DNA Mixture Interpretation Workshop | ***Dr Chris Maguire***

Identification and resolution of DNA mixtures

Introduction

Mixture analysis models

- Pros & Cons

Recognizing a DNA Mixture

- Extra peaks
- Peak imbalance
- Biochemistry refresher

Resolving mixtures

- Number of contributors
- Mixing proportion
- Component pairing
- Use of quantitative data to eliminate or include contributors

Why consider DNA mixtures?

DNA Mixtures occur in casework …

… Sexual assault/rape

Mixed body fluids : semen/vaginal; semen/blood; semen/saliva etc

… Violent assault/Homicide

Overlapping bloodstains

Blood on saliva (shirt front) – contribution from wearer

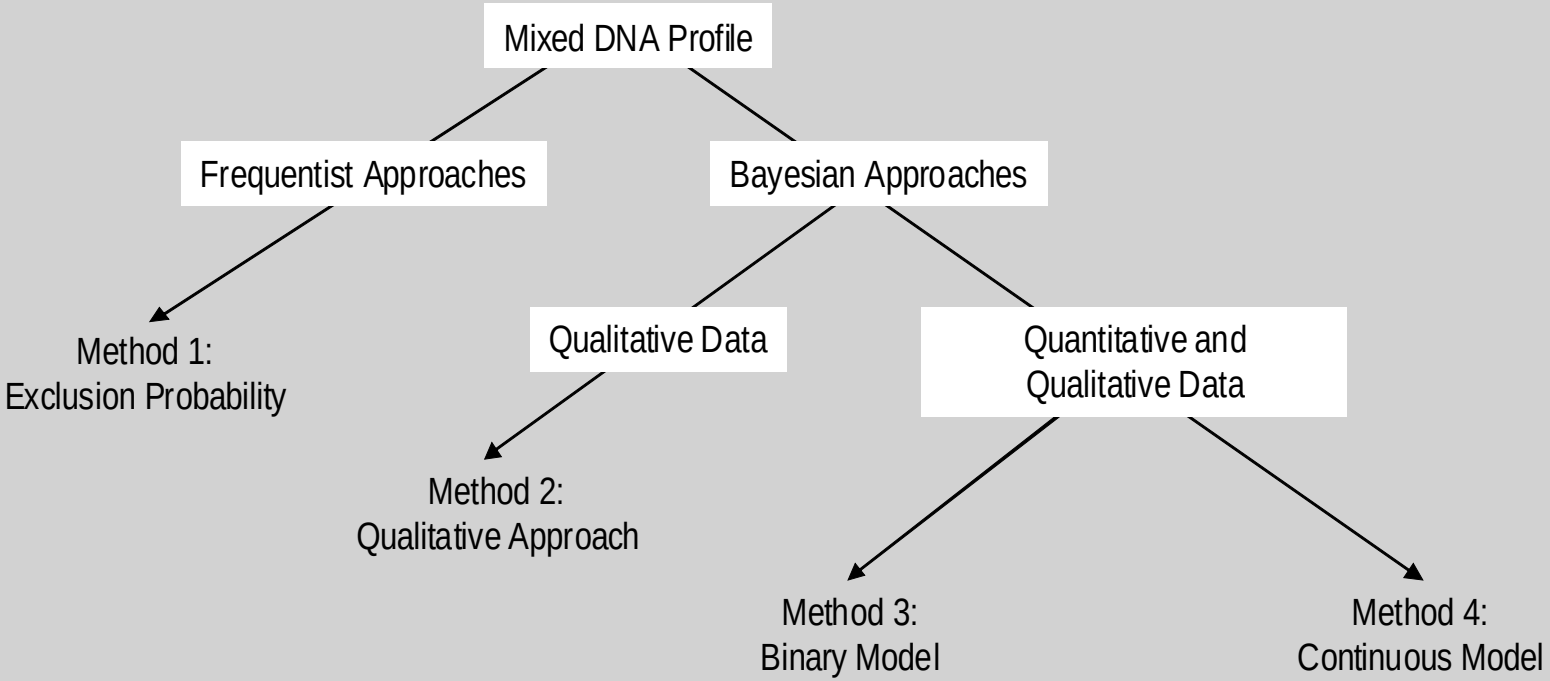
… Property Crime

Shared cigarettes

… Drug crime

Shared needles

Mixture Interpretation – Approaches



Mixture Interpretation Methods

DNA Mixture interpretation methods are divided according to their use of qualitative and/or quantitative data contained within the DNA profile presented.

