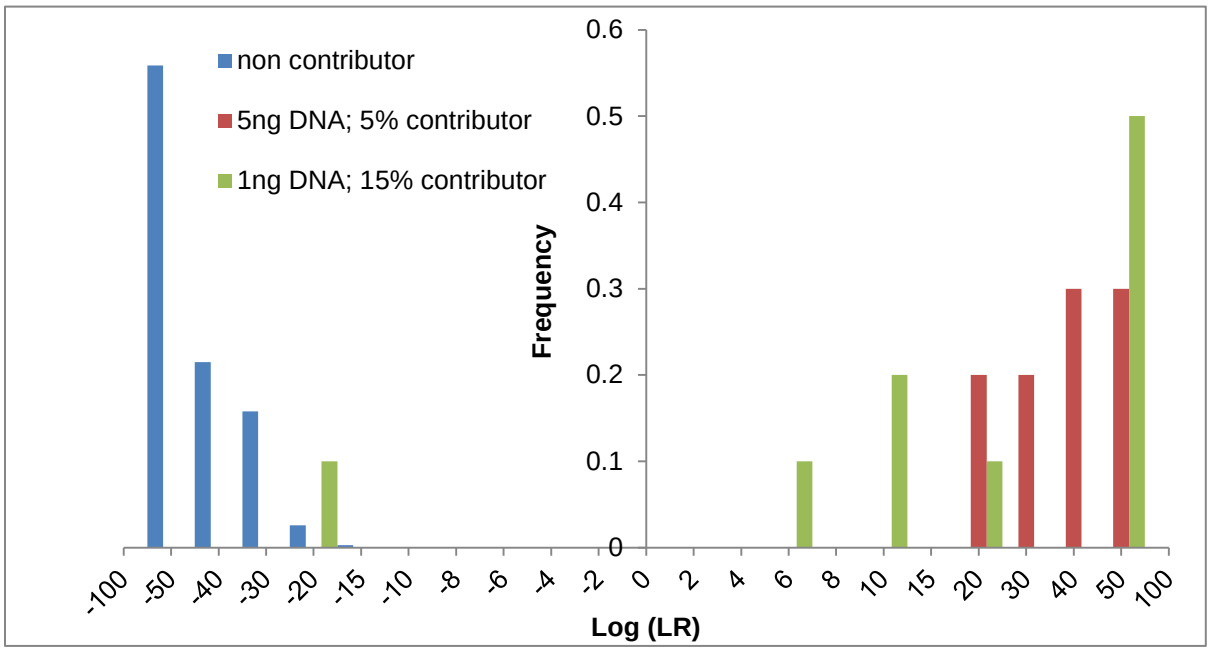


B



Comparison of the log(LR) comparing hypotheses pairs 1 & 2 with 1 & 3 for related pairs comparing two different parent-child pairs in mixtures 8 & 9 from Table 3. Likelihood ratio calculations were performed using the equations in Table 4 (hypotheses pairs 1 & 2) and Table 5 (hypotheses pairs 1 & 3). Likelihood ratios of suspects that truly contribute to the mixture are indicated by red crosses whereas those for suspects that have a relative present in the mixture but do not contribute themselves are indicated by blue circles. **A.** Likelihood ratio calculations were performed with a maximum drop-out rate of 0.5. **B.** Likelihood ratio calculations were performed with a maximum drop-out rate of 0.3.

Figure 10. The Effect of Donor Proportion on LR for Related Pairs



Distribution of log(LR) of non-donors from all four considered populations ,CEU, CHB, JPT and YRI (Blue bars), 5.0% donors to mixtures of 5.0ng initial DNA amount (red bars) and 15% donors to mixtures of 1.0ng initial DNA amount (green bars). Mixtures containing DNA amounts of 5.0ng and 1.0ng underwent WGA. Likelihood ratios were calculated utilizing hypotheses pair 1 & 3 for related pairs (Table 5) and a maximum drop-out rate of 0.3.

