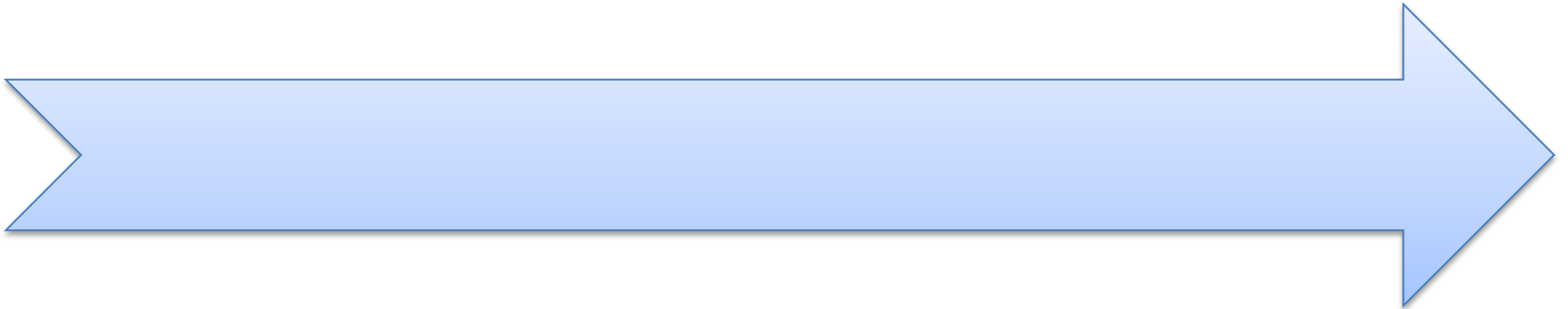


Campioni complessi, stima dell'errore e calcoli statistici

Federica Alessandrini
Medicina Legale
Università Politecnica delle Marche

Sensibilità del metodo



Complessità dei risultati

“Complex DNA profiles”

A ‘complex DNA profile’ is any profile that is, or may be, subject to allele **drop-out** and/or allele **drop-in**

“complex DNA profiles,” i.e. those profiles that are **mixtures** of two or more **contributors**, where allele drop-out, secondary transfer, and **contamination** are additional complications

“Complex DNA profiles”

A ‘complex DNA profile’ is any profile that is, or may be, subject to allele drop-out and/or allele drop-in

“complex DNA profiles” i.e. those profiles that are **mixed** (two or more **contributors**, where drop-out, secondary transfer, and contamination are additional complications

UNCERTAINTY

Elementi per l'interpretazione dei profili complessi

Principi
(teoria)



Protocolli
(validazione)



Pratica
(training e esperienza)