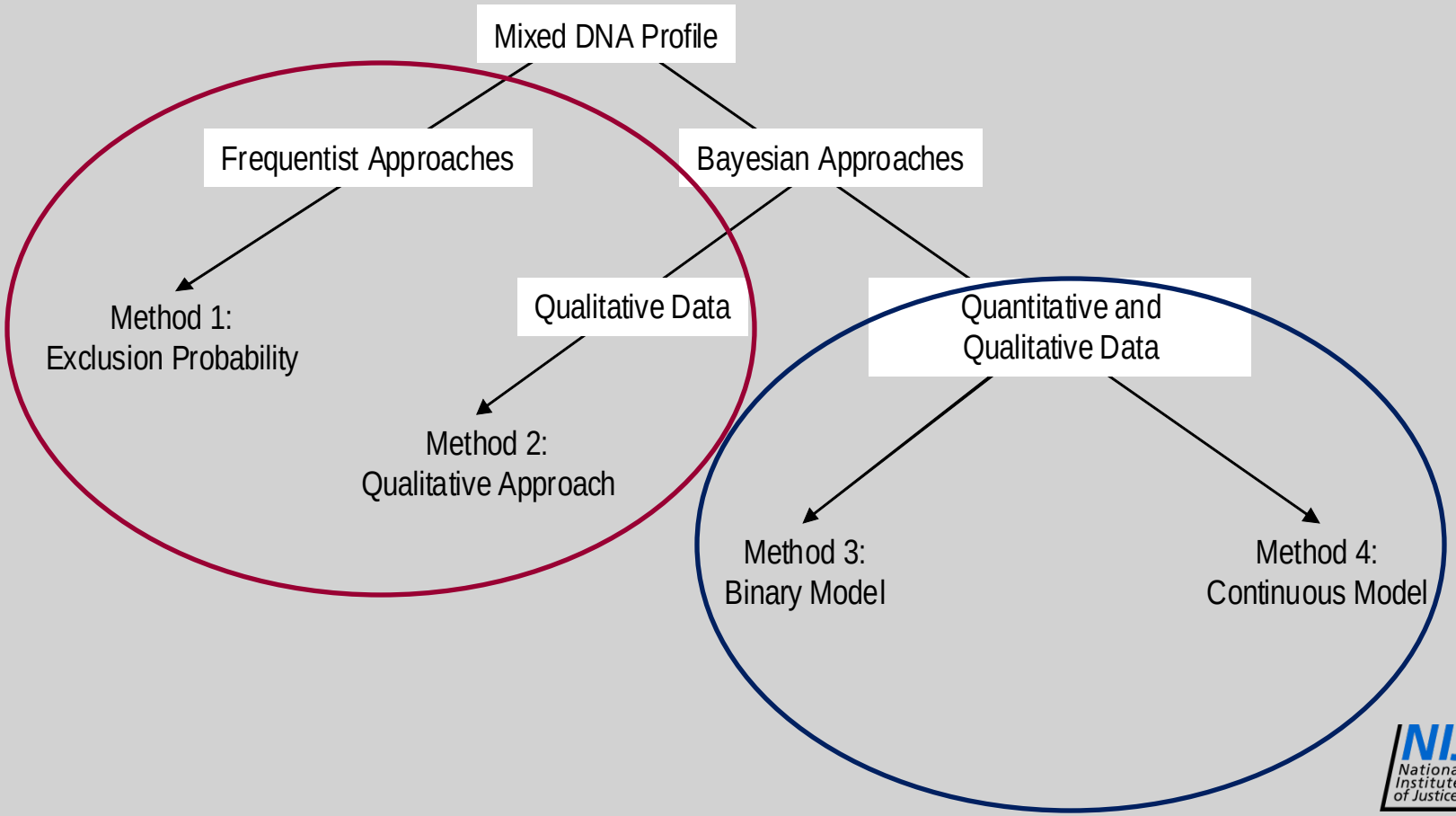
QUALITATIVE information

– which alleles are represented

QUANTITATIVE information

– the relative proportions of each allele represented

Mixture Interpretation – Methods



Exclusion Probability

Probability of Inclusion or Random Man Not Excluded (RMNE)

At any given locus (*l*) *the pI = (sum of all allele proportions)*^{2}

$$pI_l = \sum_{i=1}^{n} p(A_i)^2$$

Assumes:

- Full representation of alleles
- Hardy-Weinberg equilibrium

Probability of Exclusion at a locus (*l*) = $1 - \mathrm{pI}_l$

Exclusion Probability

Example:

Three alleles (a, b, c) at locus (l) each with frequency 0.2

$$pI_l = \sum_{i=1}^{n} p(A_i)^2$$

$pI_l = (0.2 + 0.2 + 0.2)^2 = (0.6)^2 = 0.36$

36% individuals will have allowable genotypes at this locus

(Equals sum of frequencies of aa, ab, ac, bb, bc and cc)

Exclusion Probability – Pros & Cons

Advantages

- Simplicity
  - ease of computation
  - ease of explanation
  - no assumptions with regard to no. contributors
- Generally Conservative?

Disadvantages

- Requires full representation of alleles
- Discards information contained in DNA profile
- Depends on allele calls (effect of stutter)

The Bayesian Approach

Odds and Probability – are interchangeable

$$O(A) = \frac{\Pr(A)}{\Pr(\overline{A})} = \frac{\Pr(A)}{1 - \Pr(A)}$$

$$\Pr(A) = \frac{O(A)}{O(A) + 1}$$

The Bayesian Approach

Requires us to define the hypotheses under test

Hp = Prosecution hypothesis – usually easy to define

Hd = Defence (alternate) hypothesis - ????

Hp: the POI is the donor of the DNA
Hd: The POI is **not** the donor of the DNA

The Bayesian Approach

What the jury do in their deliberations is to evaluate the scientific evidence under the two hypotheses put before them:

$$\frac{p(Hp|E)}{p(Hd|E)} = \frac{p(E|Hp)}{p(E|Hd)} \times \frac{p(Hp)}{p(Hd)}$$

The Bayesian Approach

What the jury do in their deliberations is to evaluate the scientific evidence under the two two hypotheses put before them:

$$\frac{p(Hp|E)}{p(Hd|E)} = \frac{p(E|Hp)}{p(E|Hd)} \times \text{prior odds}$$

Odds before the scientific evidence

The Bayesian Approach

What the jury do in their deliberations is to evaluate the scientific evidence under the two hypotheses put before them:

$$\text{Posterior odds} = \frac{p(E|Hp)}{p(E|Hd)} \times \text{prior odds}$$

Odds after (modified by) the scientific evidence

Odds before the scientific evidence

The Bayesian Approach

What the jury do in their deliberations is to evaluate the scientific evidence under the two hypotheses put before them ...

$$\text{Posterior odds} = \frac{p(E|Hp)}{p(E|Hd)} \times \text{prior odds}$$

Odds after (modified by) the scientific evidence

Odds before the scientific evidence

Likelihood Ratio
An expression of the scientific evidence

The Bayesian Approach

So how do we put a LR into words

What's the probability the semen on the vaginal swab came from our suspect?

Don't forget we're considering the probability of the evidence (E ie the DNA profile) given two alternate hypotheses so our words have to reflect this:

The probability of the evidence GIVEN ....

The Bayesian Approach

LR	Verbal wording	
1,000,000+		Support for H_p
100,000		
10,000		
1000		
100		
10		
1	Inconclusive	
0.1		Support for H_d
0.01		
0.001		
0.0001		
0.00001		
0.000001		

The Bayesian Approach

Requires us to define the hypotheses under test:

Vaginal swab + semen both sides agree V is represented

Hp = Prosecution hypothesis = pE | V + Suspect (POI)

Hd_1 = Defence hypothesis = pE | V + Unknown

Hd_2 = Alt defence hypothesis = pE | V + Suspect's brother

The Bayesian Approach

Requires us to define the hypotheses under test:

Victim alleges consensual intercourse with boyfriend and then raped. 3 person mixture from Vaginal swab + semen.

Hp = Prosecution hypothesis = pE | V + Boyfriend + Suspect

Hd_1 = Defence hypothesis = pE | V + Boyfriend + Unknown

Hd_2 = Alt defence hypothesis = pE | V + Uk1 + Uk2

Binary Model (described by Clayton et al 1998, FSI **91** 55-70)

Utilizes qualitative and quantitative data in DNA profile

- Requires detailed knowledge of performance of STR chemistry
- Detailed forensic validation & casework performance data

Method

- Assess profile
- Determine number of contributors
- List all possible genotype combinations at locus
- Generate a set of 'retained' genotypes
- If the genotype of suspected contributor is not in retained list he must be considered as EXCLUDED

General principles apply to any STR chemistry

NIJ National Institute of Justice

Binary Model - Assumptions

1. Peak heights (or areas) are proportional to amount DNA present
2. Mixture Proportion (or Ratio) is roughly constant across loci
3. If contributors share alleles their peak heights (areas) are cumulative (also known as DOSING)

As above this:

Requires detailed knowledge of performance of STR chemistry, which comes from

Detailed forensic validation & casework performance data

Binary Model – Method (The three R's)

1. Recognition (Appraisal)
    a) Designation of alleles
    b) Identification of a mixture
    c) Assessment of profile quality

2. Resolution (Deconvolution)
    a) List all possible genotypes
    b) Generate retained list of genotypes
    c) Is genotype of suspected contributor amongst retained list?