Conclusions

Discussion of Findings

The presence of complex DNA mixtures is common in forensic cases. When analyzed with standard STRs, results are mostly inconclusive or not admissible in court. We have previously suggested an alternative method using a few thousand SNPs to accurately determine the presence or absence of an individual in a complex DNA mixture found at the scene of crime. In the current study, using standard genotyping technologies and off-the-shelf SNP microarrays we have shown the practical applicability of this method. Our results show in general that this technology is accurate and meets the strict requirements of a forensic setting. Additionally, we have shown that this technology has a relatively high sensitivity that allows the detection of an individual that contributes as little as 5% of the total DNA in a forensic mixture.

Furthermore, we presented a thorough analysis utilizing the widely accepted likelihood ratio (LR) approach. Whereas the RMNE approach is more conservative the LR approach has higher statistical power, thereby providing very strong support for the presence of an individual in a mixture even when the relative contribution to the mixture was small and the number of contributors was high. Furthermore, non-contributors rarely exhibited a LR>1. In this study we described mixtures of up to ten contributing individuals. However, the parameter that primarily affects performance is the proportion of DNA contributed by the suspect. The effect of the number of contributors does not affect accuracy of the typing methods but does in fact have a statistical effect as previously described [1].

The recommended amount of total DNA for genotyping with the HumanCytoSNP-12 micorarray is 200ng. In most forensic cases such an amount of DNA is typically not available.

In order to analyze samples with lower amounts of DNA that more accurately reflect that present in typical forensic samples we used WGA to amplify DNA mixtures with 1.0ng, 5.0ng and 25ng of total starting DNA. Even in mixtures with only 1.0ng of total starting DNA, a robust inclusion of contributors was observed as long as they contributed 15% (only 150pg) or more to the DNA mixture. With higher DNA starting amounts, individuals contributing only 5% to the mixture were detected. This is of particular importance in some forensic cases. For example, often in sexual assault cases the amount of total DNA may be high while the individual of interest represents a very minor contribution to the mixture. We found that when 25ng of DNA is used (as the starting amount), it is possible to detect individuals with high certainty even if their contribution to the mixture was as little as 2%. Furthermore, when the recommended starting DNA amount was utilized, we were able to robustly identify all ten individuals in a single mixture where each individual contributed only 10% to the mixture. With lower starting amounts of DNA the result was more variable, only indentifying a slight majority of the contributors, but may in part have been due to the requirement of WGA to achieve the recommended starting DNA amount. Finally, with slight alterations to our likelihood ratio calculations we were able to ensure a non-contributing "suspect" is not wrongly included even if his relative is in a mixture.

We also tested our approach's ability to identify individuals in complex mixtures in the presence of inhibitors, such as hematin and humic acid, as well as when DNA has been highly degraded. Using the recommended amount of starting DNA (200ng) we could robustly identify individuals who comprised as little as 2% of the mixture. With lower starting amounts of DNA, the effects of inhibitors and degradation were more dramatic, with only individuals contributing much higher percentages (20% or greater for 5.0ng, 50% or greater for 1.0ng) to the mixtures

able to be identified with high likelihood ratios. Since the negative effects of inhibition and degradation were more dramatic with lower starting amounts of DNA, this suggests that these conditions affect WGA more than downstream SNP genotyping. An alternative WGA kit designed for formalin fixed – paraffin embedded (FFPE) tissue may perform better in this aspect and may be tested in future studies.

Based on our results we believe that the current method can be implemented in forensic laboratories relatively quickly. We used a commercially available DNA microarray that contains about 300,000 SNPs. However in the analyses we considered only 3000 of them. The immediate use of commercially available microarray as shown here has the advantage of additional statistical power over current STR technology due to the large number of SNPs since more SNPs already present on these microarrays could be added to the analysis and therefore overcome some of the aforementioned hurdles found in our studies (see Figure 3).