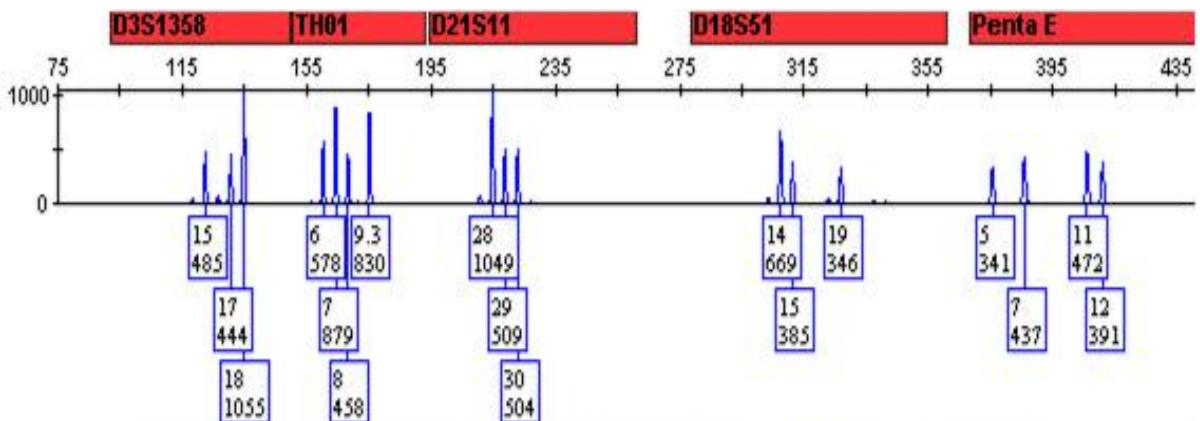
ISFG Recommendations
SWGDM guidelines



Lab SOPs

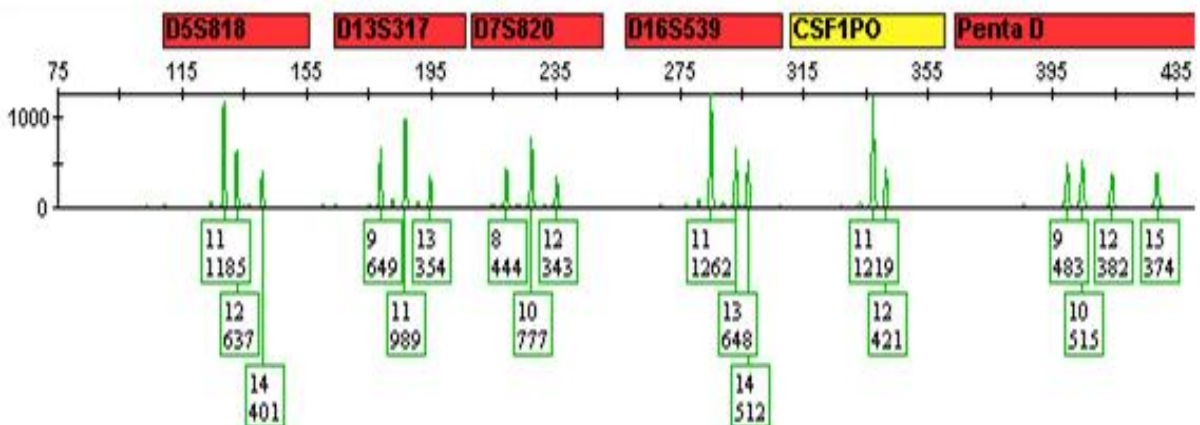


Attività all'interno
del lab



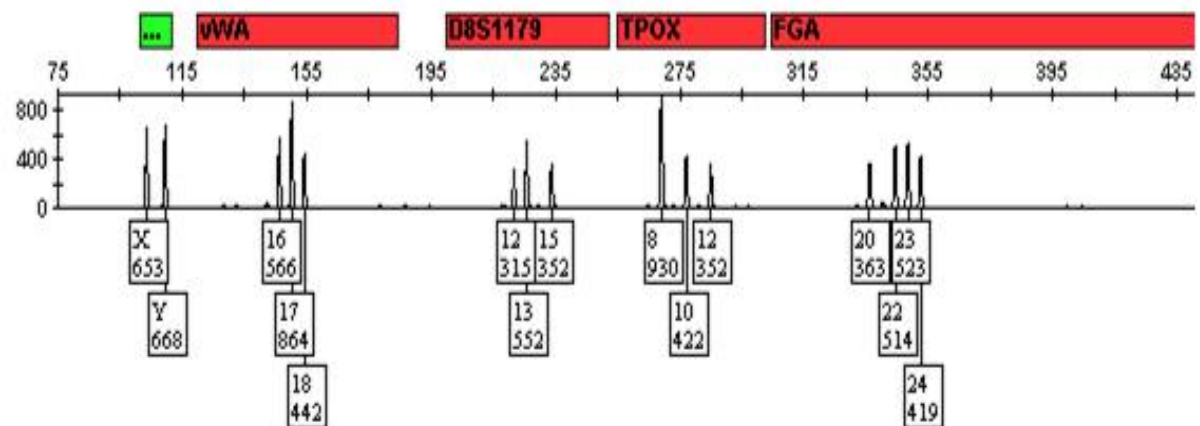
Quanti contributori?

2



Mixture ratio?

$$Mx = \frac{879+830}{578+458} = 1.65$$



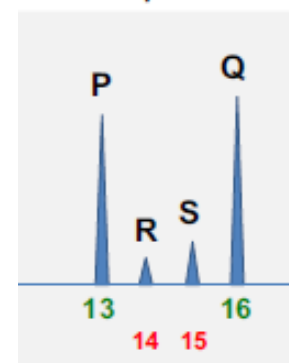
Possibili genotipi?

2	1 allele (P)	8	2 alleles (P, Q)	12	3 alleles (P, Q, R)	6	4 alleles (P, Q, R, S)
1	2 hom, 1 shared	1	hom + hom, 0 shared	3	hom + het, 1 shared	3	het + het, 0 shared
	PP PP		PP QQ		PP QR		PQ RS
		2	hom + het, 1 shared		QQ PR		PR QS
			PP PQ		RR PQ		QR PS
			QQ PQ				
		1	2 het, 2 shared	3	het + het, 1 shared		
			PQ PQ		PQ PR		
					PR QR		
					QR PQ		

Reciprocal genotype combinations

1	2 hom, 1 shared	1	hom + hom, 0 shared	3	hom + het, 1 shared	3	het + het, 0 shared
	PP PP		QQ PP		QR PP		RS PQ
		2	hom + het, 1 shared		PR QQ		QS PR
			PQ PP		PQ RR		PS QR
			PQ QQ				
		1	2 het, 2 shared	3	het + het, 1 shared		
			PQ PQ		PR PQ		
					QR PR		
					PQ QR		

Example data

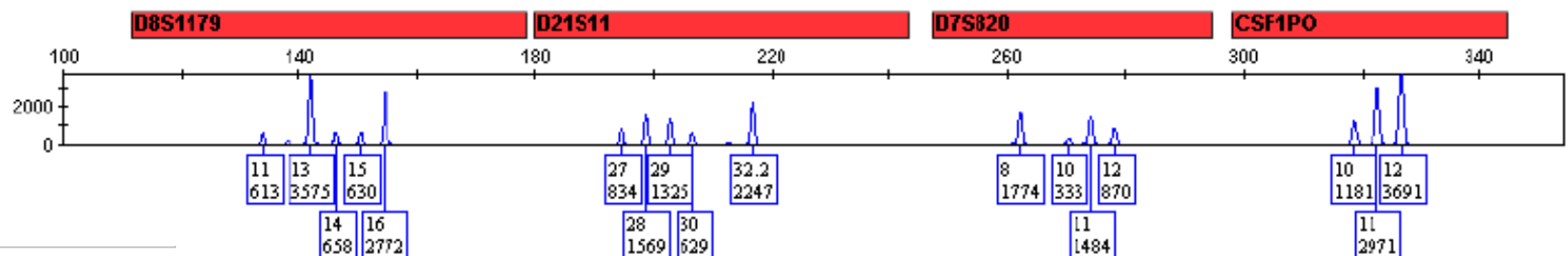


Unrestricted genotype combinations

PQ & RS 13, 16 & 14, 15	RS & PQ 14, 15 & 13, 16
PR & SQ 13, 14 & 15, 16	SQ & PR 15, 16 & 13, 14
PS & RQ 13, 15 & 14, 16	RQ & PS 14, 16 & 13, 15