3. Reporting
    a) Formulate hypotheses
    b) Evaluate strength of evidence
    c) Reporting standards

NIJ National Institute of Justice

Step 1 - Recognition

Identifying a mixed DNA profile

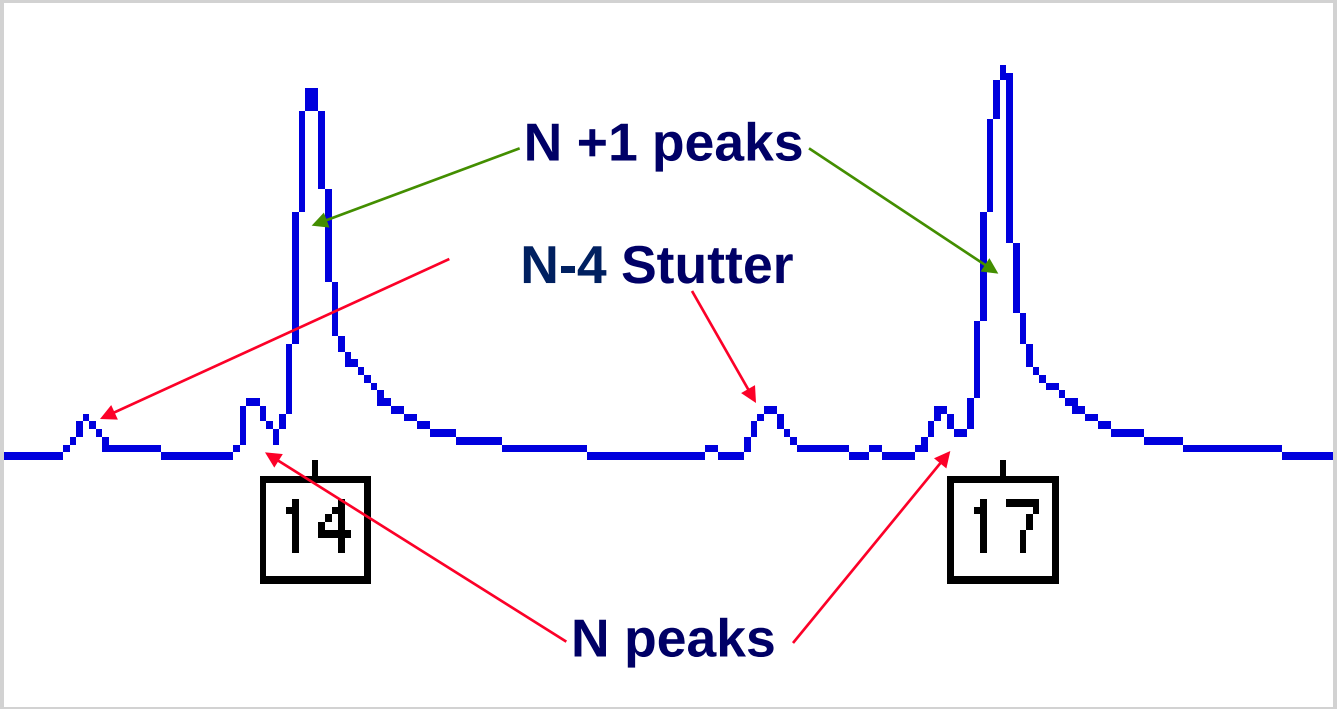
1. The presence of additional bands
    a) Does this always indicate a mixture
    b) Is it possible to have just one or two alleles at every locus?
    c) Is it likely?
    d) What else might cause additional bands

2. The presence of a pronounced heterozygous imbalance
    a) Is this always indicative of a mixed DNA profile?
    b) What else might cause heterozygous peak imbalance?

Biochemistry refresher

Potential causes of additional bands

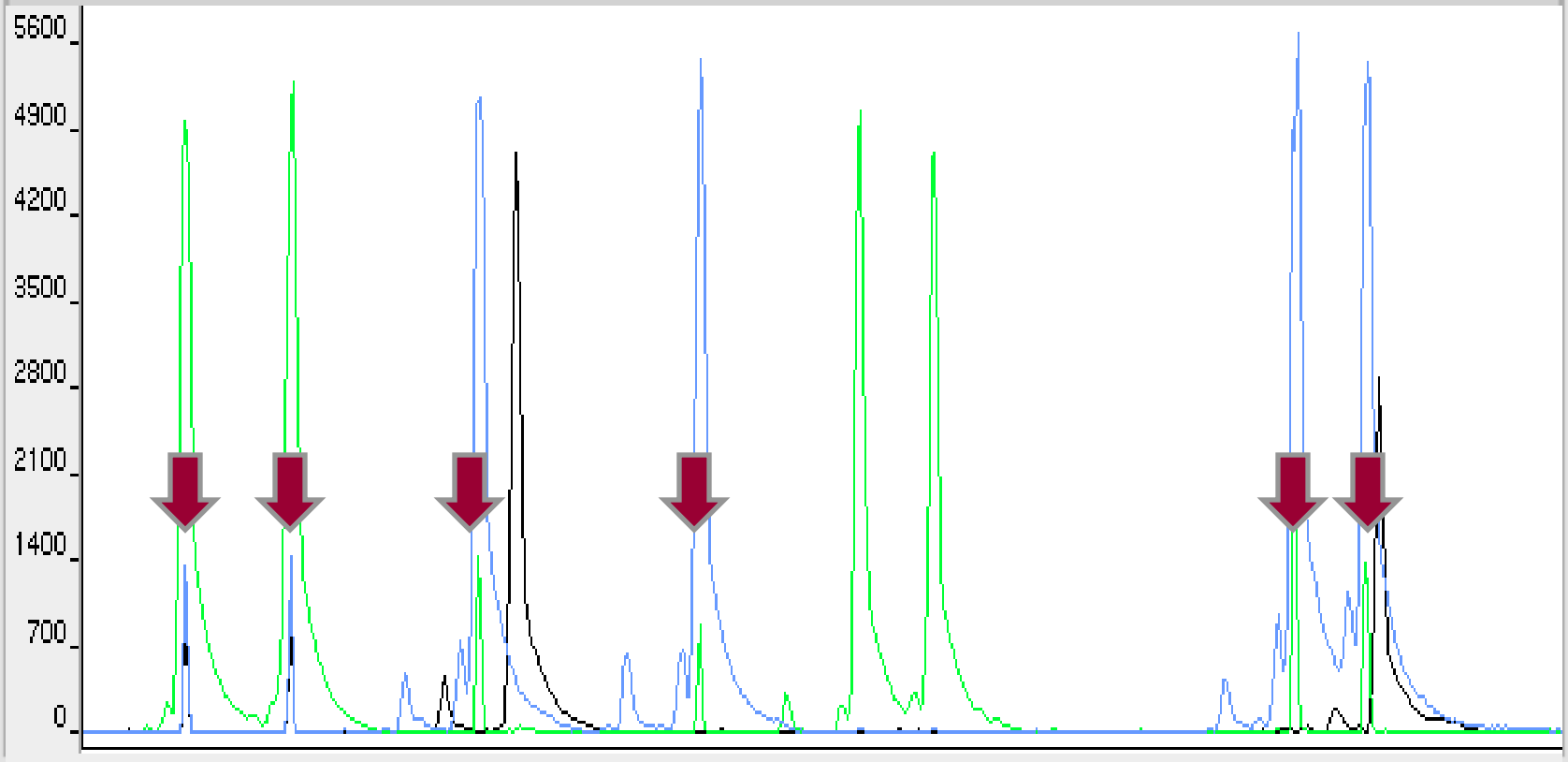
1. Stutters and N/N+1 bands



Biochemistry refresher

Potential causes of additional bands

2. Pull-up



Biochemistry refresher

Potential causes of additional bands

3. Chromosomal abnormalities (allelic mutations)

    a) Chromosomal duplications
    b) Trisomies (Aneuploides)
    c) Somatic mutations or mosaicism

