# Identifying common donors in DNA mixtures

Klaas Slooten

Netherlands Forensic Institute, VU University Amsterdam

Budapest, English Speaking Working Group of the ISFG, 2 September 2016

# Mixture evaluation

## Likelihood ratio

To evaluate whether there is evidence that a PoI $S$ with genotype $g$ has contributed to a mixture $M$, one evaluates the likelihood ratio

$$LR(M, g) = \frac{P(M \mid S = g, H_p)}{P(M \mid S = g, H_d)},$$

where

- $M$ are the mixture data that are evaluated
- $H_p$ states that $S$ is a contributor
- $H_d$ replaces $S$ by an unknown contributor

# Models

## Likelihoods

A probabilistic model is needed to compute $P(M \mid H_*)$.

## Model types

- Binary/semi-continous models consider $M$ to be the set of recorded alleles, allowing for dropout and drop-in
- Continuous models consider $M$ to be the recorded alleles and their peak heights.

Continuous models treat more data, hence need a more sophisticated probabilistic model.

# Parameter choices

## Semi-continuous

Parameters are:

- number of contributors $n$
- Dropout probability $d_1, \ldots, d_n$ per donor
- Drop-in parameter $c$

Parameters can be chosen in several ways

- Based on the mixture data, and uncontested contributors; then used for both hypotheses
- For both hypotheses separately (e.g. using Maximum Likelihood estimates)

# Likelihood ratio versus deconvolution

If we take the same model under both hypotheses, then

$$\begin{aligned} LR(M, g) &= \frac{P(M \mid S = g, H_p)}{P(M \mid S = g, H_d)} \\ &= \frac{P(S = g \mid M, H_p)}{P(S = g \mid M, H_d)}. \end{aligned}$$

## LR interpretations

The LR says:

- How much more probable the mixture data are when the PoI is supposed to be a donor than if not
- How much more likely the PoI is to have the genotype $g$ if he is a donor than if not

Without $\theta$-correction or relatedness, $P(S = g \mid M, H_d) = p_g$ (population frequency). We continue with this assumption.

# Consequences

## LR never more than for single source

$$LR(M,g)p_g = P(S = g \mid M, H_p) \leq 1 \Rightarrow LR(M,g) \leq \frac{1}{p_g}.$$

## LR distributions

Based on the mixture data we can

- Calculate the LR for every genotype
- Equivalently, calculate the probability of every genotype to be that of the searched donor
- Obtain LR-distributions both for $H_p$ and $H_d$ (for power calculations)

See also K. Slooten, T. Egeland, *Likelihood ratios and exclusion probabilities with applications to mixtures*, Int. J. Legal Medicine 130, 2016

# MixKin

Calculations have been carried out with Mathematica script MixKin.

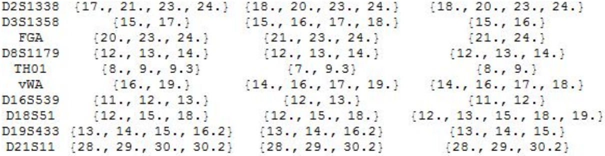
- Dropout probability per donor
- LR calculations
- Relatedness under $H_p$ (familial searching) or $H_d$ (discriminate from relative)
- LR distributions to see expected LR's for non-donors, relatives of donors and donors
- Donor genotype probability distribution
- ROC curves

See also

- K. Slooten, *Familial Searching on DNA mixtures with dropout*, Forensic Science International: Genetics **22**, 128–138, 2016
- K. Slooten, *Discriminating between donors and their relatives in complex DNA mixtures*, Forensic Science International: Genetics **21**, 95–109, 2015

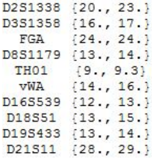
# Example

Mixture with dropout $(d_1, d_2, d_3) = (0.2, 0.3, 0.5)$ and $c = 0.05$:



