REVIEW OF BLACKBURN DNA ANALYSIS

for the
Commission of Inquiry into Forensic DNA Testing in Queensland

Dr Kirsty Wright

Version 2

STRMix™ and PowerPlex® 21 are commercial products discussed throughout this report. Both products provide high quality and reliable outcomes for forensic practitioners globally. Criticism of results produced by both products within this report are solely targeted at how Queensland Health and Forensic Science Services DNA Analysis Unit have implemented and used the technology.

Contents

Abbreviations

Abbreviation	Definition
AFP	Australian Federal Police
ANZPAA	Australia New Zealand Policing Advisory Agency
C2	Contributor two
DNA	Deoxyribonucleic acid
FASS	Forensic and Analytical Science Service
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LOD	Limit of detection
LOR	Limit of reporting
LR	Likelihood ratio
NATA	National Association of Testing Authorities
NCIDD	National Criminal Investigation DNA Database
ng	Nanogram
NIFS	National Institute of Forensic Science
OQI	Opportunity for quality improvement
PCR	Polymerase chain reaction
pg	Picogram
QHFSS	Queensland Health Forensic and Scientific Services
QPS	Queensland Police Service
QPRIME	Queensland Police Records and Information Management Exchange
RFU	Relative fluorescent units
ul	Microlitre

1.0 Summary

A range of documents were provided by the Commission of Inquiry for this review including the BLACKBURN DNA case file and full audit trail, internal validation studies and standard operational procedures. The aim of this review was to identify errors and issues in DNA analysis, interpretation, and reporting of evidence by the Queensland Health Forensic and Scientific Services DNA Analysis Unit (QHFSS) that may have affected the BLACKBURN matter, and evidence for other matters. Likely causes of errors are identified where possible. A series of recommendations are provided to enable more thorough investigation by independent experts (see Appendix A), with the overall aim of facilitating a pathway to reform and ensuring the provision of high-quality DNA evidence in Queensland.

The BLACKBURN matter was characterised by the lack of physical evidence linking [redacted] to the crime. Many crime scene samples were collected by police, but unexpectedly failed to provide DNA results. Defence offered an alternative offender, [redacted]. A scenario that was supported by a DNA mixture from BLACKBURN's pants, which QHFSS reported [redacted] as a possible contributor. It is unknown whether this evidence influenced the outcome of the trial, though it certainly supported the defence proposition of an alternative offender.

There is evidence of a poor quality culture, recklessness, and poor scientific practices within QHFSS that may have persisted for at least a decade. Serious and systemic flaws have affected QHFSS' ability to generate DNA profiles and accurately interpret them. QHFSS introduced two key processes only weeks before BLACKBURN's murder, which were validated with flawed data leading to the incorrect setting of two key thresholds and affected the accuracy of some results. However, once the error was discovered QHFSS failed to change the thresholds, re-analyse evidence, and the error was not revealed in an updated validation report. The key evidence from BLACKBURNS pants, if re-analysed with the correct thresholds, would likely result in [redacted] not being included as a contributor.

There is evidence of a catastrophic failure in QHFSS DNA profiling processes around the time 103 BLACKBURN crime scene samples were analysed which affected evidence over two to three months. It is unknown if BLACKBURN's evidence was affected. There is no record of two key suspects' DNA being compared to BLACKBURN's crime scene evidence.

A summary of issues discovered during this review is provided below.

Table 1: Summary of QHFSS DNA analysis issues.

#	Issue
1	A faulty dishwasher and defective Proteinase K had a catastrophic impact on crime scene evidence over two to three months. A period when 103 BLACKBURN crime scene samples were analysed.
2	BLACKBURN evidence with 'no DNA detected' (43 samples)[1], including samples expected to provide results.
3	BLACKBURN evidence with unexplained DNA degradation (4 samples).
4	BLACKBURN evidence with unexplained weak profiles (3 samples).
5	QHFSS fail to adequately investigate concerning results when police query the lack of DNA from key samples.
6	A warning of unexplained failure of semen samples in a 2012 QHFSS standard operational procedure suggests this occurred regularly.
7	The Krosch 2021 paper contains evidence of unexplained systemic failure of sexual assault samples analysed in 2018 and 2019 (52% failure of penis samples).
8	QHFSS use half-volume PowerPlex 21 reactions to reduce costs, despite the manufacturer recommending the full-volume reaction and advising decreasing the reaction volume can result in sub-optimal performance.
9	QHFSS quantitation thresholds were between 2.5 and 6.1 times higher than NSW at the time BLACKBURN evidence was analysed.
10	An incorrect instrument setting was used to analyse crime scene evidence for several months[2], and importantly was also used in the PowerPlex 21 and STRmix internal validations. Parts of the validation studies are invalid, potentially affecting the reliability of evidence analysed with these methods.
11	The flawed data used in the 2012 PowerPlex 21 and STRMix internal validations resulted in two critical thresholds being incorrectly set (limit of reporting and drop-in thresholds).
12	The re-issue of the PowerPlex 21 internal validation in 2013 still contained the flawed data, and did not identify the original error.

[1] Confirmation is needed on whether these samples were affected by OQI#3403 or some other faulty process.

[2] QHFSS re-analysed these samples, therefore there was no loss of information.

13	QHFSS fail to change the critical thresholds and continue analysing crime scene samples knowing the thresholds are incorrect, risking the reliability of evidence presented to the court.
14	All BLACKBURN crime scene samples were analysed with the incorrect settings causing errors. Note: these errors would affect the reported likelihood ratio, though it is unknown if it would also have changed the verbal scale used by QHFSS. It is likely to change who is reported as a contributor or non-contributor of low-level DNA mixture evidence.
15	If re-analysed with the correct thresholds, [redacted] would likely not be reported as a contributor to the key evidence (L45) from BLACKBURN's pants.
16	There is evidence in the STRMix internal validation that QHFSS were aware it was not providing expected results and causing incorrect outputs. This is likely due to the incorrect user-defined thresholds and use of half-volume reactions. QHFSS decided to implement the procedure knowing it risked incorrect results.
17	QHFSS' internal STRMix validation of four-person mixtures fails due to use of low specification computers.
18	Inaccurate classification of single contributor partial profiles as complex mixtures that cannot be interpreted (4 BLACKBURN crime scene samples).
19	Incorrectly using unlabelled sub-threshold information post-STRMix analysis to exclude 78 possible contributors and non-contributors of DNA mixtures (6 BLACKBURN crime scene samples).
20	Incorrectly reporting a mixture as unsuitable for 'meaningful interpretation' that could be analysed with STRMix for one BLACKBURN sample.
21	Inclusion of an artifact peak (poor quality information) in DNA mixture analysis of one BLACKBURN sample.
22	From July 2014 to January 2015 QHFSS blamed a 'minor mis-code' on incorrect evidence presented in court. The STRMix developers state QHFSS did not buy the updated user manual and query if they were following recommended processes.
23	In 2019 QPS reviewed cases due to uncertainty about the number of contributors within DNA mixtures for results reported between 2013 and 2018. A total of 138 cases were identified where DNA evidence was potentially significant to the prosecution. This is further proof QHFSS are not reliably interpreting DNA mixtures.
24	There is no record of two of the six key suspects' DNA being compared to BLACKBURN crime scene DNA.

2.0 Faulty dishwasher (OQI#34043)

2.1 Background

Proteinase K is a critical enzyme used in the DNA extraction process and therefore crucial to obtaining profiles from crime scene samples. Proteinase K degrades proteins in the cell wall so DNA molecules can be released into solution for subsequent profiling. Vitally, Proteinase K also inactivates nucleases, which are enzymes contained within cells that degrade DNA when released. To prevent degradation of crime scene DNA nucleases must be inactivated by Proteinase K during the extraction process. Proteinase K is most stable at pH 8, however has a working range between pH 4 and pH 12.5.

A document labelled OQI#34043 was included in the BLACKBURN DNA case file[3] and reports that one in-house prepared lot of Proteinase K was found to have a pH of 14. The opportunity for quality improvement (OQI) indicates the likely cause was an industrial dishwasher used to clean laboratory glassware not operating to full specification, and the glassware being contaminated with caustic detergent[4]. This issue was identified on 22 March 2013. When the defective Proteinase K was added to DNA extraction chemicals, the pH range was 11-12. The ramifications of this are:

1. Proteinase K enzymes are likely to be either inactivated or severely compromised at pH 14. Leading to:
    a. reduction in the quantity of DNA released from cells into solution for subsequent profiling, and
    b. reduction or prevention of nuclease deactivation leading to degradation of crime scene DNA.
2. DNA discarded during the pH dependent extraction method[5].