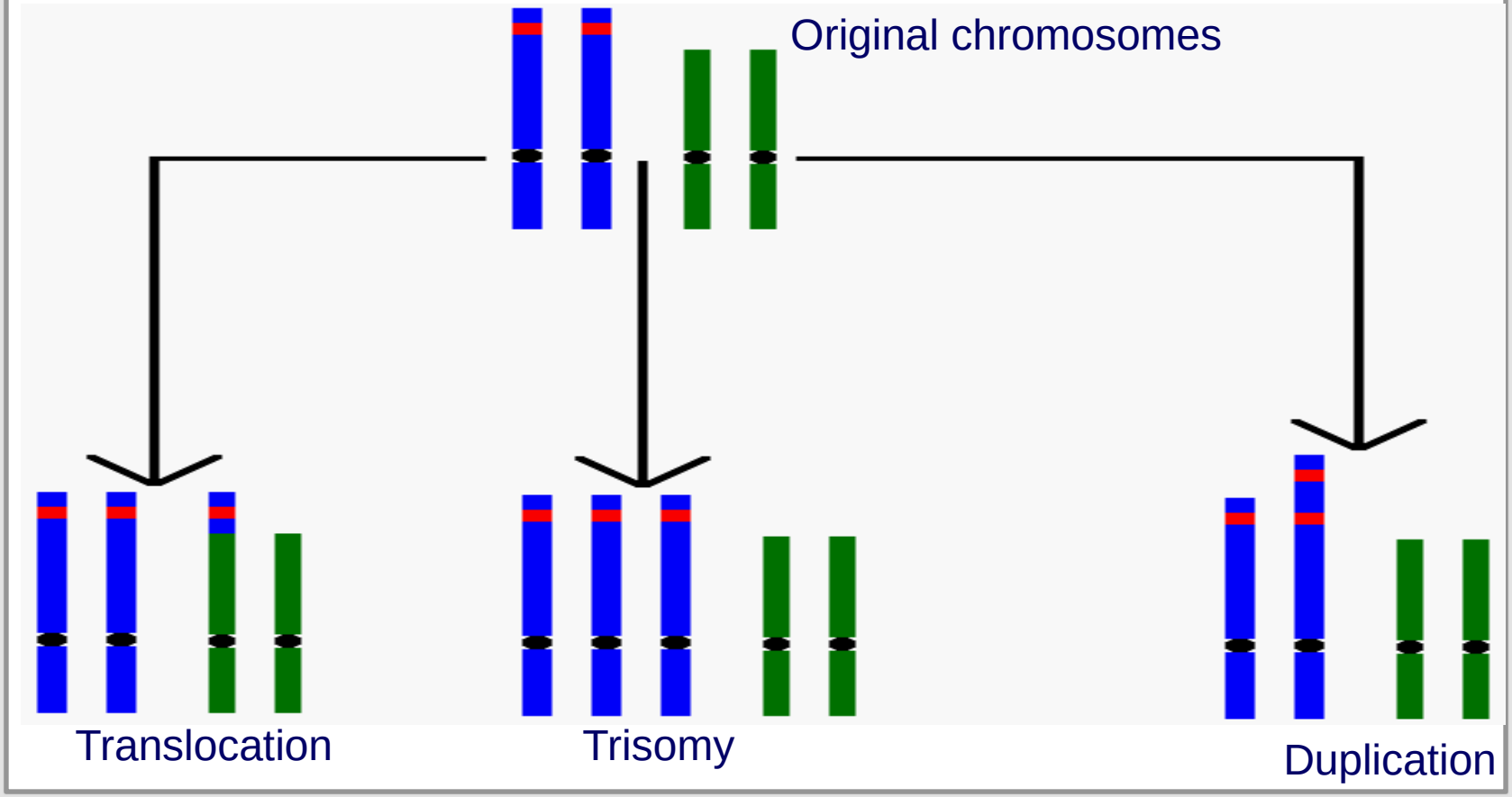
Single locus and Rare (also seen in Reference sample)

Peak heights 1:1:1 and may serve to strengthen match

XY aneuploidies may be more common (XXY or XYY)