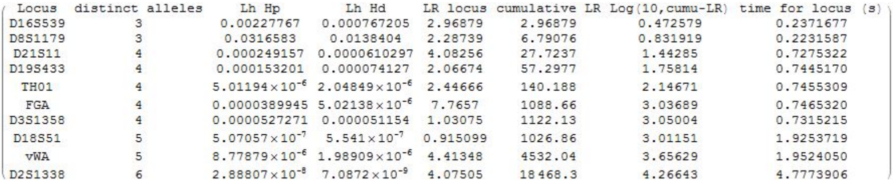
| | | | |
|---|---|---|---|
| D2S1338 | {17., 21., 23., 24.} | {18., 20., 23., 24.} | {18., 20., 23., 24.} |
| D3S1358 | {15., 17.} | {15., 16., 17., 18.} | {15., 16.} |
| FGA | {20., 23., 24.} | {21., 23., 24.} | {21., 24.} |
| D8S1179 | {12., 13., 14.} | {12., 13., 14.} | {12., 13., 14.} |
| TH01 | {8., 9., 9.3} | {7., 9.3} | {8., 9.} |
| vWA | {16., 19.} | {14., 16., 17., 19.} | {14., 16., 17., 18.} |
| D16S539 | {11., 12., 13.} | {12., 13.} | {11., 12.} |
| D18S51 | {12., 15., 18.} | {12., 15., 18.} | {12., 13., 15., 18., 19.} |
| D19S433 | {13., 14., 15., 16.2} | {13., 14., 16.2} | {13., 14., 15.} |
| D21S11 | {28., 29., 30., 30.2} | {28., 29., 30.2} | {28., 29., 30.2} |

Donor 1 (with $d_1 = 0.2$):

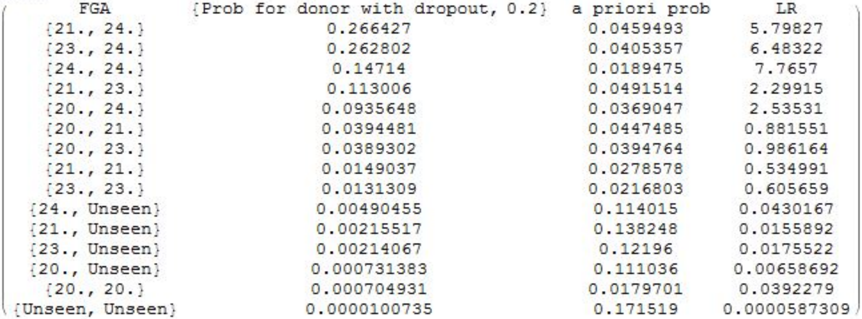


| | |
|---|---|
| D2S1338 | {20., 23.} |
| D3S1358 | {16., 17.} |
| FGA | {24., 24.} |
| D8S1179 | {13., 14.} |
| TH01 | {9., 9.3} |
| vWA | {14., 16.} |
| D16S539 | {12., 13.} |
| D18S51 | {13., 15.} |
| D19S433 | {13., 14.} |
| D21S11 | {28., 29.} |