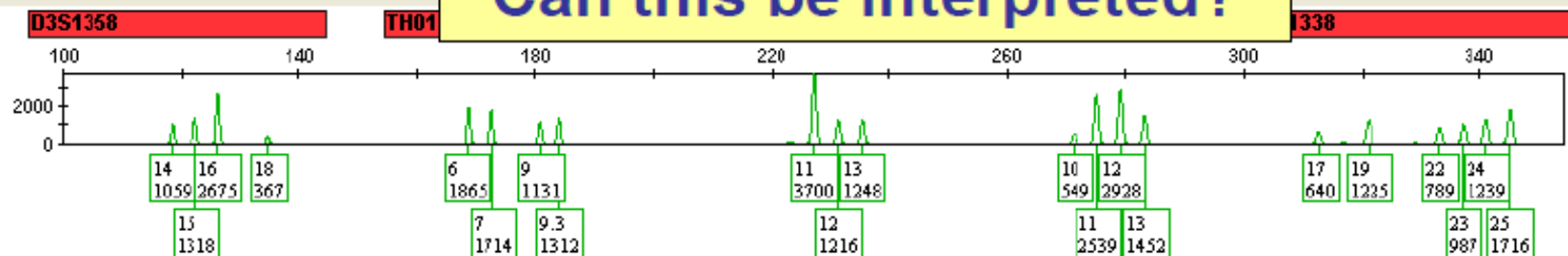
Restricted genotype combinations

PQ & RS 13, 16 & 14, 15	RS & PQ 14, 15 & 13, 16
----------------------------	----------------------------

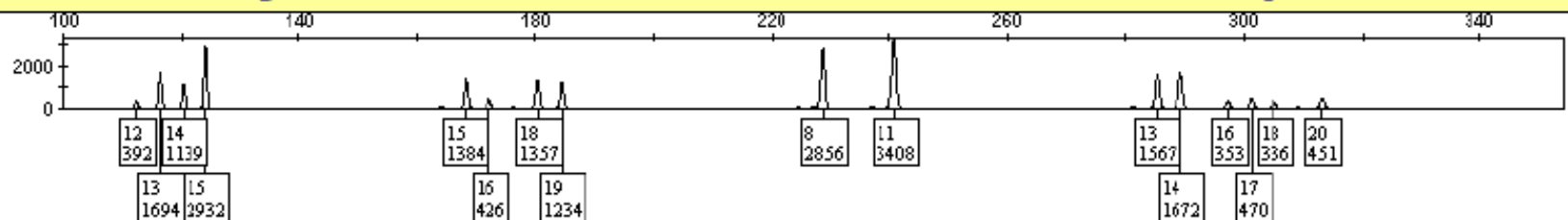


Can this be interpreted?