In addition, the Microcon concentration process (used by QHFSS to increase the chance of obtaining a profile from small quantities of DNA) is affected by pH causing adsorption of DNA on the filter membrane, and subsequent loss[6]. Further investigation is required to determine if the pH range of the affected extract would cause significant DNA losses.

[3] File 1, p1920
[4] Containing highly alkaline and corrosive chemicals.
[5] DNA binds to magnetic beads under pH7.5 during the initial stage of extraction, preventing it from being discarded during 'wash' steps.
[6] Microcon Centrifugal Devices User Guide.

Typically, an OQI is included in a case file if samples relating to that case are affected by the issue. The affected samples are usually listed in the case file, or as stated in OQI#34043 "*appropriate AUSLAB audit entries & notes have been made for all affected samples / batches"*[7]. There are no records against any crime scene samples in the BLACKBURN case file, however, handwritten notes exist for this OQI against four reference samples[8]. Reference samples normally undergo a different process to crime scene samples that does not require Proteinase K, however, these samples are listed as 'EREF' or 'MCONR'[9] on the electropherograms meaning they failed the reference profiling process so underwent the same extraction process as crime scene samples.

The four reference samples were received on 28 February 2013. It may have been at least a week before they were DNA extracted after failing the reference sample process, indicating the defective Proteinase K may have been in use early March.

The date range of this issue is not reported in the OQI, neither are the number of samples affected which are significant omissions. Incredibly, the OQI states that the defective Proteinase K was accidently used for some subsequent DNA extractions "*resulting in additional samples being affected"*.[10] Once again, there are no date ranges, however, the OQI notes this was occurring as investigation of the issue was nearing completion (the investigation was completed on 6 May 2013). Many hundreds of crime scene samples may have been affected by this issue over two to three months. Other glassware would have been affected by the faulty dishwasher during the undefined period it was contaminating items with caustic detergent. Concerns have previously been raised about the unexplained lack of DNA in the BLACKBURN matter. The first crime scene evidence was received 11 February 2013, only six weeks prior to the discovery of the defective Proteinase K. Given 108 crime scene samples were received by QHFSS prior to the faulty dishwasher issues being discovered, and before the defective Proteinase K was completely removed from use, it is possible some evidence was affected. These samples include those most likely to identify the offender.

Recommendation 1: A root cause analysis needs to be conducted by an independent quality expert to confirm the period of the faulty dishwasher use and determine how many laboratory processes and samples across all cases were affected. Those crime scene samples should be evaluated for further testing and where needed addendum statements released.

[7] File 1, p1919, paragraph 6.
[8] File 1, p1685, 1569, 1563, and 1539.
[9] EREF = Extraction Reference, meaning they underwent the same extraction process as crime scene samples because they initially failed using the process used for reference samples. MCONR = Microcon concentration of a reference sample.
[10] File 1, p1919, paragraph 5.

Table 2: BLACKBURN crime scene evidence received by QHFSS prior to the faulty dishwasher being reported and the defective Proteinase K completely removed from use.

Date Received by QHFSS	**#Samples**	**Sample Description**
11 Feb 2013	5	[redacted]' covert cig butt, Girl Guides cig butts, samples from BLACKBURN's right wrist, left and right fingernail scrapings
13 Feb 2013	25	Cig butt from drain, samples from BLACKBURN's body, samples from BLACKBURN's shirt and pants
14 Feb 2013	3	BLACKBURN's left and right fingernail clippings
22 Feb 2013	3	Samples from BLACKBURN's right and left forearm, and forehead
27 Feb 2013	30	[redacted]' car samples (including 12 x 'blood' samples), samples from BLACKBURN's shirt, pants and phone, white T-shirt with 'blood stains' found near crime scene
6 March 2013	25	Blood from gutter (including S14), samples from BLACKBURN's hairband, shirt, shoes and pants
22 March 2013	Faulty dishwasher and defective Proteinase K identified.	
12 April 2013	7	Knife samples
22 April 2013	10	BLACKBURN's shirt
Early May 2013	Defective Proteinase K completely removed from use	
Total	**108**	

If any crime scene evidence was impacted by the defective Proteinase K, the results of analysis may include:

1. 'no DNA detected' being reported;
2. a reported DNA profile, however, the raw results (electropherogram) would indicate unusually degraded DNA; or
3. a reported DNA profile, however, the profile would be weaker than expected, and would risk low-level contributors not being detected.

Evidence of all these issues exist in the BLACKBURN DNA case file. Certainly, the detectives investigating BLACKBURN's murder were not aware of the catastrophic issues occurring within QHFSS when most of the crime scene evidence was being analysed.

2.2 'No DNA detected'

There were 43 crime scene samples from the BLACKBURN case reported as 'no DNA detected', and therefore not fully tested. These include:

1. S14: Swab of blood from gutter on Boddington Street
2. Three samples of 'blood-soaked fabric' from a T-shirt found near the crime scene
3. Twelve samples from [redacted]' vehicle reported as presumptively positive for blood.

A total of 33 samples from [redacted]' vehicle were submitted for DNA testing, and 27 provided 'no DNA detected'. The absence of [redacted]' DNA on all of the vehicle's internal surfaces is unexpected.

The QHFSS Procedure for Case Management (v16) explains that samples with a quantitation concentration less than 0.0021 ng/ul are not fully tested regardless of priority and reported as 'no DNA detected'[11]. This is a valid interpretation of the quantitation results, however, means there is an unexplained absence of DNA in 43 crime scene samples, some of which should have produced DNA profiles.

The results of these samples were reported between 8 March 2013 and 2 April 2013, suggesting they may have undergone DNA extraction in the timeframe the faulty dishwasher and defective Proteinase K were in use. However, apart from S14 (taken from a visible pool of blood), it cannot be confirmed if the other samples contained biological material.

Table 3: Dates key crime scene evidence was reported in QPRIME as 'no DNA detected'. The dates they underwent DNA extraction may overlap with the faulty dishwasher (likely 5 to 10 days prior to being reported).

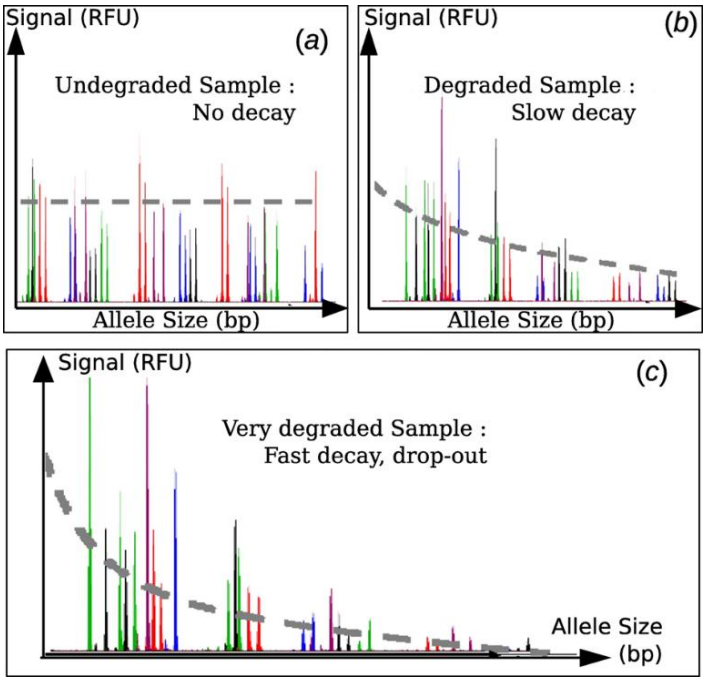
Sample	Date Reported	Sample	Date Reported
S14:'Blood from gutter	13 March 2013	V31:'Blood' door handle	8 March 2013
ML2:'Blood-soaked fabric'	11 March 2013	V32:'Blood' window wind	8 March 2013
ML4:'Blood-soaked fabric'	8 March 2013	V33:'Blood' handle to door	8 March 2013
ML 5:'Blood-soaked fabric'	2 April 2013	V34:'Blood' Door trip	8 March 2013
V14:'Blood' handbrake well	8 March 2013	V48:'Blood' steering wheel	8 March 2013
V15:'Blood' clutch pedal	8 March 2013	V49:'Blood' ignition	8 March 2013
V16:'Blood' brake pedal	8 March 2013	V50:'Blood' driver's seat	11 March 2013
V17:'Blood' accel. Pedal	8 March 2013	V51:'Blood' footwell	11 March 2013

[11] p7, paragraph 9, and p17, paragraph 10.

2.3 Unexplained degraded DNA

There is evidence of unexplained degraded crime scene DNA from the BLACKBURN case. Undegraded samples have peaks of consistent size displayed from left to right in an electropherogram. Electropherograms from degraded samples have tall peaks representing shorter DNA fragments in the left of the graph, and small peaks representing longer DNA fragments (right side of graph). The 'ski-slope' curve is an indication of DNA degradation, where longer fragments of DNA are most affected. Degradation typically causes breaks in the DNA strands. The strands become shorter as degradation progresses. While DNA degradation can occur naturally over many years or under extreme environmental conditions, the crime scene samples in the BLACKBURN case do not meet those criteria.