# FGA: donor is (24, 24)

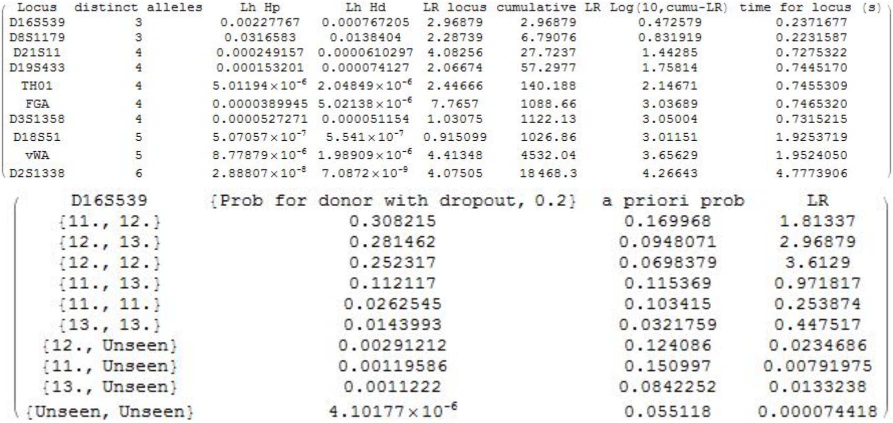


| Locus | distinct alleles | Lh Hp | Lh Hd | LR locus | cumulative LR | Log(10,cumu-LR) | time for locus (s) |
|---|---|---|---|---|---|---|---|
| D16S539 | 3 | 0.00227767 | 0.000767205 | 2.96879 | 2.96879 | 0.472579 | 0.2371677 |
| D8S1179 | 3 | 0.0316583 | 0.0138404 | 2.28739 | 6.79076 | 0.831919 | 0.2231587 |
| D21S11 | 4 | 0.000249157 | 0.0000610297 | 4.08256 | 27.7237 | 1.44285 | 0.7275322 |
| D19S433 | 4 | 0.000153201 | 0.000074127 | 2.06674 | 57.2977 | 1.75814 | 0.7445170 |
| TH01 | 4 | $5.01194\times10^{-6}$ | $2.04849\times10^{-6}$ | 2.44666 | 140.188 | 2.14671 | 0.7455309 |
| FGA | 4 | 0.0000389945 | $5.02138\times10^{-6}$ | 7.7657 | 1088.66 | 3.03689 | 0.7465320 |
| D3S1358 | 4 | 0.0000527271 | 0.000051154 | 1.03075 | 1122.13 | 3.05004 | 0.7315215 |
| D18S51 | 5 | $5.07057\times10^{-7}$ | $5.541\times10^{-7}$ | 0.915099 | 1026.86 | 3.01151 | 1.9253719 |
| vWA | 5 | $8.77879\times10^{-6}$ | $1.98909\times10^{-6}$ | 4.41348 | 4532.04 | 3.65629 | 1.9524050 |
| D2S1338 | 6 | $2.88807\times10^{-8}$ | $7.0872\times10^{-9}$ | 4.07505 | 18468.3 | 4.26643 | 4.7773906 |



| FGA | {Prob for donor with dropout, 0.2} | a priori prob | LR |
|---|---|---|---|
| {21., 24.} | 0.266427 | 0.0459493 | 5.79827 |
| {23., 24.} | 0.262802 | 0.0405357 | 6.48322 |
| {24., 24.} | 0.14714 | 0.0189475 | 7.7657 |
| {21., 23.} | 0.113006 | 0.0491514 | 2.29915 |
| {20., 24.} | 0.0935648 | 0.0369047 | 2.53531 |
| {20., 21.} | 0.0394481 | 0.0447485 | 0.881551 |
| {20., 23.} | 0.0389302 | 0.0394764 | 0.986164 |
| {21., 21.} | 0.0149037 | 0.0278578 | 0.534991 |
| {23., 23.} | 0.0131309 | 0.0216803 | 0.605659 |
| {24., Unseen} | 0.00490455 | 0.114015 | 0.0430167 |
| {21., Unseen} | 0.00215517 | 0.138248 | 0.0155892 |
| {23., Unseen} | 0.00214067 | 0.12196 | 0.0175522 |
| {20., Unseen} | 0.000731383 | 0.111036 | 0.00658692 |
| {20., 20.} | 0.000704931 | 0.0179701 | 0.0392279 |
| {Unseen, Unseen} | 0.0000100735 | 0.171519 | 0.0000587309 |

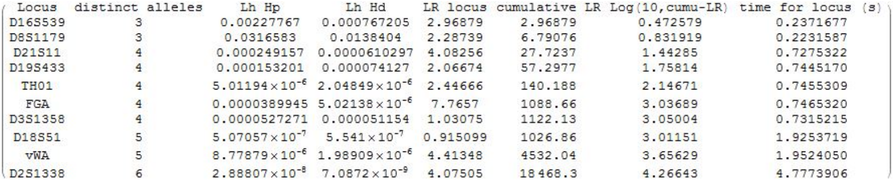
# D16S539: donor is (12, 13)