☐ Mark Sample for Deletion

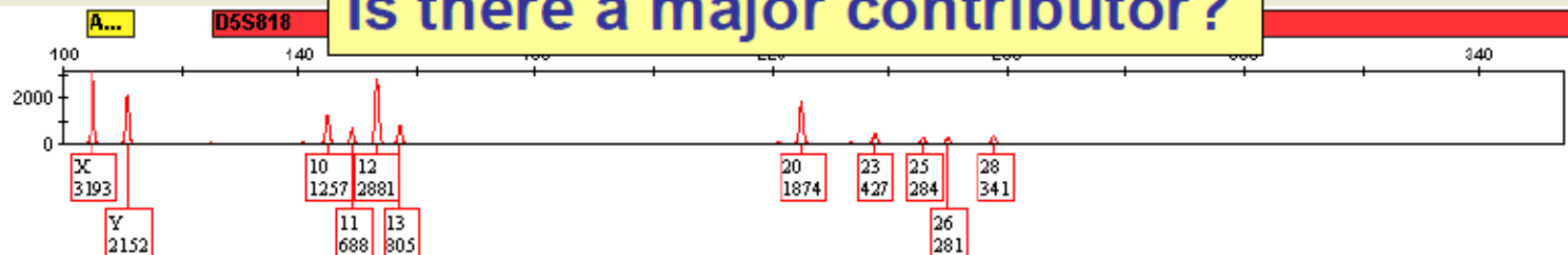


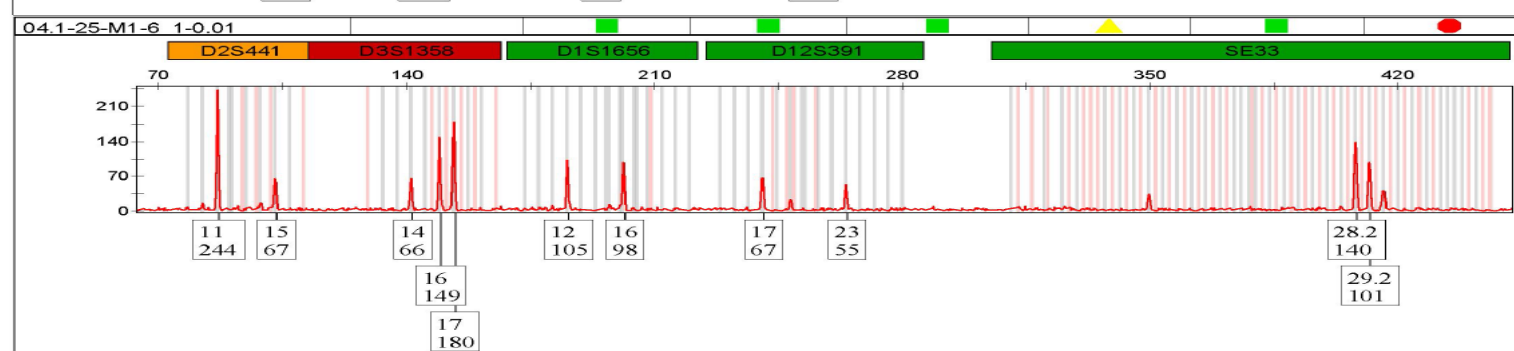
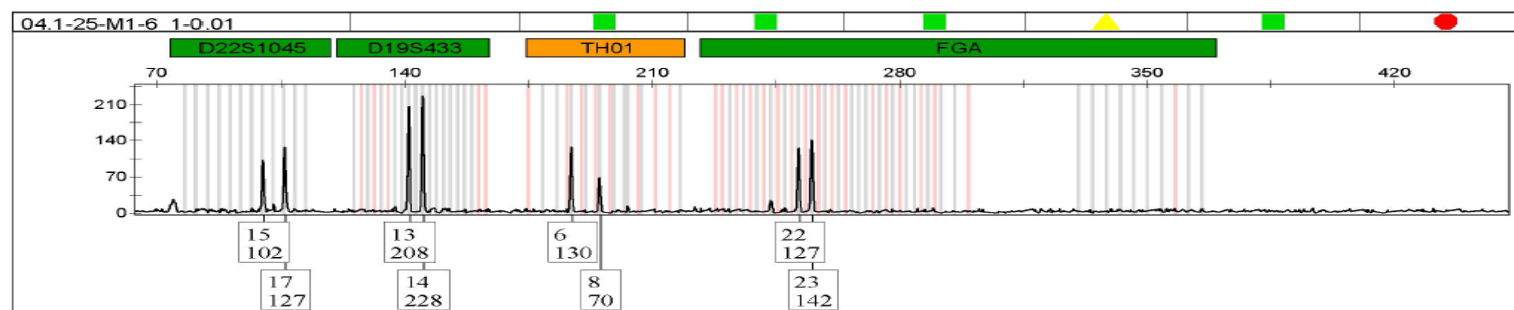
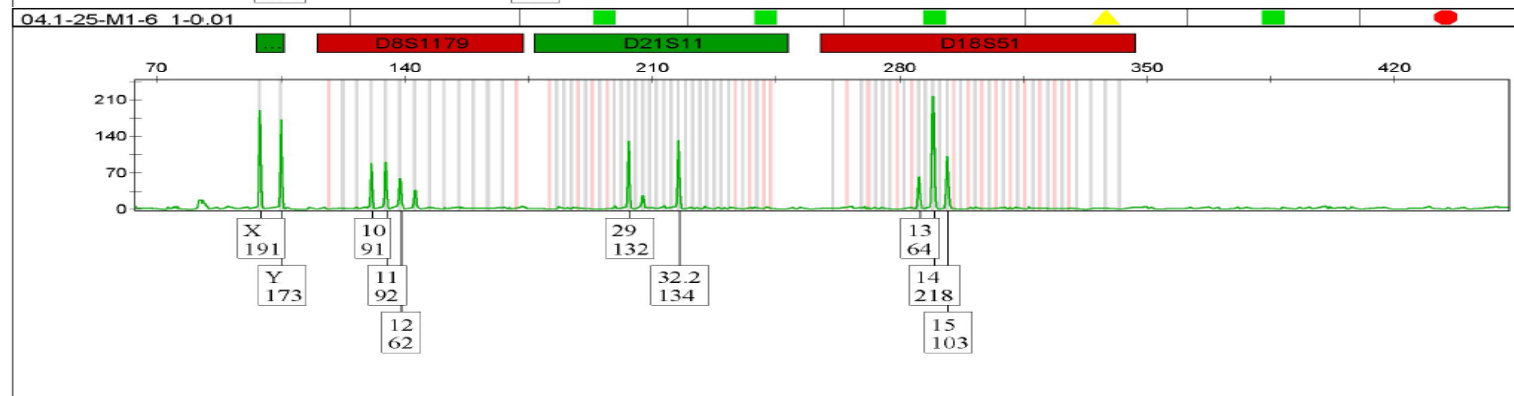
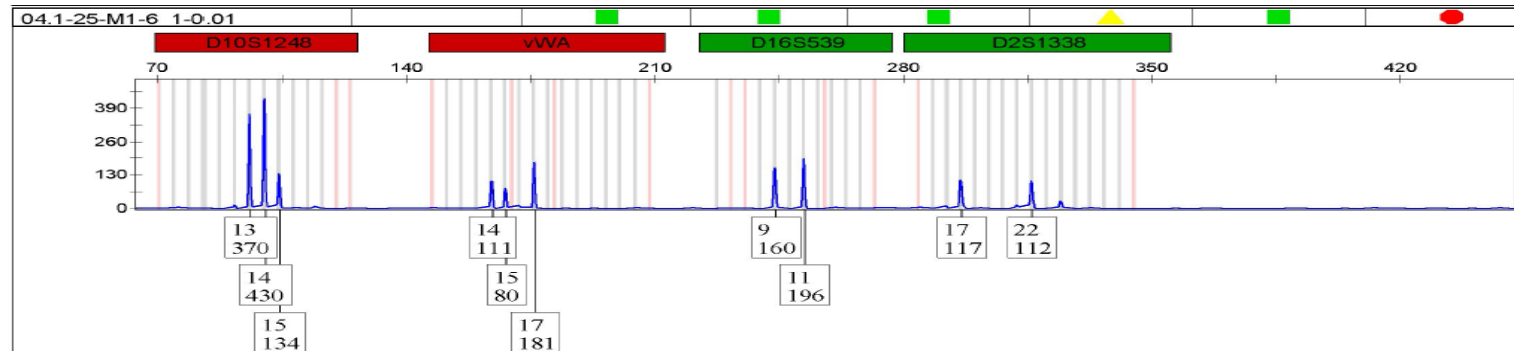
How many contributors assumed for interpretation?



Is there a major contributor?

☐ Mark Sample for Deletion





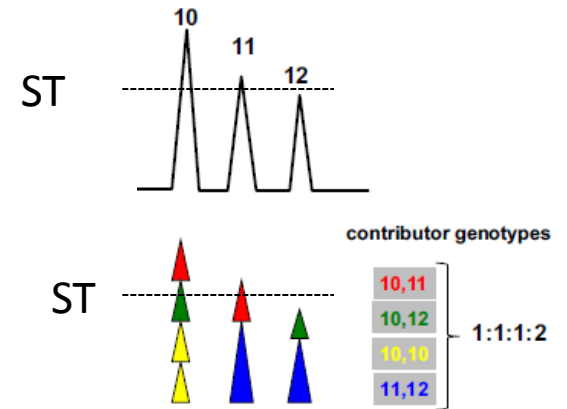
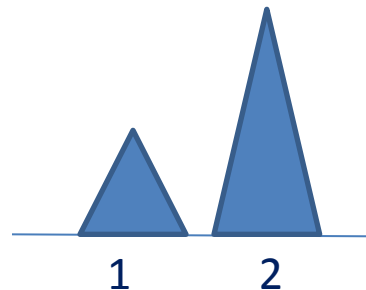
Sfide interpretative nei profili complessi