Figure 1: Electropherograms displaying a) a normal DNA profile, b) and c) artificially degraded DNA.[12]



There are numerous electropherograms that indicate the BLACKBURN crime scene DNA has suffered unexplained degradation. They all fall within the date range of the faulty dishwasher and use of the defective Proteinase K.

[12] https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3074-0/figures/1

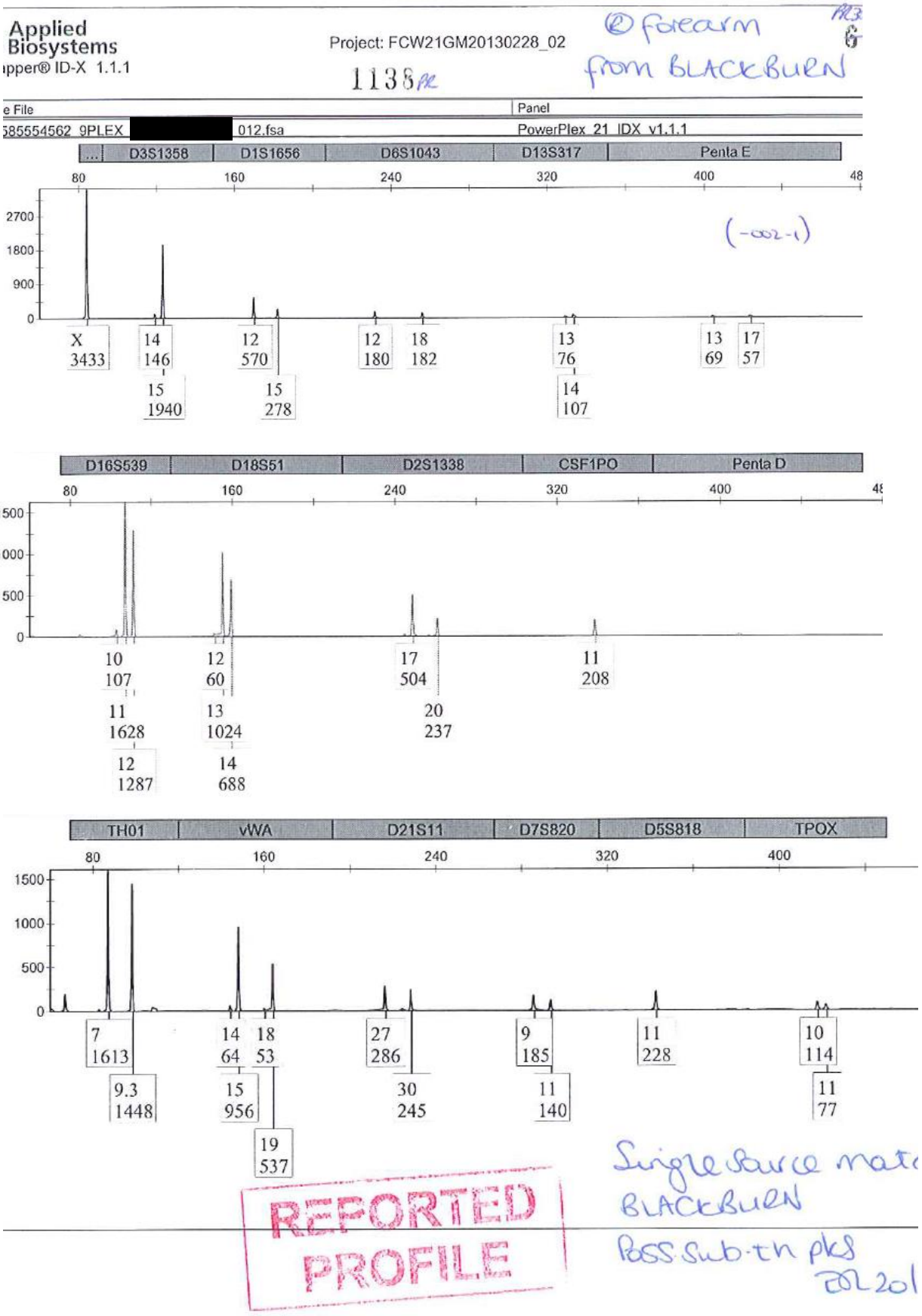
Table 4: Crime scene samples with unexplained DNA degradation and the corresponding date on the electropherograms.

Sample	Date	Sample	Date
Tapelift: R. forearm	28 February 2013	L9a: 'Bloodstained fabric' shirt (BLACKBURN)	3 May 2013
Tapelift: Forehead	28 February 2013	L6a: 'Bloodstained fabric' shirt (BLACKBURN)	3 May 2013
S15: Swab of blood from gutter Boddington Street	19 March 2013	L1a: 'Bloodstained fabric' shirt (BALCKBURN)	3 May 2013
L14a: 'Bloodstained fabric' shirt (BLACKBURN)	3 May 2013	L3a: 'Bloodstained fabric' shirt (BALCKBURN)	3 May 2013

Samples taken directly from a person's skin and from 'bloodstained' fabric should contain many hundreds of good quality cells. Therefore, these samples should have provided good quality DNA profiles unaffected by degradation. The electropherograms indicate the DNA has been degraded, which may be explained if they were affected by the defective Proteinase K.

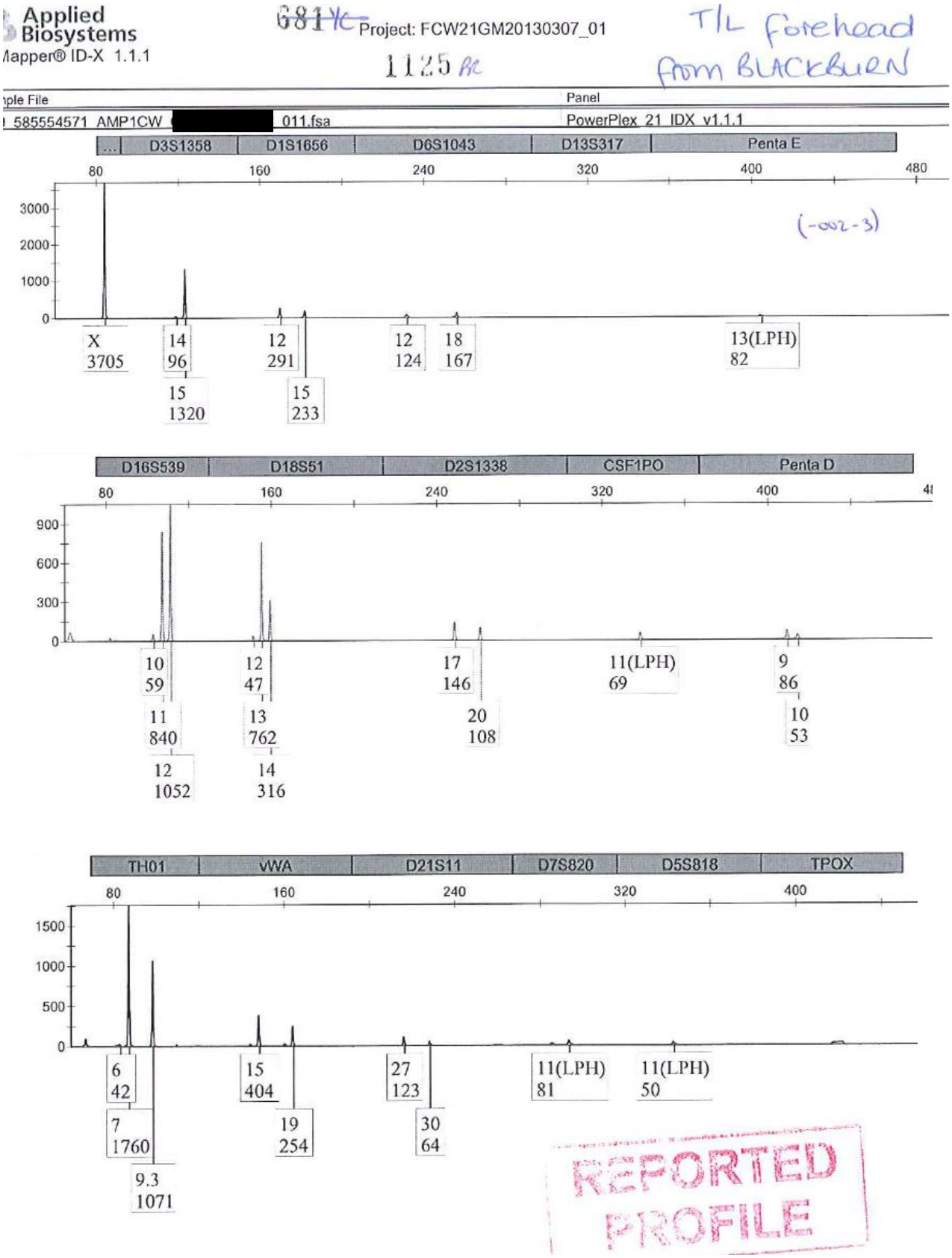
The electropherogram from a sample of BLACKBURN's right forearm was generated on 28 February 2013 (File 2, p1138). Note the significant decrease in peak heights from left to right in the electropherogram (the bottom number for each labelled allele is the peak height) indicating degradation.