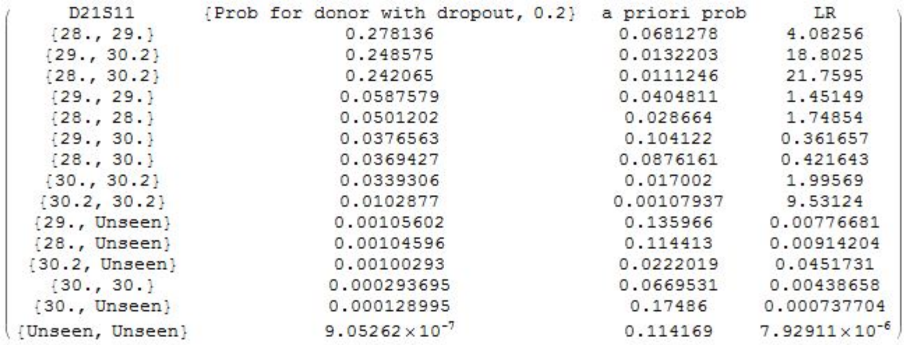
| Locus | distinct alleles | Lh Hp | Lh Hd | LR locus | cumulative LR | Log(10,cumu-LR) | time for locus (s) |
|---|---|---|---|---|---|---|---|
| D16S539 | 3 | 0.00227767 | 0.000767205 | 2.96879 | 2.96879 | 0.472579 | 0.2371677 |
| D8S1179 | 3 | 0.0316583 | 0.0138404 | 2.28739 | 6.79076 | 0.831919 | 0.2231587 |
| D21S11 | 4 | 0.000249157 | 0.0000610297 | 4.08256 | 27.7237 | 1.44285 | 0.7275322 |
| D19S433 | 4 | 0.000153201 | 0.000074127 | 2.06674 | 57.2977 | 1.75814 | 0.7445170 |
| TH01 | 4 | $5.01194\times10^{-6}$ | $2.04849\times10^{-6}$ | 2.44666 | 140.188 | 2.14671 | 0.7455309 |
| FGA | 4 | 0.0000389945 | $5.02138\times10^{-6}$ | 7.7657 | 1088.66 | 3.03689 | 0.7465320 |
| D3S1358 | 4 | 0.0000527271 | 0.000051154 | 1.03075 | 1122.13 | 3.05004 | 0.7315215 |
| D18S51 | 5 | $5.07057\times10^{-7}$ | $5.541\times10^{-7}$ | 0.915099 | 1026.86 | 3.01151 | 1.9253719 |
| vWA | 5 | $8.77879\times10^{-6}$ | $1.98909\times10^{-6}$ | 4.41348 | 4532.04 | 3.65629 | 1.9524050 |
| D2S1338 | 6 | $2.88807\times10^{-8}$ | $7.0872\times10^{-9}$ | 4.07505 | 18468.3 | 4.26643 | 4.7773906 |

| D16S539 | {Prob for donor with dropout, 0.2} | a priori prob | LR |
|---|---|---|---|
| {11., 12.} | 0.308215 | 0.169968 | 1.81337 |
| {12., 13.} | 0.281462 | 0.0948071 | 2.96879 |
| {12., 12.} | 0.252317 | 0.0698379 | 3.6129 |
| {11., 13.} | 0.112117 | 0.115369 | 0.971817 |
| {11., 11.} | 0.0262545 | 0.103415 | 0.253874 |
| {13., 13.} | 0.0143993 | 0.0321759 | 0.447517 |
| {12., Unseen} | 0.00291212 | 0.124086 | 0.0234686 |
| {11., Unseen} | 0.00119586 | 0.150997 | 0.00791975 |
| {13., Unseen} | 0.0011222 | 0.0842252 | 0.0133238 |
| {Unseen, Unseen} | $4.10177\times10^{-6}$ | 0.055118 | 0.000074418 |

# D21S11: donor is (28, 29)



| Locus | distinct alleles | Lh Hp | Lh Hd | LR locus | cumulative LR | Log(10,cumu-LR) | time for locus (s) |
|---|---|---|---|---|---|---|---|
| D16S539 | 3 | 0.00227767 | 0.000767205 | 2.96879 | 2.96879 | 0.472579 | 0.2371677 |
| D8S1179 | 3 | 0.0316583 | 0.0138404 | 2.28739 | 6.79076 | 0.831919 | 0.2231587 |
| D21S11 | 4 | 0.000249157 | 0.0000610297 | 4.08256 | 27.7237 | 1.44285 | 0.7275322 |
| D19S433 | 4 | 0.000153201 | 0.000074127 | 2.06674 | 57.2977 | 1.75814 | 0.7445170 |
| TH01 | 4 | $5.01194\times10^{-6}$ | $2.04849\times10^{-6}$ | 2.44666 | 140.188 | 2.14671 | 0.7455309 |
| FGA | 4 | 0.0000389945 | $5.02138\times10^{-6}$ | 7.7657 | 1088.66 | 3.03689 | 0.7465320 |
| D3S1358 | 4 | 0.0000527271 | 0.000051154 | 1.03075 | 1122.13 | 3.05004 | 0.7315215 |
| D18S51 | 5 | $5.07057\times10^{-7}$ | $5.541\times10^{-7}$ | 0.915099 | 1026.86 | 3.01151 | 1.9253719 |
| vWA | 5 | $8.77879\times10^{-6}$ | $1.98909\times10^{-6}$ | 4.41348 | 4532.04 | 3.65629 | 1.9524050 |
| D2S1338 | 6 | $2.88807\times10^{-8}$ | $7.0872\times10^{-9}$ | 4.07505 | 18468.3 | 4.26643 | 4.7773906 |