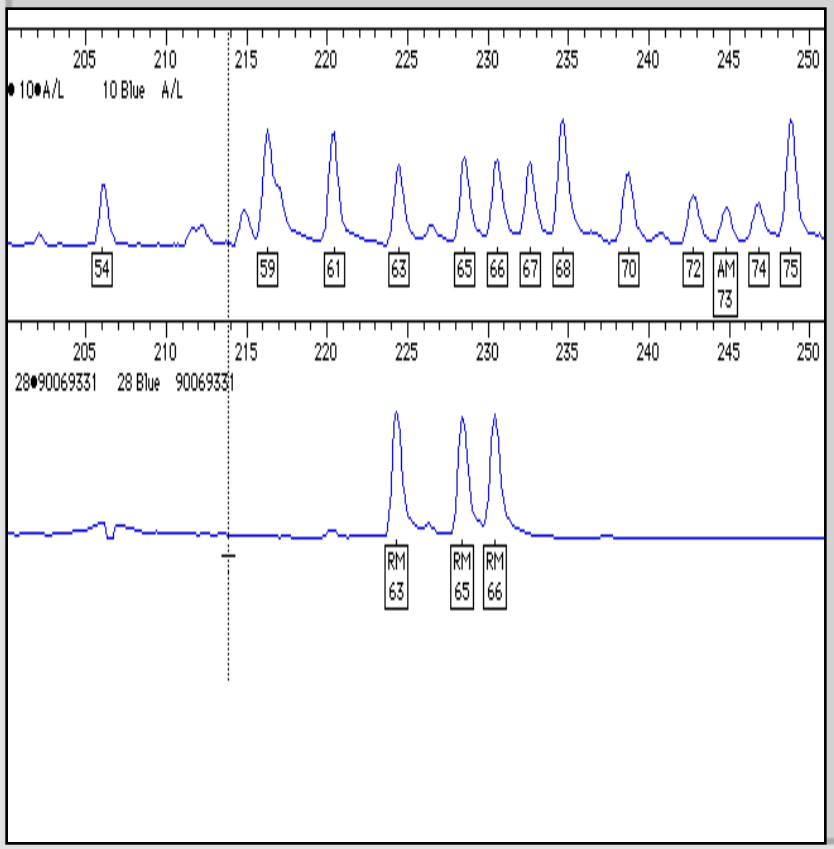
Biochemistry refresher

Chromosomal rearrangements

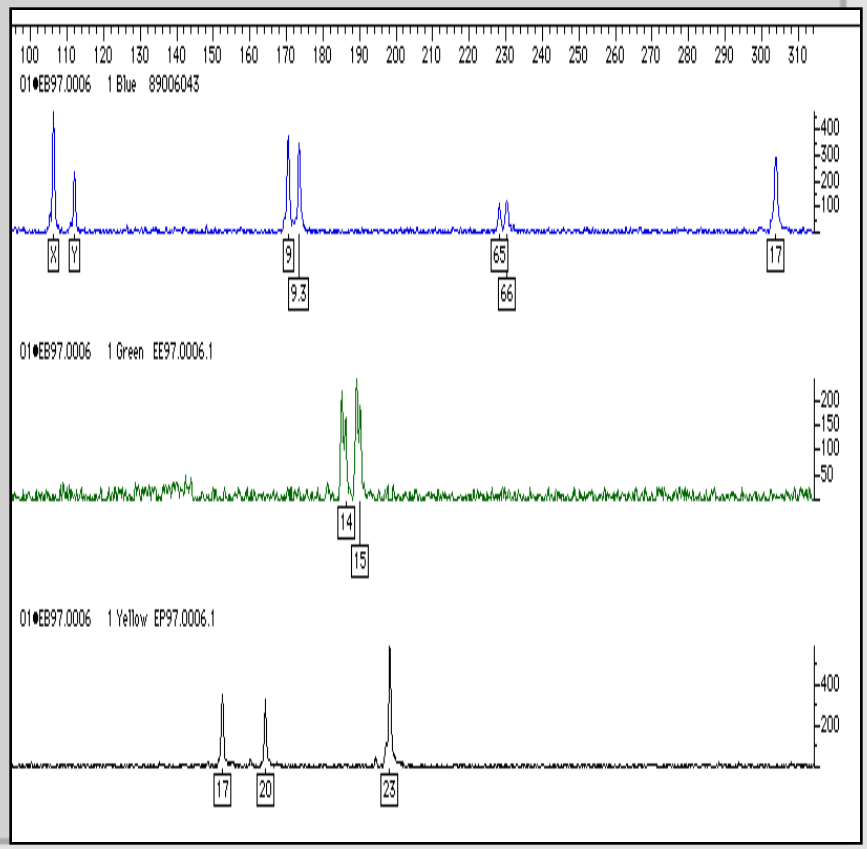


Biochemistry refresher

D21 Trisomy

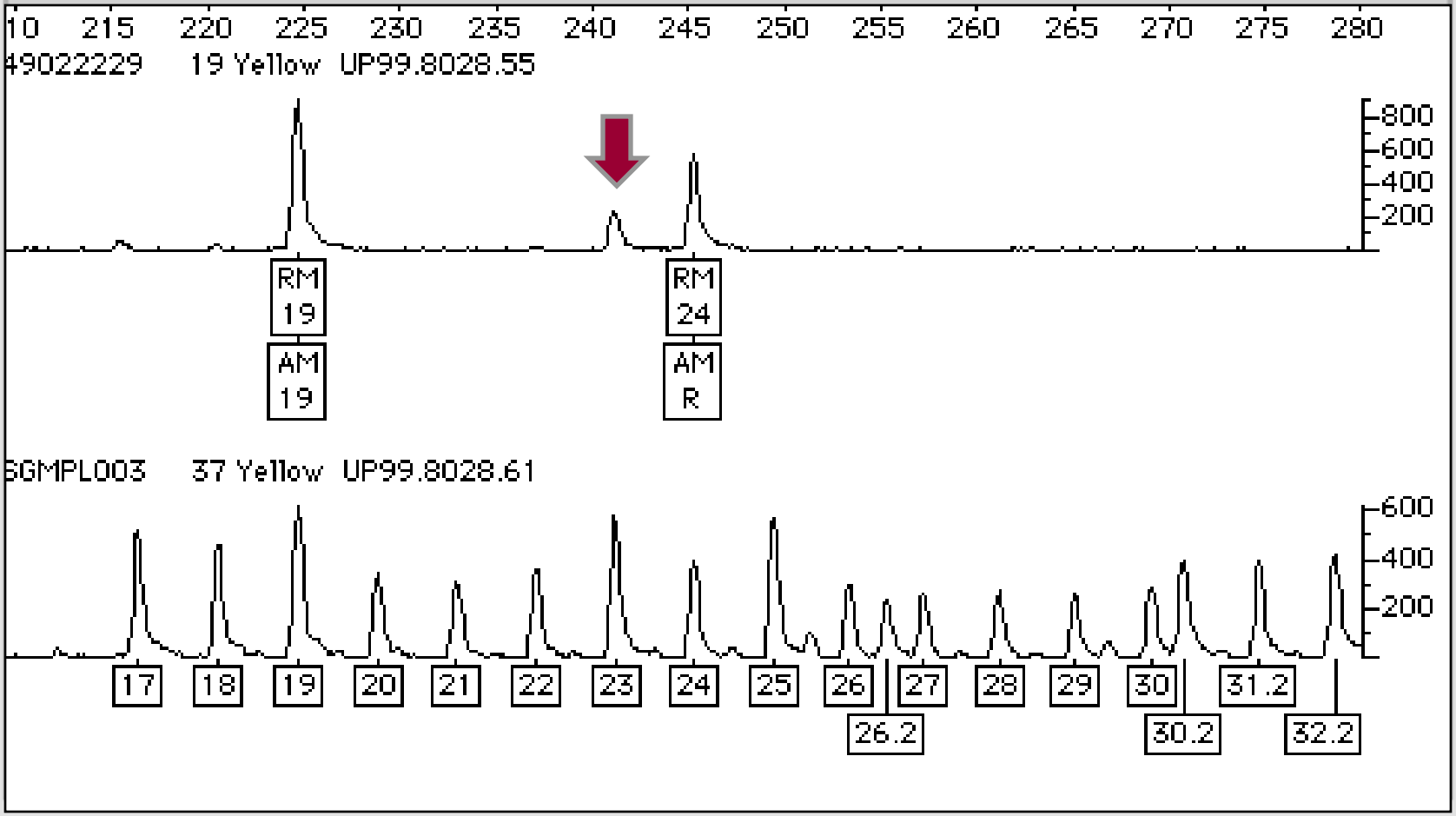


Kleinfelters (XXY)



Biochemistry refresher

Somatic Mutations (or mocaisism)