Figure 2: Electropherogram of the tapelift from BLACKBURN's right forearm.



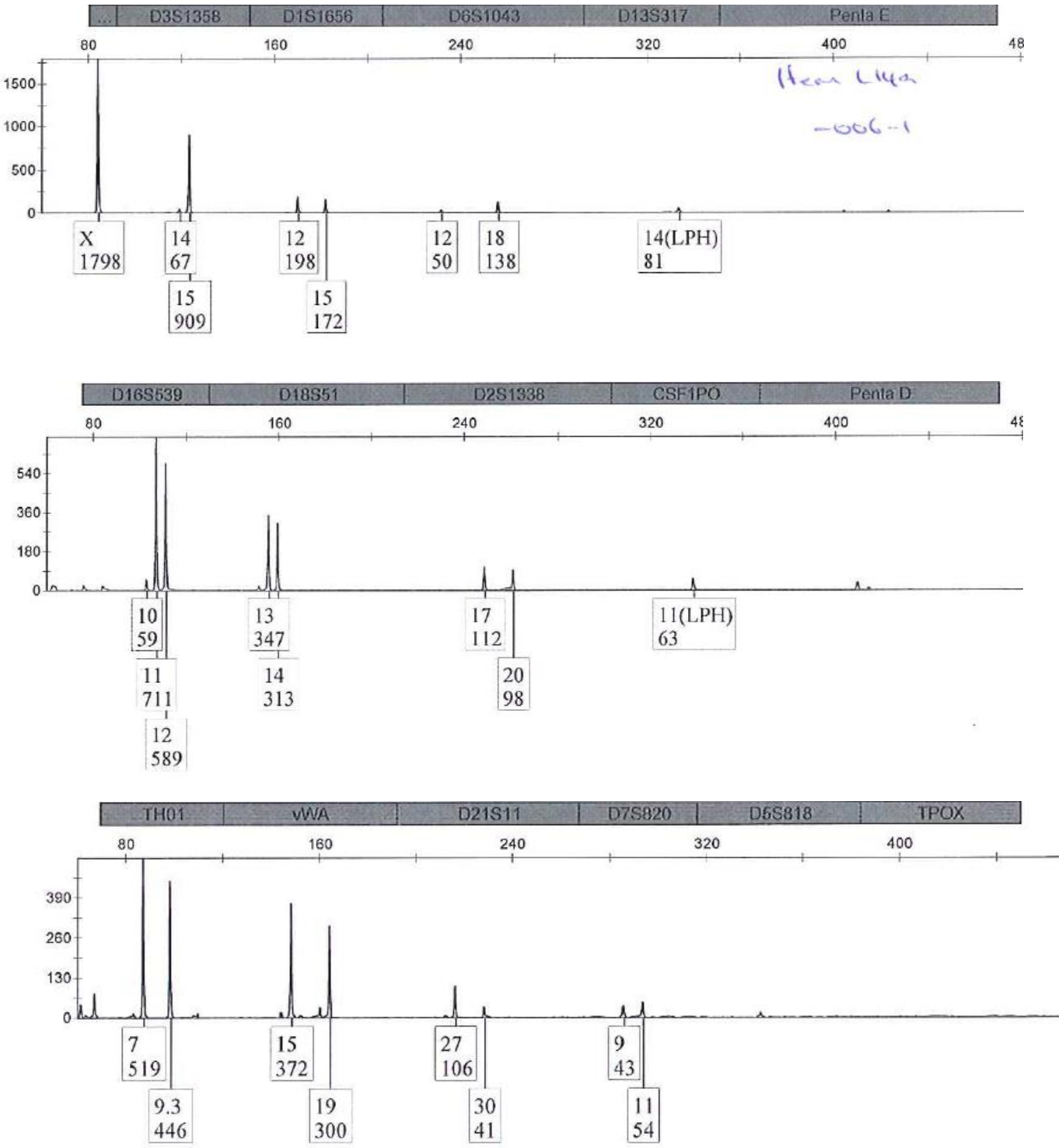
The electropherogram from a sample of BLACKBURN's forehead was generated on 7 March 2013 indicates unexplained DNA degradation (File 2, p1125).

Figure 3: Electropherogram of the tapelift taken from BLACKBURN's forehead.



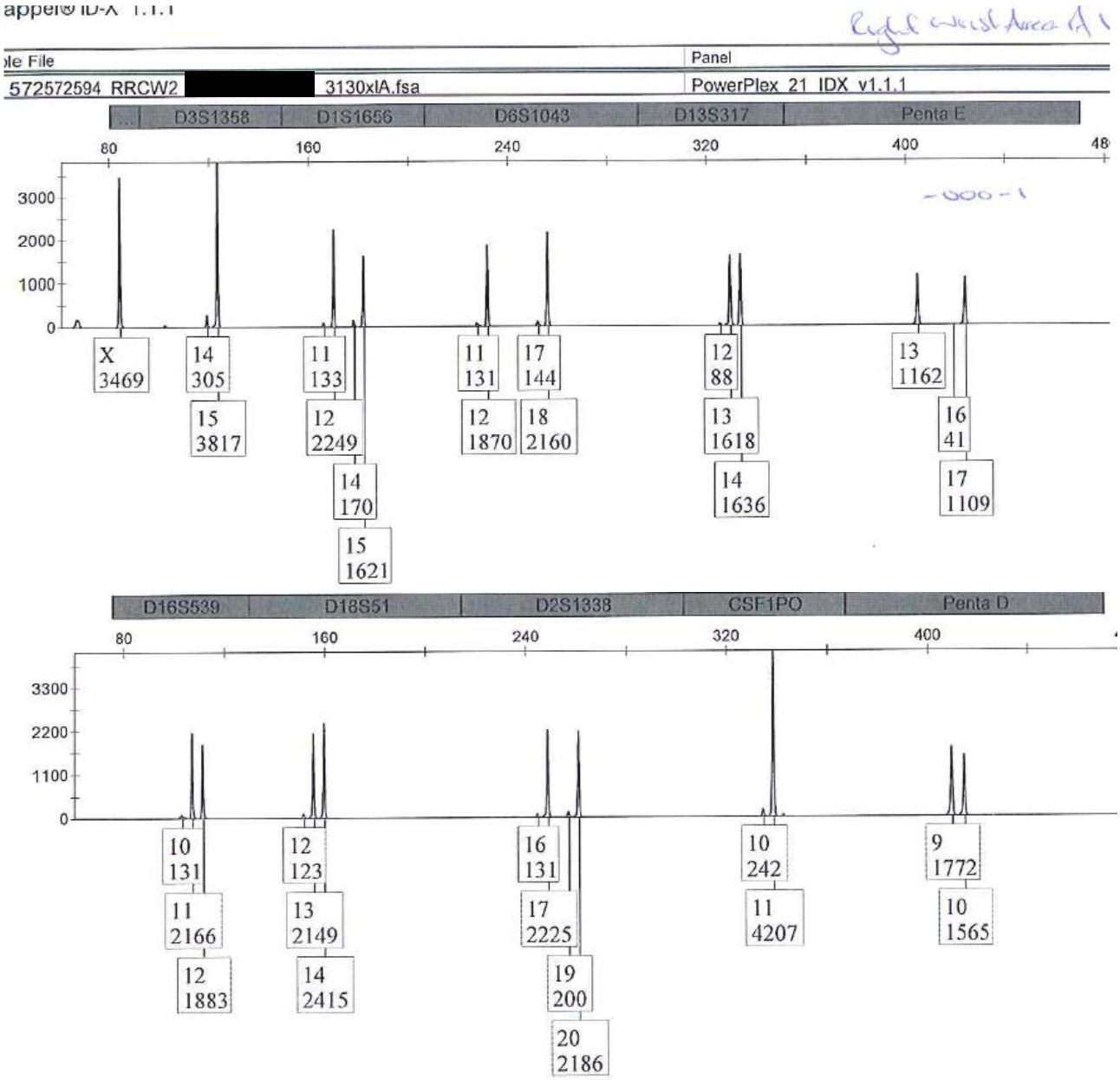
The electropherogram from a sample of bloodstained fabric was generated on 3 May 2013 indicates unexplained DNA degradation (File 5, p0421).