✓ Allele sharing

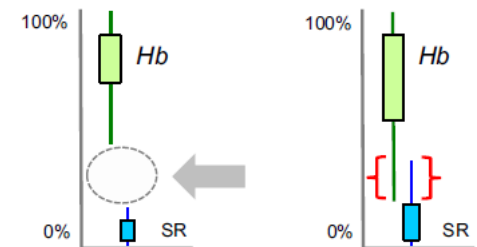
✓ Valore di ST

✓ LT-DNA

effetti stocastici:
sbilanciamento eterozigoti
drop out/drop in
altezza stutter >15%



(a) PHRs with optimal DNA amounts (e.g., 1 ng) (b) PHRs with low DNA amounts (e.g., 100 pg)



Drop in e drop out

Fenomeni stocastici che si verificano in caso di LT-DNA

Drop out: un allele non viene amplificato ($h \text{ picco} < \text{soglia analitica}$)
→ allele presente nel riferimento ma non nella traccia

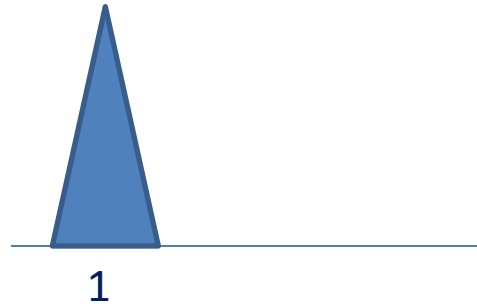
Drop in: presenza di **uno/max due** alleli «estranei» nel profilo
→ allele presente nella traccia ma non nel riferimento

Drop in ≠ contaminazione

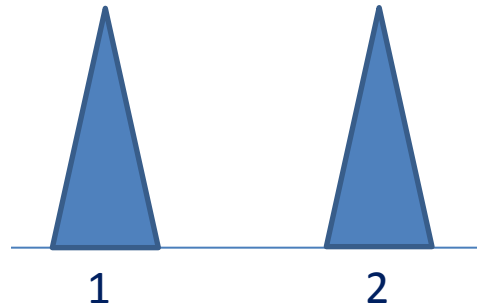
- La contaminazione può essere spiegata introducendo un altro contributore sconosciuto
- I drop in sono eventi indipendenti, la contaminazione è un evento dipendente (c'è stato un trasferimento di DNA)

Drop out

- Traccia



- Riferimento

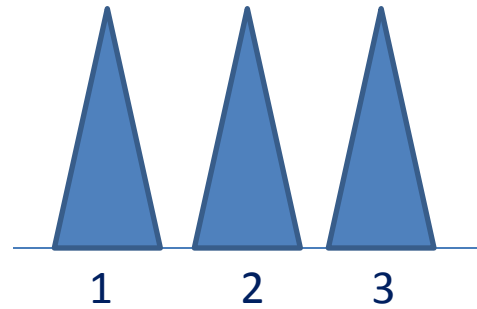


Match?

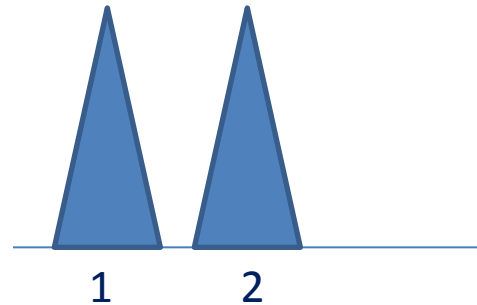
L'allele 2 deve «essere scomparso» (drop out) nella traccia

Drop in

- Traccia



- Riferimento



Match?

L'allele 3 deve essere «comparso» (drop in) nella traccia

Sfide interpretative nei campioni complessi

✓ **# di contributori** (maximum allele count non adatto)

True # of contributors

		1	2	3	4	5
		<i>2 alleles</i>	<i>4 alleles</i>	<i>6 alleles</i>	<i>8 alleles</i>	<i>10 alleles</i>
6	CODIS 13	1.75×10^{-40}	6.34×10^{-9}	0.161	0.946	0.999
	22 STRs	0 ($<10^{-99}$)	9.59×10^{-21}	5.32×10^{-5}	0.188	0.860
5	CODIS 13	9.78×10^{-33}	2.10×10^{-6}	0.414	0.990	
	CODIS22	6.36×10^{-61}	7.01×10^{-15}	0.00484	0.610	
4	CODIS 13	7.02×10^{-25}	0.00052	0.786		
	22 STRs	3.50×10^{-46}	3.49×10^{-9}	0.165		
3	CODIS 13	8.42×10^{-17}	0.05949			
	22 STRs	5.77×10^{-31}	0.00043			
2	CODIS 13	1.70×10^{-8}				
	22 STRs	2.05×10^{-15}				