Step 1 - Recognition

Identifying a mixed DNA profile

2. The presence of a pronounced heterozygous imbalance

    a) Unequal amplification efficiency - Processivity of Taq polymerase

    b) Degraded template DNA

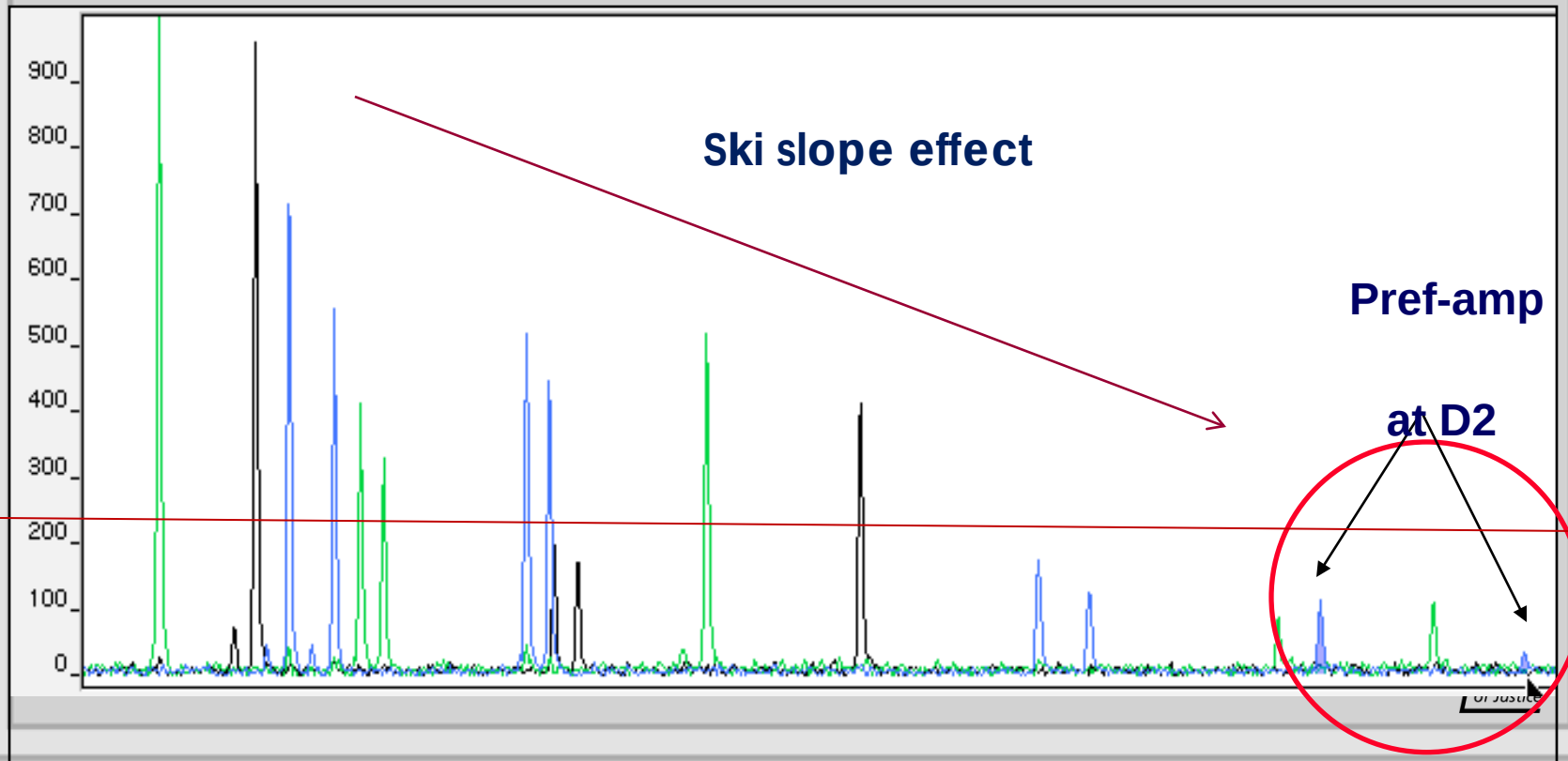
    c) Stochastic variation

    d) Primer binding site mutations

Biochemistry refresher

Potential causes of heterozygote peak imbalance

2. Degradation of template DNA



Biochemistry refresher

Potential causes of heterozygote peak imbalance

3. Stochastic variation – Lt DNA effect

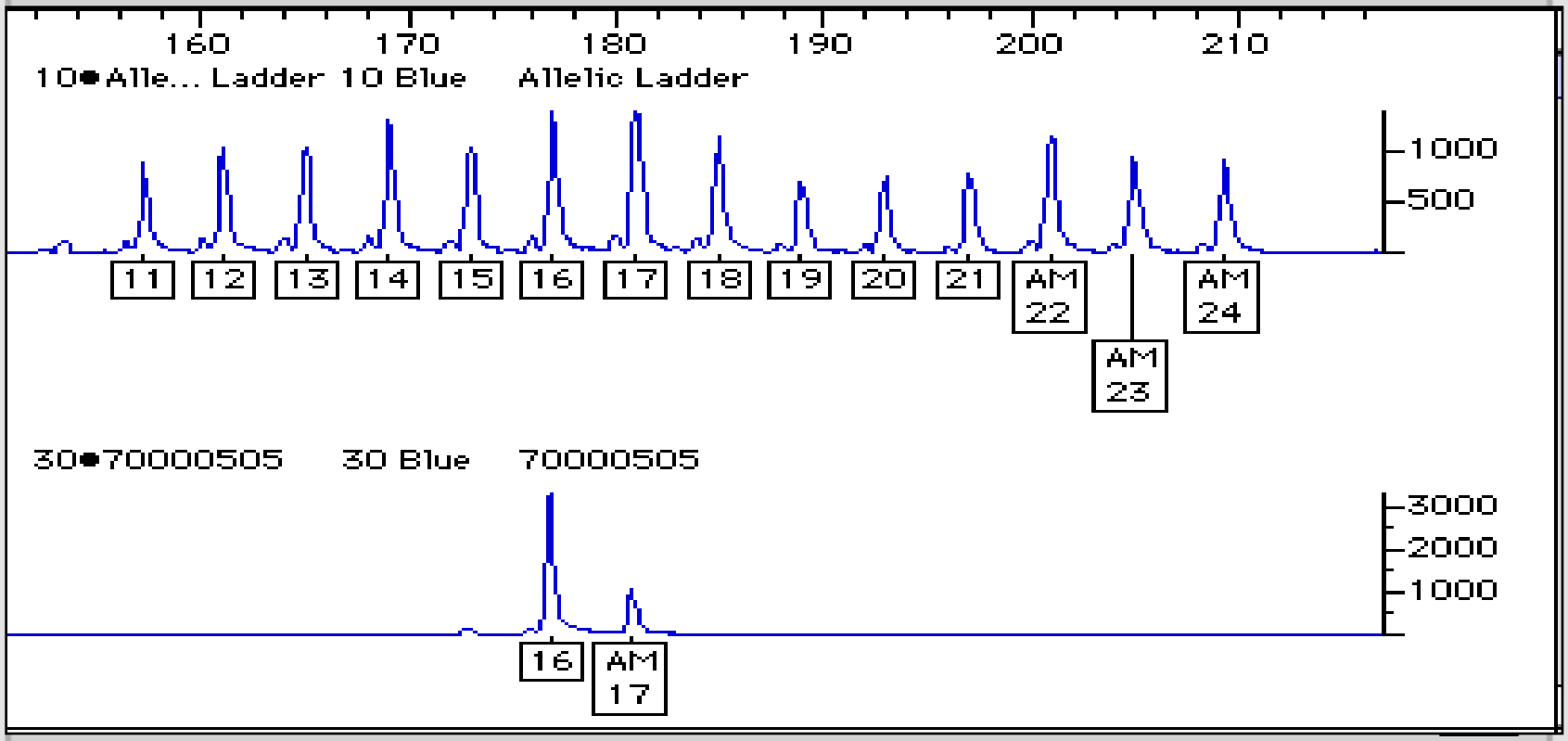
    a) Unequal numbers of template molecules due to sampling variation

    b) PCR process preserves asymmetry

Biochemistry refresher

Potential causes of heterozygote peak imbalance

4. Primer binding site mutations



Step 1 – Recognition: Recap

Identifying a mixed DNA profile

1. Is it a mixture – extra bands and/or peak imbalance

2. Assessment of the DNA profile – Quality appraisal
   a) Amplification efficiency – peak height (rfu)
   b) Degraded?
   c) Is it a low level mixture – does it matter?