Figure 4: Electropherogram of L14a, 'bloodstained fabric' from BLACKBURN's shirt.



Interestingly, a tapelift taken from BLACKBURN's right wrist provided a good quality DNA profile with no degradation demonstrating results expected from skin samples (File 2, p1495). This sample, however, was profiled on 11 February 2013[13], seventeen days prior to the right forearm and forehead samples. Is it possible the defective Proteinase K was not used on the right wrist sample processed in early February, but was used on the right forearm and forehead samples processed in late February, thereby explaining the DNA degradation? If these samples were not affected by degradation, could the offender's DNA be identified?

Figure 5: Electropherogram of the tapelift taken from BLACKBURN's right wrist.



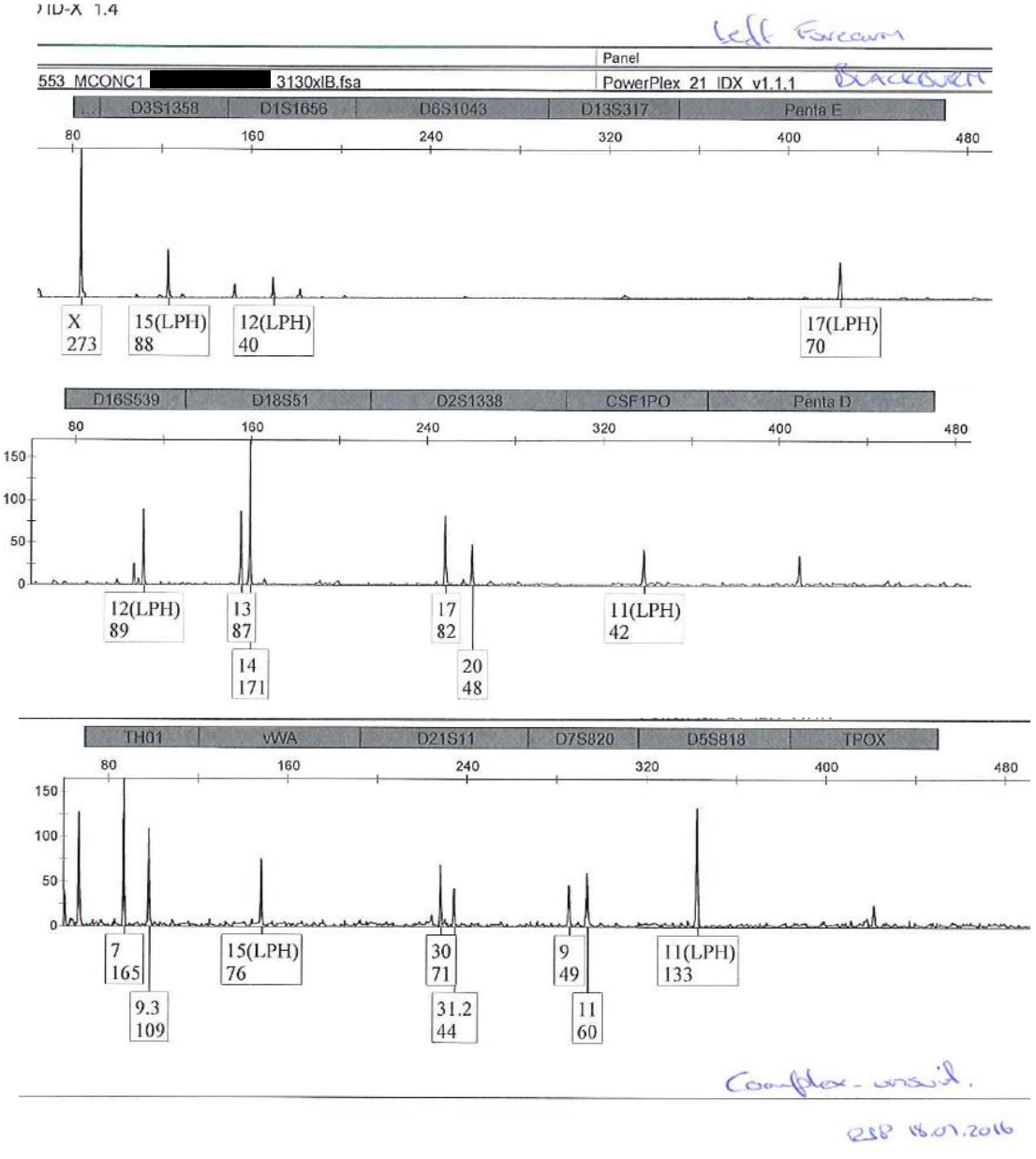
[13] The tapelift from the right wrist was submitted to QHFSS on 11 February 2013, with a request from Inspector Carstenson for a five-day turnaround (File 1, p1748). The tapelifts from the right forearm and forehead were submitted to QHFSS on 22 February 2013 (p2, paragraph 9 of PARRY's DNA statement).

2.4 Unexplained weak crime scene profiles

The third sign of DNA degradation is a 'weak' DNA profile, that is, a profile with smaller electropherogram peaks than expected or a partial DNA profile. OQI#34043 states for affected samples *"the quantification values observed were in the range of 0.01-0.1ng/ul, whereas typically the positive extraction control yields values in the range of 1-3ng"*. Positive control samples of known quality that were affected by the defective Proteinase K therefore had 30 to 100 times less DNA available after extraction. Crime scene samples often contain very small quantities of DNA, therefore a 30 to 100 times reduction in DNA is considered catastrophic.

The tapelift from BLACKBURN's left forearm is an example of an unexplained weak profile (File 2, p1114). Swabs of blood from Boddington Street (S15 and S16), are also unexplained weak profiles given they were taken from visible bloodstains.

Figure 6: Electropherogram of the tapelift from BLACKBURN's left forearm demonstrating an unexplained weak profile.

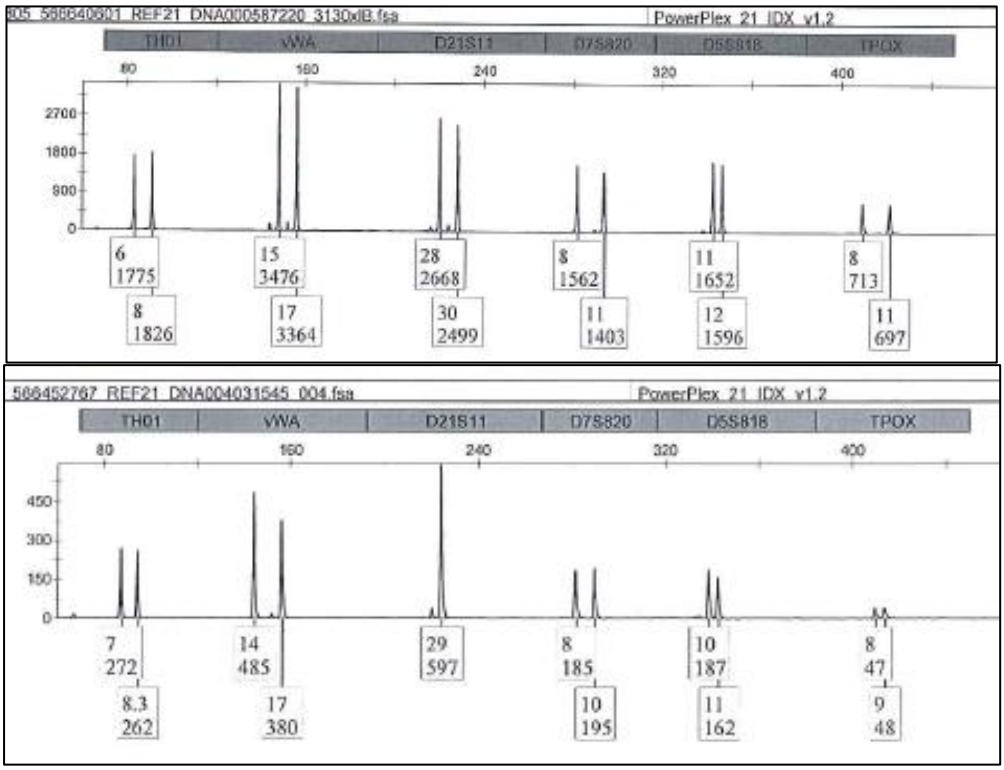


This sample was received on the same date as the tapelifts from the right forearm and forehead on 22 February 2013 that also show unexplained degradation. It was first reported as 'no DNA detected' on 28 February 2013 and was subsequently reworked in 2014 upon request from police (using Microcon concentration), though still provided an unexplained weak partial profile.