Implications for Policy and Practice

SNP genotyping with current commercially available microarrays, such as the HumanCytoSNP-12 (Illumina®) microarray utilized in our studies, enables robust identification of individuals minimally contributing to complex DNA mixtures that include up to 10 contributors. Relatively low amounts of DNA starting material (down to 1.0ng) are efficiently amplified with standard WGA procedures using currently available kits and can be genotyped accurately on this microarray. Using our statistical methods, forensic labs could immediately improve their ability to identify individuals in complex mixtures in a cost effective manner as compared to currently available STR amplification kits and procedures. This could have the

largest impact on identification of individuals on touch evidence or mass disasters, where a high number of individuals could be present.

Implications for Further Research

In partnership with Illumina®, development of a focused forensic SNP microarray based on the Infinium HD chemistry utilized by the HumanCytoSNP-12 microarray that was used in our research and containing only the 3000 SNPs we used for our analysis is being planned for further research. Using a smaller microarray, the required starting DNA amount would be greatly lowered (2.0ng or less) thereby lessening or eliminating the hurdles discovered in our study. Since most of the problems uncovered related to the utilization of whole genome amplification, such as the effect of inhibitors on WGA, it is important that an alternative be established in order to eliminate this need. For example, when the recommended amount of starting DNA was used in our research (200ng) individuals were robustly identified in the presence of all inhibitors when they comprised as little as 2% of the mixture; with a focused forensic microarray this same result could be accomplished with as little as 2.0ng of total DNA where an individual contributing 2.0ng only contributes 40pg. Alternatively, if the HumanCytoSNP-12 microarray were continued to be utilized, a different whole genome amplification kit designed for use with more difficult samples (*i.e.* FFPE tissue) may be tested that may better alleviate these effects.

References

1. Voskoboinik, L. and A. Darvasi, *Forensic identification of an individual in complex DNA mixtures.* Forensic Sci Int Genet, 2011. **5**(5): p. 428-35.
2. Benschop, C., H. Haned, and T. Sijen, *Consensus and pool profiles to assist in the analysis and interpretation of complex low template DNA mixtures.* Int J Legal Med, 2011.
3. Carracedo, A., et al., *Focus issue-Analysis and biostatistical interpretation of complex and low template DNA samples.* Forensic Sci Int Genet, 2012.
4. Gill, P., et al., *Interpretation of complex DNA profiles using empirical models and a method to measure their robustness.* Forensic Sci Int Genet, 2008. **2**(2): p. 91-103.
5. Homer, N., et al., *Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays.* PLoS Genet, 2008. **4**(8): p. e1000167.
6. Egeland, T., et al., *Complex mixtures: a critical examination of a paper by Homer et al.* Forensic Sci Int Genet, 2012. **6**(1): p. 64-9.
7. Sean X. C. , J.S.L., *Statistical Applications of the Poisson-Binomial and conditional Bernoulli distribution.* Statistica Sinica, 1997. **7**: p. 875-892.
8. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
9. Zeng, Z., et al., *Evaluation of 96 SNPs in 14 populations for worldwide individual identification.* J Forensic Sci, 2012. **57**(4): p. 1031-5.
10. Fondevila, M., et al., *Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies.* Forensic Sci Int Genet, 2012.
11. Ruiz, Y., et al., *Further development of forensic eye color predictive tests.* Forensic Sci Int Genet, 2012.
12. Walsh, S., et al., *Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence.* Forensic Sci Int Genet, 2011. **5**(5): p. 464-71.

13. Branicki, W., U. Brudnik, and A. Wojas-Pelc, *Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype.* Ann Hum Genet, 2009. **73**(2): p. 160-70.

14. Matise, T.C., Chen F., Chen W., De La Vega F.M., Hansen M., He C., et al., *A second-generation combined linkage physical map of the human genome*. Genome Res. 17 (2007) 1783-1786.

Dissemination of Research Findings

The findings from this research grant have been disseminated to the NIJ by virtue of this Final Technical Report. To date, they have not been disseminated elsewhere, however a manuscript is currently in revision stages for submission to FSI:Genetics. In addition, a poster is in preparation for the 24th Annual Promega International Symposium on Human Identification.