3. Do the case circumstances allow conditioning?

Step 1 – Recognition: Classification

Classification of mixed DNA profiles

1. GREEN – allelic peaks around or above 400 rfu

2. AMBER – peak height between 150 and 400 rfu. Lab guidelines become progressively less robust

3. RED – peak heights below 150 rfu - potential for incomplete representation, drop-out, drop-in high stutters, exaggerated peak imbalance

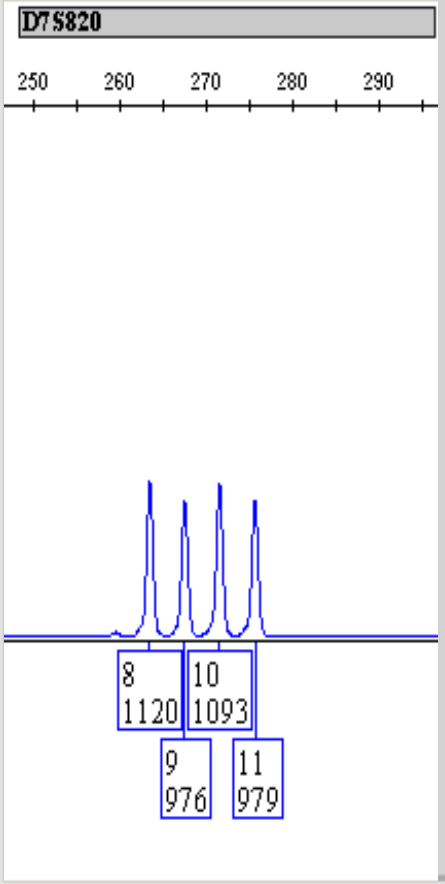
NIJ National Institute of Justice

Step 2– Resolution

NOTE Unless conditioning on known profile this step is conducted without knowledge of DNA profiles of any reference samples

Step 2– Resolution

Conditioning depends on circumstances



Two contributors

Scenario 1 – Vaginal Swab + semen; donor 8, 10

What's the genotype of the male contributor?

Scenario 2 – i/s front male underpants; donor 8, 10

What's the genotype of the second contributor?

Scenario 3 – mixed body fluid stain from a carpet

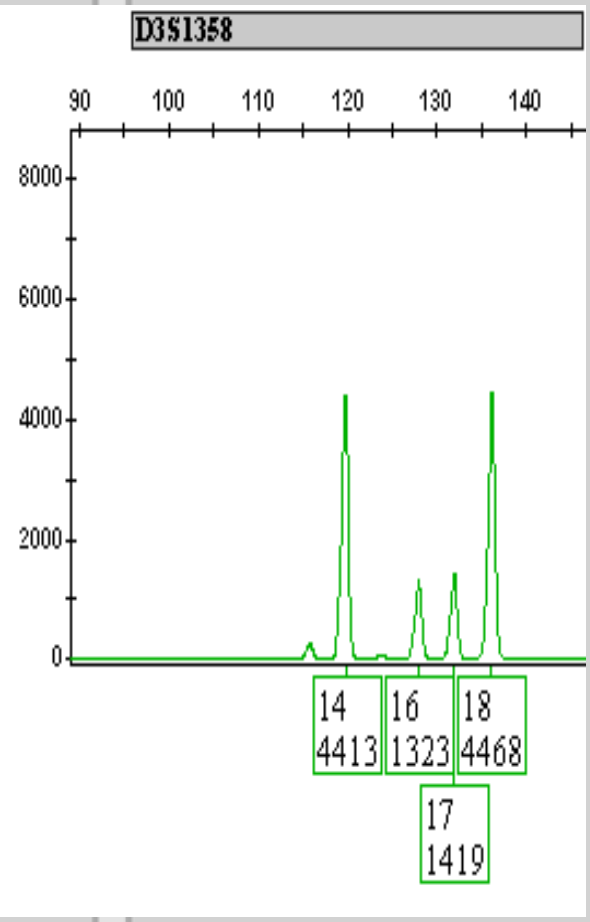
Can we condition on anything?

Step 2– Resolution

Logical progression for the analysis

1. Assess the MINIMUM number of contributors (no. bands)
2. Estimate the mixing proportion (or ratio) – M_X
3. Using observed alleles list all pairwise combinations
4. Use peak height/area data and M_X to ELIMINATE genotypes not supported by the data
5. Compare with reference samples

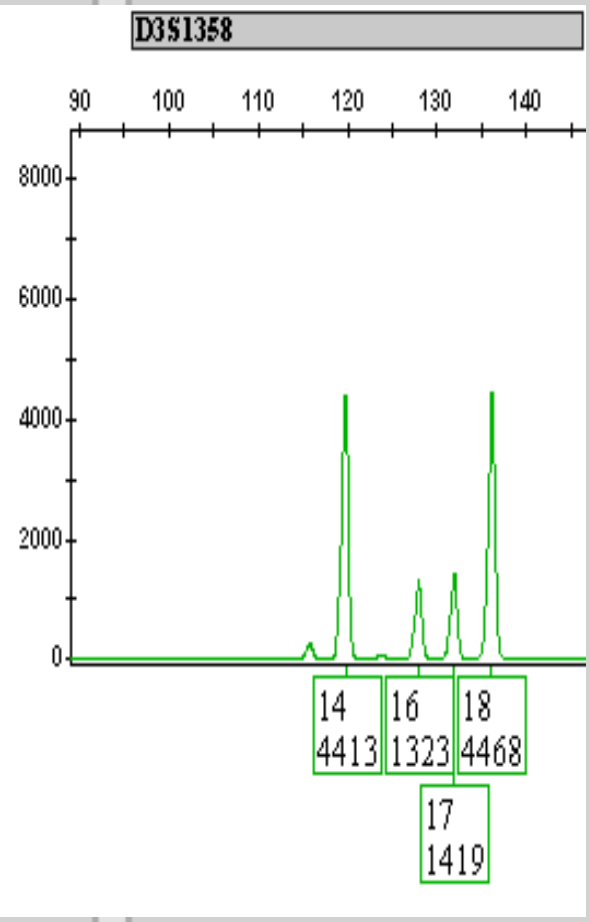
Step 2– Resolution



Logical progression for the analysis

1. Assess the MINIMUM number of contributors
2. Estimate the mixing proportion (or ratio) – M_X
3. Using observed alleles list all pairwise combinations
4. Use peak height/area data and M_X to ELIMINATE genotypes not supported by the data
5. Compare with reference samples