Sfide interpretative nei campioni complessi

✓ # di
possibili
genotipi

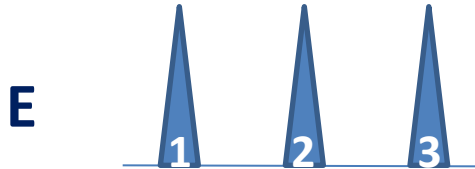
	1 contributor either hom or het	#	2 contributors 7 classes	#	3 contributors 23 classes	#	4 contributors 41 classes	#
1 allele (1 hom, 0 het)	homozygote (hom)	1	(a) 2 hom, 1 shared	1	(a) 3 hom, 1 shared	1	(a) 4 hom, 1 shared	1
2 alleles (2 hom, 1 het)	heterozygote (het)	1	(a) hom + hom, 0 shared (b) hom + het, 1 shared (c) hom + het, 2 shared	1 2 1	(a) 2 hom + hom, 0 shared (b) 2 hom + het, 1 shared (c) hom + hom + het, 2 shared (d) hom + 2 het, 2 shared (e) 3 het, 2 shared	2 2 1 2 1	(a) 3 hom + hom, 1 shared (b) 2 hom + 2 hom, 2 shared (c) 3 hom + het, 1 shared (d) 2 hom + hom + het, 2 shared (e) 2 hom + 2 het, 2 shared (f) hom + hom + 2 het, 2 shared (g) hom + 3 het, 2 shared (h) 4 het, 2 shared	2 1 2 2 1 1 2 1
3 alleles (3 hom, 3 het)	— possible tri-allele?	—	(a) hom + het, 1 shared (b) het + het, 1 shared	3 3	(a) hom + hom + hom, 0 shared (b) hom + hom + het, 0 shared (c) hom + hom + het, 1 shared (d) hom + 2 het, 2 shared (2:2:2) (e) hom + het + het, 2 shared (4:1:1) (f) hom + het + het, 2 shared (3:2:1) (g) het + het + het, 2 shared (3:2:1) (h) het + het + het, 3 shared (2:2:2)	1 3 6 3 3 6 6 1	(a) 2 hom + hom + hom, 1 shared (b) 3 hom + het, 1 shared (c) hom + hom + hom + het, 2 shared (d) hom + hom + het + het, 2 shared (e) hom + 2 het + het, 2 shared (f) hom + het + het + het, 3 shared (g) 3 het + het, 2 shared (h) 2 het + 2 het, 3 shared (i) 2 het + het + het, 3 shared	3 3 3 5 9 3 6 2 3
4 alleles (4 hom, 6 het)	—	—	(a) het + het, 0 shared	3	(a) hom + hom + het, 0 shared (b) hom + het + het, 1 shared (3:1:1:1) (c) hom + het + het, 1 shared (2:2:1:1) (d) 2 het + het, 2 shared (2:2:1:1) (e) het + het + het, 1 shared (3:1:1:1) (f) het + het + het, 2 shared (2:2:1:1)	6 12 12 6 4 12	(a) hom + hom + hom + hom, 0 shared (b) 2 hom + hom + het, 1 shared (4:2:1:1) (c) hom + hom + hom + het, 1 shared (3:2:2:1) (d) hom + hom + 2 het, 2 shared (2:2:2:2) (e) hom + hom + het + het, 2 shared (3:2:2:1) (f) het + het + het + het, 4 shared (2:2:2:2) (g) het + het + het + het, 3 shared (3:2:2:1) (h) 2 hom + het + het, 2 shared (4:2:1:1) (i) hom + het + het + het, 3 shared (3:2:2:1) (j) 2 het + het + het, 3 shared (3:2:2:1) (k) 2 het + het + het, 3 shared (4:2:1:1) (l) 2 het + het + het, 2 shared (3:3:1:1)	1 12 12 6 24 2 8 12 24 24 12 24
5 alleles (5 hom, 10 het)	—	—	—	—	(a) hom + het + het, 0 shared (b) het + het + het, 1 shared	15 30	(a) hom + hom + hom + het, 0 shared (b) hom + hom + het + het, 1 shared (c) hom + het + het + het, 2 shared (d) het + het + het + het, 3 shared (2:2:2:1:1) (e) 2 het + het + het, 2 shared (3:2:1:1:1)	10 30 30 4 11
6 alleles (6 hom, 15 het)	—	—	—	—	(a) het + het + het, 0 shared	15	(a) hom + hom + het + het, 0 shared (b) hom + het + het + het, 1 shared (c) het + het + het + het, 2 shared	30 48 13
7 alleles (7 hom, 21 het)	—	—	—	—	—	—	(a) hom + het + het + het, 0 shared (b) het + het + het + het, 1 shared	70 35
8 alleles (8 hom, 28 het)	—	—	—	—	—	—	(a) het + het + het + het, 0 shared	105

Weight of the Evidence (WoE)

4.1. The laboratory **must perform statistical analysis in support of any inclusion** that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.

(2010 SWGDAM STR Interpretation Guidelines)

Approcci statistici al WoE



Allele-centrico

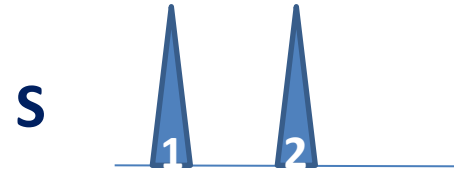
✓ RMNE (CPI)

La probabilità che una persona a caso non possa essere esclusa dall'aver contribuito alla mistura

$$\text{RMNE} = (p_1 + p_2 + p_3)^2$$

RMNE non tiene conto del profilo di S

$p_j = 0.2$, $\text{RMNE} = 0.36$: circa 1 persona su 3 può aver contribuito ad E



Genotipo-centrico

✓ RMP

La probabilità che una persona a caso possa avere lo stesso profilo di S

$$\text{RMP} = 2p_1p_2$$

✓ LR

Rapporto tra due probabilità

$$\frac{\Pr(\text{Evidence} | H_p)}{\Pr(\text{Evidence} | H_d)}$$

Approccio bayesiano all'interpretazione dei profili genetici complessi

Metodi
statistici
di base

Metodi statistici avanzati

Compatibilità/
Esclusione

Probabilistici: tutto è possibile

Modello
Binario

Modello
Semicontinuo

Modello
Continuo

$$LR = \frac{1}{2p_a p_b} \quad \text{or} \quad LR = \frac{0}{2p_a p_b}$$

(= inclusion) (= exclusion)

$$LR = \frac{n}{2p_a p_b} \quad \text{for} \quad 0 < n < 1$$

Binary LR

Semicontinuos
LR (drop in drop
out)

Continuos LR (drop in,
drop out, altezze dei
picchi)

Aumenta la complessità
Sempre più difficile da spiegare

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill ^{a,*}, C.H. Brenner ^b, J.S. Buckleton ^c, A. Carracedo ^d, M. Krawczak ^e, W.R. Mayr ^f,
N. Morling ^g, M. Prinz ^h, P.M. Schneider ⁱ, B.S. Weir ^j

Forensic Science International: Genetics 6 (2012) 679–688



Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



DNA commission of the International Society of Forensic Genetics:
Recommendations on the evaluation of STR typing results that may
include drop-out and/or drop-in using probabilistic methods

P. Gill ^{a,b,*}, L. Gusmão ^c, H. Haned ^d, W.R. Mayr ^e, N. Morling ^f, W. Parson ^g, L. Prieto ^h,
M. Prinz ⁱ, H. Schneider ^j, P.M. Schneider ^k, B.S. Weir ^l