2.5 Unexplained weak reference profiles

Over half of the reference DNA profiles received for the BLACKBURN matter did not produce 'strong profiles'[14], and approximately 15% of these failed to generate a profile the first attempt[15]. Reference DNA samples are considered clinical grade samples, meaning they typically contain large amounts of good quality DNA. Approximately 2% to 5% of these samples may fail to produce a profile the first attempt, however, the others should provide strong good quality profiles (tall peaks in the electropherograms).

Figure 7: Electropherograms of reference samples. Top: represents a strong profile. Bottom: represents a weak profile as defined by peak height. The bottom number in each box is the peak height.



Large numbers of failed or weak reference profiles indicates there are underlying issues within QHFSS. Reference samples undergo a different process to crime scene samples that do not require Proteinase K. However, it should be explored whether glassware used in any other

[14] It would be expected peak heights should be 500RFU or over consistently across all loci.
[15] Identified by handwritten notes 'previous runs not reported', or 'EREF'.

process to generate reference DNA profiles may have been affected by the faulty dishwasher. The first reference sample that generated a weak profile was received on 20 February 2013, however, other failed or weak profiles were obtained from samples received throughout 2013 and as late as July 2014.

2.6 Summary

The impact the faulty dishwasher and defective Proteinase K would have on crime scene evidence is catastrophic. The date range of the issue and the number of samples affected should have been reported in OQI#34043. Critical crime scene evidence totalling 108 samples were received prior to the issue being detected and the defective Proteinase K being completely removed from use. Crime scene evidence from the BLACKBURN matter have unexpectedly failed, have unexplained degradation or unexplained weak profiles.

It seems unlikely that only *one* measuring cylinder and *one* lot of Proteinase K for *one* laboratory process was affected by the faulty dishwasher. For example, DTT (a chemical used for DNA extraction) is prepared in a similar manner to Proteinase K. The full extent of the malfunctioning dishwasher on laboratory processes needs to be thoroughly investigated by an independent quality expert to determine the time range of the issues and exactly how many and which crime scene samples were affected.

The four reference samples nominated as being affected by this OQI were received on **28 February 2013**, and likely underwent the secondary extraction process using Proteinase K in early March. It is therefore certain that crime scene samples processed in **early March** were at risk of being affected. BLACKBURN's crime scene samples received on 11 and 13 February 2013 provided good quality DNA profiles. These were reported to QPRIME prior to and on 20 February, therefore would have been processed between 11 and 18 February[16]. Issues with crime scene samples first appear in items received on **22 February** with profiling first completed by 28 February[17]. It is therefore possible the faulty dishwasher started affecting crime scene samples between **12 February and 24 February**,[18] however, was not detected until 22 March 2013 (approximately 4 to 6 weeks later). The failure to remove the defective Proteinase K from use may have resulted in a **two to three month** period in total where crime scene evidence may have been affected.

[16] With DNA extractions likely conducted between 11 and 14 February 2013

[17] Tapelift right forearm, tapelift forehead, left forearm, S14, S15, and S16 swabs of blood upper gutter verge Boddington Street (received 22 February 2013) all provide poor quality or no DNA profiles.

[18] 24 February is the latest the samples received on 22 February 2013 would have undergone DNA extraction to enable them to be completed by 28 February 2013.

A total of **103 crime scene samples** were received between 12 February and prior to May (when the defective Proteinase K was completely removed from use). These include evidence from [redacted]' car including the 12 'blood' samples, samples from BLACKBURN's body, clothing, shoes, phone, the 'blood-soaked fabric' samples from the white T-shirt found near the crime scene, the blood samples from the gutter, and one of the two knives submitted.

OQI#34043 states that: "*All affected samples, where substrate remained following the initial extraction, were re-extracted. Whilst the initial DNA extraction performed sub-optimally (low DNA yields) any DNA profiles obtained from either the original extraction or the re-extraction (or pooled samples) are reportable"*[19]. Reporting DNA evidence affected by OQI#34043 is misleading to the police and courts given the ground truth of crime scene samples is unknown. Destruction of crime scene DNA prevents the entire evidence being discovered, including information that may link to the victim or offender. A clear and thorough explanation of the risks associated with accepting this DNA evidence would be required. For example, the tapelift from BLACKBURN's right forearm was a weak partial profile (which was not apparent in the DNA statement[20]), though if affected by the defective Proteinase K, may have generated a DNA mixture of two people (victim and offender). The weak partial profile therefore would have been reported and incorrectly accepted as the entirety of evidence available.

For crime scene samples with no remaining material after DNA extraction the evidence cannot be retested. This is a catastrophic outcome. It is unclear how many samples this relates to. In some circumstances QHFSS may have retained the substrate after extraction in a 'spin basket'[21]. It is uncertain whether these samples still exist, though if included in the original extraction with the defective Proteinase K, they are at risk of containing no DNA, or degraded DNA.

The possible ramifications of the faulty dishwasher affecting DNA evidence, but QHFSS not disclosing this include:

1. failed DNA samples that may have contained DNA from offenders or victims;
2. failed DNA samples that may have contained exculpatory evidence;
3. not detecting DNA from offenders or victims if they were the second or third contributor in a DNA mixture (i.e., only detecting the DNA of one person);
4. the loss of opportunity to rework crime scene samples to produce more informative results;
5. loss of opportunity for police to collect more crime scene samples for analysis; and

[19] File 1, p1919, paragraph 8.
[20] The tapelift from the right forearm was reported as a single contributor profile matching to BLACKBURN.
[21] QHFSS Procedure for Case Management (v16), p 13, section 6.2.7.

6. presentation of failed DNA test results in court that should not have been submitted as evidence, favouring an incorrect scenario and potentially affecting the outcome of trials.

Other evidence exists of systemic failures within QHFSS processes causing catastrophic outcomes for crime scene evidence. QHFSS reveal in the Procedure for Case Management (v16) document that "*semen samples have also been observed to return an NSD profile after initial extraction with no indication of inhibition*[22]". This procedure was last updated on 12 November 2012, though it is unknown whether this observation was present in previous versions of the document. The frequency of failed semen samples must have occurred regularly to warrant inclusion of this warning in the procedure. Semen is a rich source of DNA, and the comment explicitly states no inhibition was detected (in the quantitation stage) meaning a catastrophic failure is occurring.

There is a trend within QHFSS of samples rich in DNA unexplainably failing to produce DNA profiles. The 2021 Krosch paper[23] also contains evidence of unexplained systemic failure of sexual assault samples analysed in 2018 and 2019 (52% failure from penis samples). This is evidence QHFSS knew vital evidence was failing to provide DNA results, and they were incorrectly reporting 'no DNA detected' to the police and courts. It is therefore possible other systemic and catastrophic failures in the laboratory's processes were occurring at the same time BLACKBURN's crime scene evidence was processed which prevented vital evidence from being found.

Recommendation 2: An independent investigation is required to examine the failure of semen samples, including determining when the issues first started, the cause of the failure, and to identify all samples affected.

Recommendation 3: The success rate of QHFSS DNA analysis requires close examination to uncover the real scope of the flawed testing and identify samples that require re-testing. The following analysis of QHFSS data needs to be undertaken by an independent expert:

a) determine the success rate of each sample type (item and collection method as per Krosch paper);

b) determine the success rate of samples submitted for DNA analysis presumptively positive for blood;

[22] P10, paragraph 2. Note: NSD = no size data, that is, a failed DNA profile after full testing.
[23] M. Krosch. Variation in forensic DNA profiling success among sampled items and collection methods: a Queensland perspective. *Australian Journal of Forensic Sciences*, (2021) 53:6, 612-625, DOI: 10.1080/00450618.2020.1759687.

c) determine the success rate of samples submitted for DNA analysis presumptively positive for semen;

d) determine the success rate of samples submitted for DNA analysis confirmed positive for semen (confirmed through microscopy);

e) determine the success rate of samples taken from obvious stains of biological fluid (presumptively positive);

f) how many samples were reported to QPS as 'INCORRECT RESULT' by QHFSS; and

g) conduct a trend analysis on profile success rates (by item and collection method) from 2010 onwards.

These are critical performance measures that should be reported for each year. The Forensic Register can easily provide information on submitted samples that were positive for biological fluids (b-d) yet yielded no profile. Similarly, the Forensic Register will have recorded the appearance of the stain (e), and how many 'Incorrect Results' have been reported (f).