Step 2– Resolution



Logical progression for the analysis

Major Contributor	Minor Contributor
14, 18	16, 17

5. **Compare with reference samples**

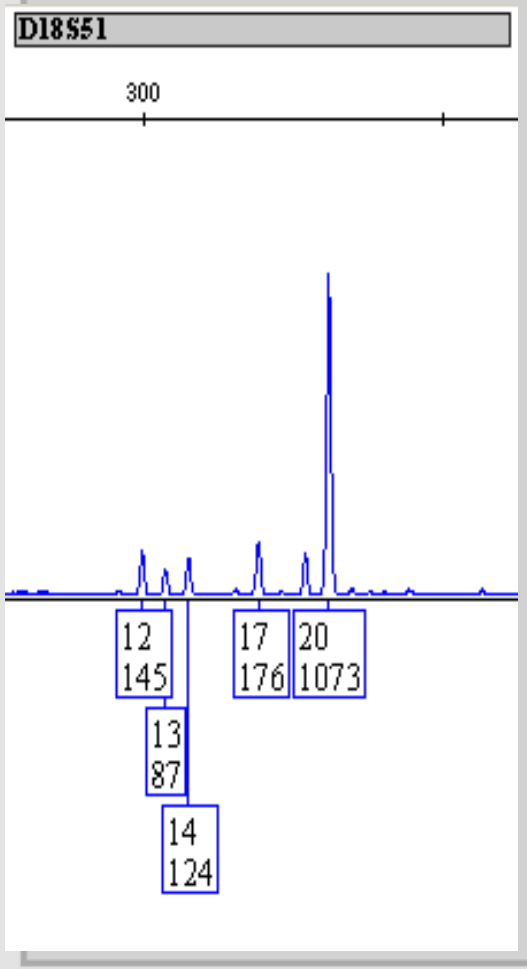
Step 2– Resolution

Logical progression for the analysis

1. Assess the MINIMUM number of contributors (no. bands)
   a) If there are two contributors what is the maximum no. bands?
   b) If you see five or six bands at a locus what does this imply?

2. Mixing Proportions (M_X) or Mixing Ratio (M_R)

Step 2– Resolution



How many contributors?

Three persons or more?

Is the peak in the 19 position a stutter or an allele?

Is this a major 20, 20 with other low level contributors

Could this be a 5 person mixture?

(12, 20 + 13, 20 + 14, 20 + 17, 20 + 19, 20)

Step 2– Resolution

There are three fundamental observations:

1. Within a locus the proportion peak height/area reflects the amount of template DNA from each contributor

2. If bands are shared between contributors the peak height/areas are cumulative (approx) – DOSING

3. In a well amplified (GREEN) mixed DNA profile the mixing proportion (M_X) is fairly constant

Step 2– Resolution

Mixing Ratio (MR)	Contributor 1 (MX)	Contributor 2 (MX)
10:1		
5:1		
4:1		
3:1		
2:1		
1:1		
1:2		
1:3		
1:4		
1:5		
1:10		

Step 2– Resolution

The Amelogenin locus may indicate a MALE/FEMALE mixture

Ratio of Components		Dosage of observed products	
Male (X,Y)	Female (X,X)	X allele	Y allele
10	1		
5	1		
4	1		
3	1		
2	1		
1	1		
1	2		
1	3		
1	4		
1	5		
1	10		

Step 3 – List all genotype combinations

Take the observed allele designations and list out all pair-wise combinations.

1. Try A,B,C,D — 6 pair-wise combinations
2. Try A,B,C — 12 pair-wise combinations
3. Try A,B — 7 pair-wise combinations

See: Component Pairing exercise sheets

Step 4 – Generate retained genotype list

We only want to retain those genotype combinations which are supported by the observed data.

At each locus:

1. Assess every genotype combination against:
    a) Consistency with the observed Mixing Proportion (M_X)
    b) Observed Heterozygote peak imbalance
2. Mark every genotype combination as Pass or Fail
3. Generate retained genotype list

Step 4 – Generate retained genotype list

Try These

1.	Four alleles at a locus	A, B, C, D
2.	Three alleles at a locus	A, B, C
3.	Two alleles at a locus	A, B
4.	One allele at a locus	A, A

Step 4 – Generate retained genotype list

1. Four alleles at a locus A, B, C, D

Contributor 1	Contributor 2

Step 4 – Generate retained genotype list

2. Three alleles at a locus A, B, C

Contributor 1	Contributor 2
A, A	B, C
A, B	A, C
A, B	B, C
A, B	C, C
A, C	B, B
A, C	B, C
A, C	A, B
B, C	A, B
B, C	A, C
B, C	A, A
B, B	A, C
C, C	A, B

Step 4 – Generate retained genotype list

3. Two alleles at a locus A, B

Contributor 1	Contributor 2
A, A	B, B
A, A	A, B
A, B	A, A
A, B	B, B
A, B	A, B
B, B	A, B
B, B	A, A

Step 4 – Generate retained genotype list

4. One alleles at a locus A, A

Contributor 1	Contributor 2
A, A	A, A

Step 4 – Generate retained genotype list

Software support?

Pref Amp Tolerence	Mixing Proportion Tolerence	Homozygote
60%	15%	60

Weight: Maximum	Weight: Minimum	Weight: Mean
18% 5:1	14% 6:1	16% 5:1

Database Consolidation is on

-			Possible Contributors				Pref Amp Rule				Mix Prop Rule		-				
Locus	Allele	Area	Contributor 1		Contributor 2		Contributor 1		Contributor 2		Mix Est		RC	Contributor 1		Contributor 2	
D3S1358	15	5367	16	16	15	15	-	Y	-	Y	18% 5:1	Y	Include	16	16	15	15
	16	24242	15	16	15	15	22%	N	100%	Y	-64% >10:1	N	-	-	-	-	-
	-	-	15	16	15	16	22%	N	22%	N	-	Y	-	-	-	-	-
	-	-	16	16	15	16	100%	Y	452%	Y	36% 2:1	N	-	-	-	-	-
	-	-	15	15	16	16	-	Y	-	Y	82% 1:5	N	-	-	-	-	-
	-	-	15	15	15	16	100%	Y	22%	N	164% <1:10	N	-	-	-	-	-
	-	-	15	16	16	16	452%	Y	100%	Y	64% 1:2	N	-	-	-	-	-
	Database Consolidation for D3S1358													16	16	15	15

http://www.promega.com/profiles/802/ProfilesinDNA_802_08.pdf

Step 5 – Compare with Reference Samples

You have defined your retained genotype list – you cannot change your mind when the reference samples are revealed

1. If the suspect's DNA profile does not match one of the retained genotypes he is EXCLUDED as a contributor

2. If suspect's DNA profile does match one of the retained genotypes this should be confirmed at all loci

3. This process should be repeated by second scientist as a blind exercise to ensure objectivity

4. This type of analysis is built into some DNA Expert systems (FSS-i^3)

Summary

Outline models – particularly RMNE and binary methods

Recognizing DNA mixtures – extra bands, Het peak imbalance

Biochemistry refresher – other things cause artifacts

Resolving mixtures – series of logical steps

NB. Must be objective – so no sneaky peeks at suspect profile

Contact Information

Dr Chris Maguire

Northumbria University Centre for Forensic Science
Ellison Building Rm. D104
Northumberland Road
Newcastle upon Tyne
NE1 8ST

+44 191 243 7627
chris.maguire@northumbria.ac.uk