Stima del drop out e del drop in

Basata su studi di validazione del metodo

✓ **Probabilità di drop out: $Pr(D)$**

determinata empiricamente tramite analisi di diluizioni seriali di DNA; regressione logistica

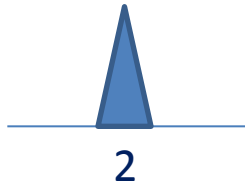
(Tvedebrink et al FSI Genetics 2009, 20012)

✓ **Probabilità di drop in: $Pr(C)$**

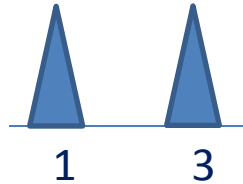
determinata empiricamente dall'analisi dei controlli negativi

LR considerando drop out e drop in

Traccia



Sospettato



$H_p: S$
 $H_d: U$

Affinchè sia vera H_p :

- ✓ Gli alleli 1 e 3 hanno subito drop out
- ✓ l'allele 2 ha subito drop in

$$P(E | H_p) = P(\text{drop-out 1}) * P(\text{drop-out 3}) * P(\text{dropin 2}) = d * d * cp_2$$

$P(\text{dropin}) = c$, $P(\text{dropin allele 2}) = cp_2$

Affinchè sia vera H_d lo sconosciuto può avere qualsiasi genotipo:

Genotipo	Dropout	Dropin	Probabilità del genotipo
2,2	$(1-d')$	$(1-c)$	p_2^2
2,Q	$(1-d)d$	$(1-c)$	$2p_2p_Q$
Q,Q	d'	cp_2	p_Q^2

$$P(E | H_d) = p_2^2(1-c)(1-d') + 2p_2p_Q(1-d)d(1-c) + p_Q^2d'cp_2$$



Q = ogni allele diverso da quello presente nella traccia
Assumendo che il locus abbia 5 alleli, ciascuno con frequenza p_i :
 $Q = \{2,3,4,5\}$ e $p_Q = p_2 + p_3 + p_4 + p_5$

$$LR = \frac{d * d * cp_2}{p_2^2(1-c)(1-d') + 2p_2p_Q(1-d)d(1-c) + p_Q^2d'cp_2}$$

SOFTWARES

✓ Modelli semicontinui:

Forensim/LRmix/LRmix studio (H. Haned, NFI)

LikeLTD (D. Balding)

LabRetriever (N Rudin, K. Inman, K. Lohmueller)

Forensic Statistic Tools FST (A. Mitchell, OCME)

.....

✓ Modelli continui

TrueAllele (M. Perlin)

STRmix(J. Buckleton et al, ESR, New Zeland)

LiRA (R. puch-Solis, LGC)

DNA-View (C. Brenner)

.....

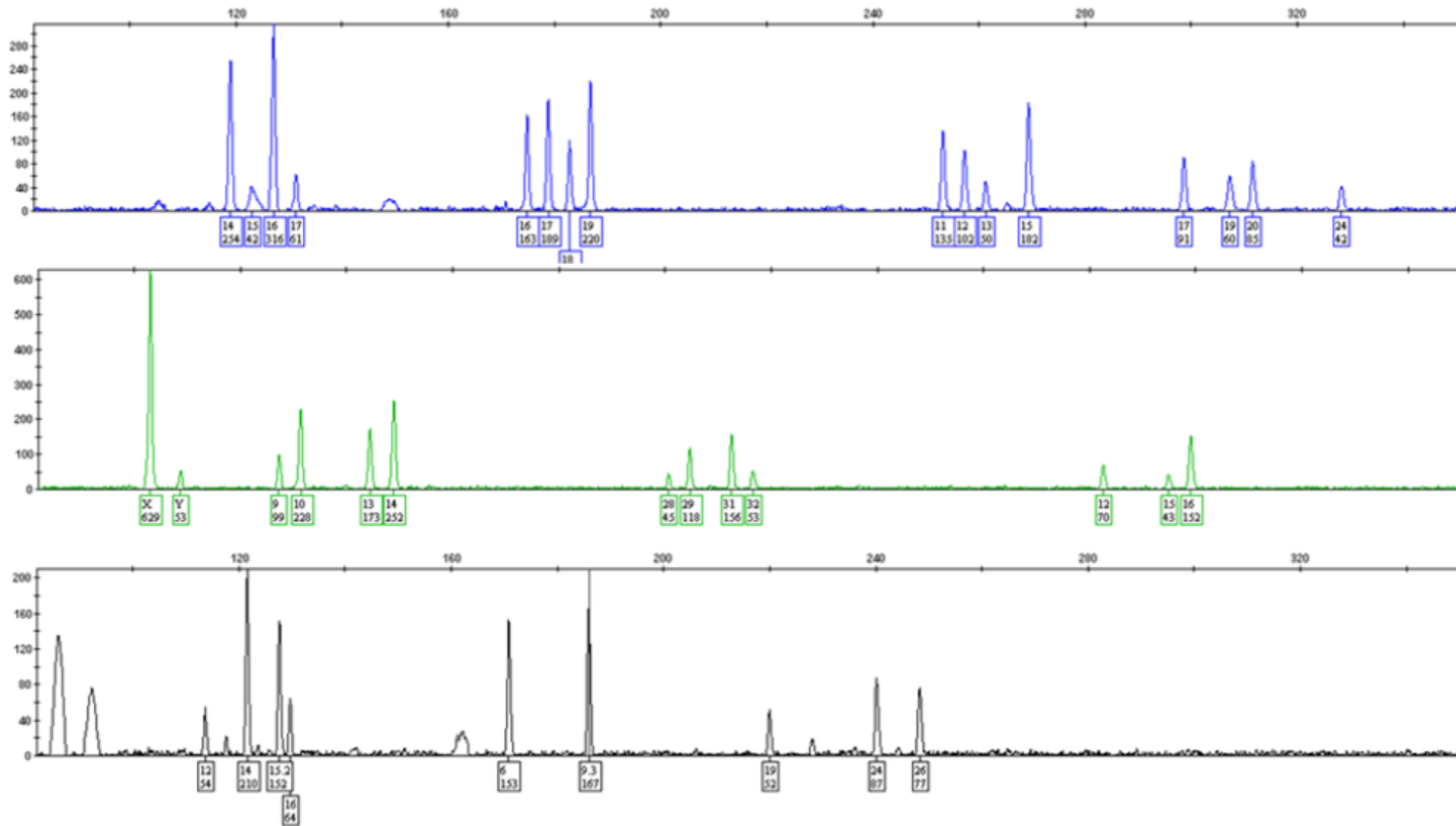
	EVIDENCE				S1		V	
Marker	Allele1	Allele2	Allele3	Allele4	Allele1	Allele2	Allele1	Allele2
D3S1358	15	16	17	18	17	18	16	17
VWA	14	16	19		16	19	14	16
D16S539	9	10	13		9	10	9	13
D2S1338	20	22	23	25	22	25	22	23