3.0 QHFSS quantitation thresholds

There has been much discussion about the quantitation thresholds used by QHFSS which determines whether crime scene evidence is fully tested. The research paper by Dr Krosch published in 2021[24] stated the QHFSS threshold was 0.0088 ng/ul, which is twice as high as the threshold used by the New South Wales (NSW) Forensic and Analytical Scientific Services (FASS) laboratory (0.004 ng/ul)[25]. Concentration will be used as the unit to compare thresholds between QHFSS and the NSW FASS laboratory rather than the number of cells. Despite QHFSS using full and half volume PowerPlex reactions (which require less DNA than full reactions), it is the concentration of the sample that determines whether it is fully tested. A laboratory's threshold setting is based on information from internal validation studies and is also a business decision. Therefore, at a certain DNA concentration management have determined that a crime scene sample is highly unlikely to provide a reportable DNA profile. The decision is made to stop testing and save resources which can be used on other samples more likely to generate a result. This is an acceptable practice used throughout the world.

QHFSS procedures used at the time of testing BLACKBURN's crime scene samples state their quantitation threshold was actually higher than 0.0088 ng/ul. If a crime scene sample did not reach a concentration of 0.01 ng/ul, it would not be tested further and reported as 'DNA insufficient for further processing'[26]. This concentration is 2.5 times greater than the NSW threshold and requires further investigation. The concentration QHFSS reports as 'no DNA detected' is <0.0021 ng/ul.

The QHFSS PowerPlex 21 internal validation studies (2012 and 2013) report concentrations less than 0.0088 ng/ul "*may result in increased stochastic effects*[27]". Stochastic effects include imbalances in allele peak height within a locus, and peak drop-out (a missing peak where expected) and can confuse interpretation of profiles leading to error. The validation study shows QHFSS were obtaining DNA profiles at well below the 0.01 ng/ul threshold. At three times below this on average 41/42 alleles were obtained, however, the stochastic effects made interpretation of these samples unreliable. Therefore, there is ample crime scene DNA below the QHFSS quantitation threshold, enough to obtain complete DNA profiles, however, the stochastic effects they were observing led to the high threshold.

Interestingly, QHFSS chose to use a 'half-volume' PCR reaction (12.5ul) as the default for all crime scene samples. Although the PowerPlex 21 developmental validation states "*the*

[24] M. Krosch. Variation in forensic NDA profiling success among sampled items and collection methods: A Queensland perspective. *Australian Journal of Forensic Sciences*, (2021) 53:6, 612-625, DOI:10.1080/00450618.2020.1759687. See p613, paragraph 4.
[25] E. Prasad, et al., Trace recovery rates from firearms and ammunition as revealed by casework data. *Australian Journal of Forensic Sciences*, (2021). DOI: 10.1080/00450618.2021.1939783. See p5, paragraph 1.
[26] QHFSS Procedure for Case Management (v16), p7, paragraph 10.
[27] QHFSS PowerPlex 21 Amplification of extracted DNA validation 2012, p63, paragraph 10.

recommended reaction volume of the PowerPlex 21 system is 25ul[28]", which is the 'full-volume'. The chemicals used in profiling kits are the most expensive part of the profiling process which would have been a major factor in this decision. QHFSS also chose to use 30 cycles for PCR conditions (the maximum), which is likely to have caused increased stochastic effects. The PowerPlex 21 Technical Manual recommends users optimise their protocol, including cycle number, however QHFSS did not try different cycle numbers as part of their validation study to reduce stochastic effects. It should be noted that the NSW FASS laboratory use the recommended 25ul reaction volume and 29 cycles[29]. The PowerPlex 21 Technical Manual (troubleshooting section) also recommends using fewer cycles if increased stochastic effects exist, and suggests the issues could be caused if:

> "*The reaction volume was too low. This system is optimized for a final reaction volume of 25ul. Decreasing the reaction volume can result in suboptimal performance*.[30]"

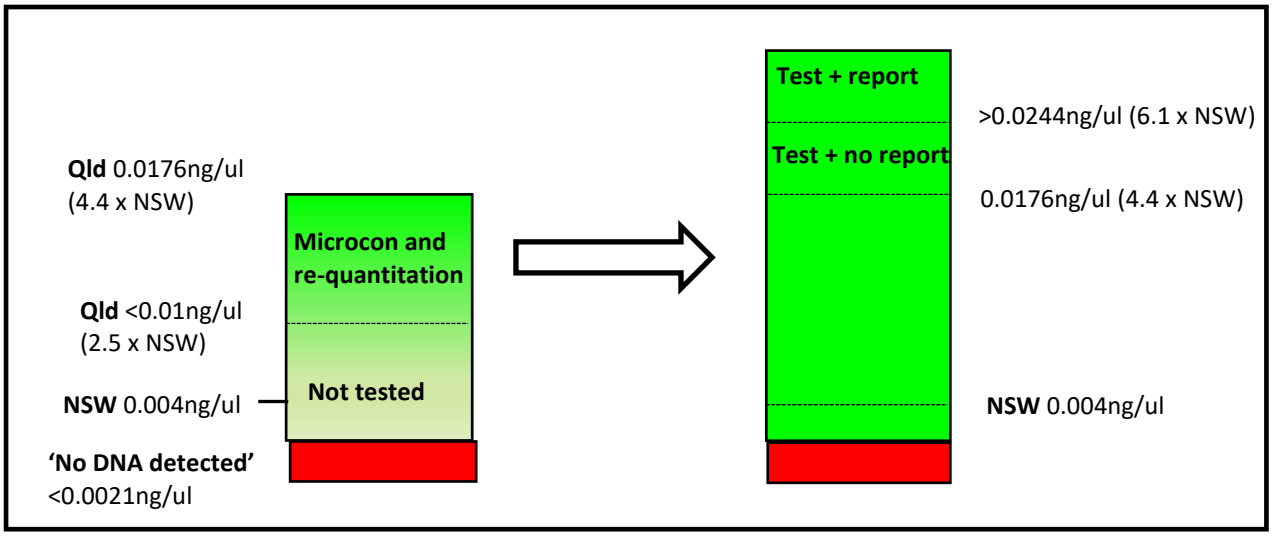
Although QHFSS set their quantitation threshold at 0.01 ng/ul, this did not result in all crime scene samples above this concentration being fully tested. The QHFSS Procedure for Case Management (v16) details additional thresholds in a convoluted process that was introduced on 11 December 2012 to coincide with the implementation of PowerPlex 21. If a crime scene sample was between 0.01 ng/ul and 0.0176 ng/ul, it would undergo Microcon concentration, then be re-quantitated. The upper range of 0.0176 ng/ul is 4.4 times the concentration required by NSW to proceed to amplification with PowerPlex 21.

After re-quantitation, if the crime scene sample was:

- between 0.0176 ng/ul and 0.0244 ng/ul the sample would be amplified at half volume (12.5ul), though the QHFSS procedure states "*currently these results are not reported back in an EXH*[31]".
- above 0.0244 ng/ul the sample would be amplified at half volume and reported. This is 6.1 times the concentration required by NSW.
- between 0.01 ng/ul and 0.0176 ng/ul the case manager would determine if further testing was required.

[28] Ensenberger, M. et al. Developmental validation of the PowerPlex 21 system. *Forensic Science International:Genetic*s 9 (2014) 169-178.
[29] E. Prasad, et al., Trace recovery rates from firearms and ammunition as revealed by casework data. *Australian Journal of Forensic Sciences*, (2021). DOI: 10.1080/00450618.2021.1939783. See p5, paragraph 1 and 2.
[30] Technical Manual: PowerPlex 21 System for use on the Applied Biosystems Genetic Analysers, p54.
[31] QHFSS Procedure for Case Management (v16), p7, paragraph 12. It is uncertain if these results were routinely released to police.

Figure 8: QHFSS quantitation thresholds to determine how crime scene samples were tested.



It appears QHFSS knew the PowerPlex 21 method used in the internal validation was not producing optimal results due to stochastic effects, but rather than troubleshoot the issues by reducing cycle number and using the recommended PCR volume, they instead decided to raise the quantitation threshold and implement a convoluted re-work process. This would have led to many thousands of crime scene samples failing to be fully tested because they fell below the excessively high QHFSS threshold.

The QHFSS procedure required a DNA concentration between 2.5 and 6.1 times greater than another jurisdiction's threshold at the time crime scene samples from BLACKBURN's case were being processed. Although QHFSS reported 43 samples as 'no DNA detected' (below 0.0021 ng/ul), surprisingly no samples were reported as 'DNA insufficient for further processing' (between 0.0021 ng/ul to 0.01 ng/ul).

Despite fully testing the sample from BLACKBURN's left forearm (originally reported as 'no DNA detected') and obtaining a profile (albeit a partial profile), QHFSS did not fully test any of the other 43 samples, including the critical samples from [redacted]' car and the 'blood soaked' samples from the white T-shirt.