| D21S11 | {Prob for donor with dropout, 0.2} | a priori prob | LR |
|---|---|---|---|
| {28., 29.} | 0.278136 | 0.0681278 | 4.08256 |
| {29., 30.2} | 0.248575 | 0.0132203 | 18.8025 |
| {28., 30.2} | 0.242065 | 0.0111246 | 21.7595 |
| {29., 29.} | 0.0587579 | 0.0404811 | 1.45149 |
| {28., 28.} | 0.0501202 | 0.028664 | 1.74854 |
| {29., 30.} | 0.0376563 | 0.104122 | 0.361657 |
| {28., 30.} | 0.0369427 | 0.0876161 | 0.421643 |
| {30., 30.2} | 0.0339306 | 0.017002 | 1.99569 |
| {30.2, 30.2} | 0.0102877 | 0.00107937 | 9.53124 |
| {29., Unseen} | 0.00105602 | 0.135966 | 0.00776681 |
| {28., Unseen} | 0.00104596 | 0.114413 | 0.00914204 |
| {30.2, Unseen} | 0.00100293 | 0.0222019 | 0.0451731 |
| {30., 30.} | 0.000293695 | 0.0669531 | 0.00438658 |
| {30., Unseen} | 0.000128995 | 0.17486 | 0.000737704 |
| {Unseen, Unseen} | $9.05262\times10^{-7}$ | 0.114169 | $7.92911\times10^{-6}$ |

# Two mixtures

## Deconvolute each mixture

- Mixture $M$: obtain probabilities $P_{\vec{d},c}(D = g \mid M)$ based on the chosen dropout probabilities $\vec{d}$ and drop-in $c$
- Mixture $M'$: obtain probabilities $P_{\vec{d'},c'}(D' = g \mid M')$ based on the chosen dropout probabilities $\vec{d'}$ and drop-in $c'$

## Match the donors

- $H_1$: $D = D'$, i.e., common donor
- $H_2$: $D \neq D'$, i.e., no common donor
- Then the LR becomes

$$\begin{aligned} LR(M, M') &= \frac{P(M, M' \mid H_1)}{P(M, M' \mid H_2)} \\ &= \sum_g \frac{P_{\vec{d},c}(D = g \mid M) P_{\vec{d'},c'}(D' = g \mid M')}{p(g)}. \end{aligned}$$

# Properties

## Special case: person-mixture

If $M'$ is a single source trace (e.g. reference sample) with genotype $g_0$ then

$$LR(M, M') = \sum_g \frac{P_{\vec{d},c}(D = g \mid M) P_{\vec{d'},c'}(D' = g \mid M')}{p(g)}$$

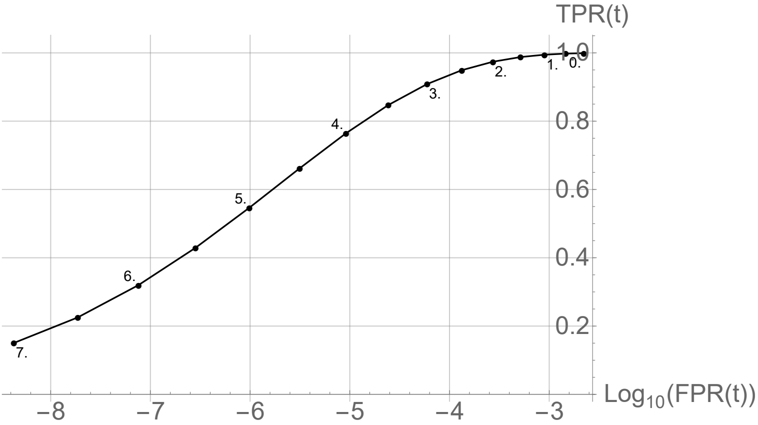
reduces to

$$LR(M, g_0) = \frac{P_{\vec{d},c}(D = g_0 \mid M)}{p(g_0)} = \frac{P_{\vec{d},c}(M \mid D = g_0)}{P_{\vec{d},c}(M)},$$

which is the previously discussed LR to test contribution of a PoI.