Diverso # di contributori (Hp ancorata)

Hp	Hd	log(LR)
S1+V+U	V+U	2.431

Stesso # di contributori (Hp non ancorata)

Hp	Hd	log(LR)
S2+V+U	V+2U	3.026



	EVIDENCE				S1		S2		V		Alleli Unici
Marker	Allele1	Allele2	Allele3	Allele4	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	
AMEL	X	Y			X	Y	X	Y	X	Y	
D3S1358	14	16	17		16	17	15	17	14	16	4
VWA	16	17	18	19	16	18	18	19	17	19	4
D16S539	11	12	13	15	12	13	12	12	11	15	4
D2S1338	17	19	20	(24)	19	20	17	18	17	24	4
D8S1179	9	10	13	14	9	13	13	13	10	14	4
D21S11	(28)	29	31	32	28	32	30	30	29	31	5
D18S51	12	(15)	16		12	15	12	20	16	16	4
D19S433	12	14	15.2	16	12	16	12	15	14	15.2	5
TH01	6	9.3			6	9.3	6	9.3	6	9.3	2
FGA	19	24	26		19	21	20	21	24	26	5

Hp	Hd	log(LR)
S1+S2+V	V+2U	5.496

Stima dell'errore: p-value per LR

p-value per LR: probabilità di osservare un valore di LR grande almeno come quello osservato se H_0 è vera

1. Calcolare la $LR(S)$
2. Sostire al profilo di S tutti i possibili profili genetici (diversi da S) di persone prese a caso nella popolazione (profili che possono verificarsi sotto H_d) e ricalcolare la LR per ognuno di essi, senza cambiare nessun parametro.
3. Considerare solo quei profili con $LR \geq LR(S)$
4. p -value è la somma delle frequenze genotipiche di questi profili e indica la probabilità che una persona presa a caso possa avere una $LR \geq LR(S)$

Stima dell'errore: performance test (LRmix)

Confronta la $LR(S)$ con le LR di individui presi a caso (H_d è vera):

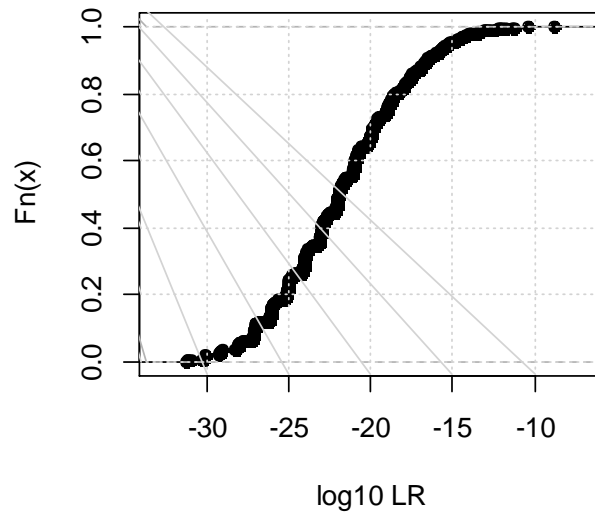
- ✓ p-value considera tutti i possibili profili genetici di individui presi a caso
- ✓ performance test considera un campione di profili genetici (es. 10 000)

Hp	Hd	log(LR)
S1+S2+V	V+2U	5.496

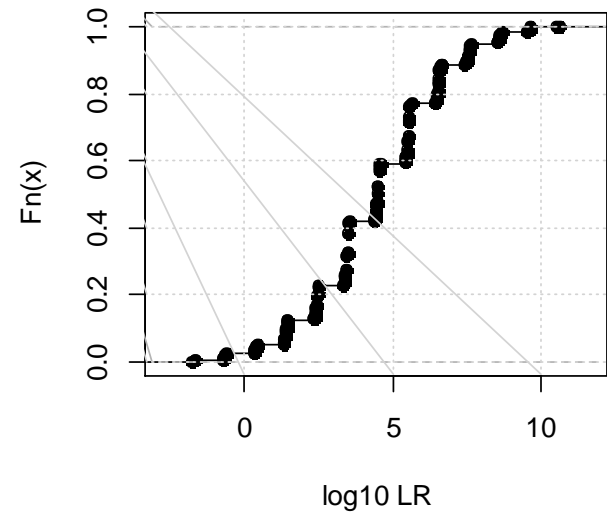
S1 sostituito con random men

S2 sostituito con random men

Empirical distribution function



Empirical distribution function



Hp	Hd	log(LR)
S1+V+U	V+U+U	8.929
S2+V+U	V+2U	-4.774
S1+V	V+U	13.09
S2+V	V+U	-17.38

Conclusioni

- ✓ **Determinazione empirica della probabilità degli effetti stocastici**
- ✓ **Metodo esploratorio:**
 - considerare più coppie di ipotesi
 - ipotesi semplici (un sospettato alla volta)
 - il # di contributori in H_p e H_d non deve essere necessariamente lo stesso
- ✓ **Validazione dei software**

Conclusioni

**QUALITA'
DEL
PROFILO GENETICO**

“If you cannot explain your evidence to someone that is not from the field (like a judge) and you need a lot of technical excuses to report something, then the result is not good. You should leave it on your desk and not take it to court.”

(Peter Schneider, Roma 2012)