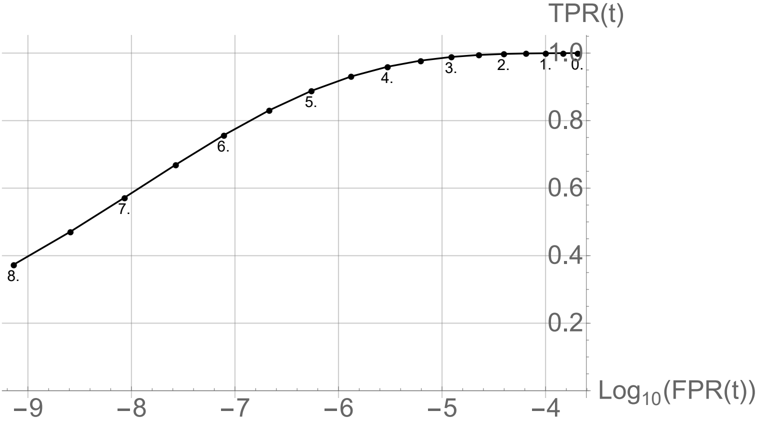
Therefore $LR(M, M')$ can be seen as a natural extension of the LR method, for comparison of any pair of traces, each of which may but need not be a mixed one.

# Power for two-person mixtures, no dropout, on the 15 NGM loci



True positive rate $TPR(t) = P(LR(M, M') \geq t \mid H_1)$ and false positive rate $FPR(t) = P(LR(M, M') \geq t \mid H_2)$ for various thresholds $t$; dots labelled by $Log_{10}(t)$.

# Power for two-person mixtures, no dropout, on the 21 GlobalFiler loci

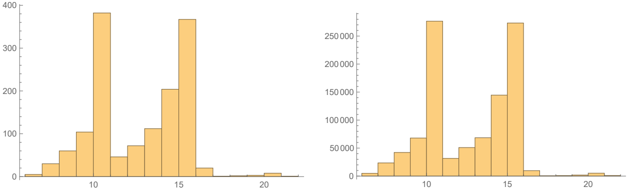


True positive rate $TPR(t) = P(LR(M, M') \geq t \mid H_1)$ and false positive rate $FPR(t) = P(LR(M, M') \geq t \mid H_2)$ for various thresholds $t$; dots labelled by $Log_{10}(t)$.

# Dutch DNA database

## Data

1417 two-person mixtures



(a) Histogram of the number of loci for which the mixtures from the database are typed.

(b) Histogram of the number of loci for which the mixture pairs are compared.

Figure 1: Breakdown of database mixtures and mixture comparisons according to number of loci

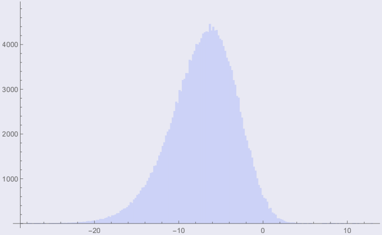
# Results

## Parameters

- Dropout $(0, 0.5)$, $c = 0$, test for major donor to be the same
- Reason: sometimes a partly derived profile is entered instead of whole mixture

## Results

$1,417 \cdot 1,416/2 = 1,003,236$ LR's calculated, of which 204,870 non-zero:



# Matches above $Log_{10}(LR)$-threshold $t$

| t | Matches |
|---|---|
| 0 | 3558 |
| 1 | 1288 |
| 2 | 410 |
| 3 | 118 |
| 4 | 43 |
| 5 | 32 |
| 6 | 26 |
| 7 | 26 |
| 8 | 22 |
| 9 | 16 |
| 10 | 12 |
| 11 | 7 |
| 12 | 6 |

# Follow-up

## True positives

- All matches with $LR > 10^5$ were investigated
- In all cases, these turned out to indeed correspond to mixtures with a common donor

## False negatives

- Two pairs were not found above the threshold
- These had $LR$ equal to 6 resp. 1200

# Conclusions

## Summary

- LR calculation for contribution of a donor amounts to deconvolution of the mixture
- One can then 'match' two deconvoluted mixtures
- This yields LR for the mixtures to have/not to have a common donor
- This LR is a generalization of the trace-person LR usually considered
- Provides additional investigative information to connect cases with each other.
- False positive rate can be controlled by LR-threshold

## Contact

k.slooten@nfi.minvenj.nl