Recommendation 4: An independent review is required of QHFSS quantitation thresholds in relation to the PowerPlex 21 validation data and any other internal research performed. Ideally the QHFSS PowerPlex 21 method requires optimisation to reduce stochastic effects, and a new quantitation threshold set. Evaluation is needed of critical crime scene samples that were previously not fully tested to determine if they should be further analysed.

4.0 Incorrect setting used for capillary electrophoresis (OQI#34817)

4.1 Background

The QHFSS laboratory has two Genetic Analyzers that performs capillary electrophoresis on crime scene and reference samples. These are internally labelled as 'A' and 'B'. On 8 July 2013 Genetic Analyzer B was identified as having an incorrect setting. Specifically, the injection time was set at 3 seconds instead of 5 seconds. The injection time affects how much DNA from a sample is automatically removed by a needle-like system from a small tube and analysed by the instrument. Therefore, a greater volume of sample potentially containing DNA is injected into the instrument using a 5 second injection setting, than with 3 seconds.[32]

If more DNA is available for analysis, it increases the chance of detecting a DNA profile from an individual, particularly when their DNA is present in only small quantities (this includes minor contributors in DNA mixtures). Smaller peak heights are likely to be observed in electropherograms from samples affected by this error, or peaks may fall below the reporting threshold preventing some or all DNA information being obtained. This is confirmed in OQI#34817 that reports: "*There were however, instances where the 5 second injection time run showed additional peaks that were above the limit of reporting (LOD[33] = 50 RFU) that were not above the LOD in the 3 second injection time run*". The OQI also confirms that peak heights were on average 1.75 times higher using the correct injection time (between 1.3 to 2.48 times higher)[34].

The incorrect setting had been used since the introduction of PowerPlex 21 in December 2012 (several months prior) and was also used in the validation of PowerPlex 21 and STRMix software. In 2013 QHFSS reported analysing over 22,000 items[35], therefore many thousands of crime scene samples were likely affected over several months. Fortunately, there is enough sample available for re-analysis on the Genetic Analyzer[36]. If each affected sample was identified and re-analysed using the correct setting, then no DNA information would have been lost. The OQI states '*all affected samples and batches were identified, and appropriate AUSLAB batch audit entries and specimen notes were made*'[37].

[32] The PowerPlex 21 User Manual states an injection time of between 3 seconds and 22 seconds may be used.
[33] LOD = limit of detection. Note that QHFSS limit of reporting (LOR) is 40 RFU, not 50 RFU as stated in the OQI.
[34] File 1, p1916, paragraphs 3 and 4.
[35] DOH-DL 15/16-041 https://www.health.qld.gov.au/system-governance/contact-us/access-info/disclosure-logs/2015-16
[36] The lab could either use the remaining PCR product or re-analyse the original plate run on Genetic Analyzer B if it was retained. In both circumstances there is plenty of sample available.
[37] File 1, p1916, paragraph 8.

4.2 BLACKBURN evidence QHFSS listed against OQI#34817

There are four specimen notes and two handwritten notes indicating affected crime scene samples in the BLACKBURN DNA case file. They were first reported in QPRIME in February 2013, though have electropherograms with analysis dates in July 2013. This indicates the samples were subsequently re-analysed with the correct settings. All samples were originally reported as 'single contributor', which did not change after re-analysis, therefore no additional information was gained.

Table 5: BLACKBURN crime scene samples listed against OQI#34817.

Sample	Date first reported	Reworked electropherogram
Cig butts from Girl Guide's hut	14 Feb 2013	19 July 2013
BLACKBURN's right wrist	14 Feb 2013	19 July 2013
BLACKBURN's right fingernail scrapings	14 Feb 2013	19 July 2013
BLACKBURN's left fingernail scrapings	14 Feb 2013	19 July 2013
BLACKBURN's shirt (L16)	19 Feb 2013	17 July 2013
BLACKBURN's right palmate	19 Feb 2013	17 July 2013

Interestingly, there are 21 crime scene samples received around the same time as the above items, which were also reported in QPRIME in February 2013, though had additional reworks and electropherograms with analysis dates in July 2013. Three of these were originally reported as single contributor profiles in February, then later changed to two-person mixtures. The other 18 samples were reported in February as single contributor, and the July electropherogram also indicated a single contributor. Review of the BLACKBURN audit trail reveals 35 samples were re-analysed due to being affected by OQI#34817[38].

This indicates OQI entries are not always made against each affected sample in a case file, which is an incorrect practice and misleading. Samples involved in all OQIs need to be readily identified in the DNA case file to provide DNA experts acting on behalf of defence the opportunity to properly evaluate DNA results, and for the limitations and reliability of any affected results to be fully articulated in DNA statements and court testimony. Did QHFSS also fail to record crime scene samples affected by the faulty dishwasher OQI in the case file?

[38] Excel spreadsheet 'FSS.001.002.6261'. See column K.

4.3 Impact on PowerPlex 21 and STRMix internal validations

The PowerPlex 21 internal validation informed user-defined settings for STRMix. The issue of main concern in relation to OQI#34817 is incorrect settings were used on samples involved in the PowerPlex 21 and STRMix internal validations. The validation studies used both A and B Genetic Analyzers. Internal validation is required before any new instrument or method (including software) is used on crime scene and reference samples to determine whether it is fit for purpose, working accurately and reliably prior to introduction, to understand the limits of the technology, to determine analysis and interpretation thresholds, and define settings that should be used for the instrument, software, or method. It is unclear from QHFSS documents whether use of the incorrect setting to generate internal validation data has had a detrimental effect on the implementation of PowerPlex 21 and STRMix. If so, results generated from crime scene evidence using PowerPlex 21 and STRMix since their implementation in December 2012 (regardless of which Genetic Analyzer was used) may not be valid and risk being incorrect or should be reported with limitations.

All Australian jurisdictions were expected to implement a new DNA profiling kit by the end of 2012, an expectation directed by the Australian and New Zealand Policing Advisory Agency (ANZPAA), National Institute of Forensic Science (NIFS).[39] PowerPlex 21 and STRMix were two major changes to the QHFSS DNA analysis process in 2012, and STRMix is a complex program requiring thorough and accurate internal validation. It is likely both validation studies took many months to complete.

'Drop-in' and artifacts are peaks in the electropherogram that do not represent DNA from the crime scene. Internal validation will define two critical profile interpretation parameters: the limit of reporting (LOR) and the drop-in threshold. The reliable internal validation of the PowerPlex 21 kit is therefore key to accurate profile interpretation and mixture analysis using STRMix. STRMix developers emphasise '*that it is the multiplex-interpretation method couplet that requires validation*'[40].

OQI#34187 states an amended PowerPlex 21 validation report was released (in December 2013) using new data and QHFSS management determined that "*some additional data was obtained (but not utilised for the validation report), the PP21 validation report will be re-issued with some data split into 3 second and 5 second injection time data, and some additional data obtained from 3 second injection times included*[41]". They note that concordance data, LOR, stutter and drop-out rates were not significantly affected. Importantly, there is no mention of drop-in rates or thresholds, peak height ratios, mixture, and dilution analysis.

[39] QHFSS PowerPlex 21 Amplification of extracted DNA validation. 2012, p7.
[40] Duncan Taylor et al. Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Science International: Genetics, (2016) 20, 6-19 (p6)
[41] File 1, p1915, paragraph 3.

It is not clear from the OQI what the change in validation data was or how it might impact profile interpretation and STRMix outputs. It is not clear whether the updated validation included STRMix analysis. It is not clear whether any changes were made to standard operational procedures involving profile interpretation, or if any changes were made to STRMix user-defined settings. If changes were made to either procedures or STRMix settings, then samples analysed and reported prior to this would require re-analysis and where needed, results updated, and the police and courts informed. This would be a large undertaking given issue were first detected several months after the incorrect setting was first used.

4.3.1 Incorrect limit of reporting and drop-in thresholds

An examination of the QHFSS PowerPlex 21 internal validation studies from 2012 and 2013 was performed to determine if any changes in analytical or interpretation thresholds were required after removal of the 3 second injection data. Surprisingly, results for determining the LOR threshold, the most critical threshold in the DNA profiling process, are identical between studies.

Table 6: Baseline results for amplification at 12.5ul identical between the QHFSS 2012 and 2013 PowerPlex 21 internal validation studies[42].

		3130xl A 12.5µL	3130xl B 12.5µL	Overall 3130xl A & B 12.5µL
Fluorescin (Blue)	μ_{PK}	3.10	2.19	2.64
	σ_{PK}	3.66	2.72	2.99
	LOD	14.07	10.36	11.59
	LOR	39.67	29.42	32.49
JOE (Green)	μ_{PK}	4.46	2.69	3.62
	σ_{PK}	4.41	2.86	3.86
	LOD	17.70	11.26	15.22
	LOR	48.60	31.28	42.27
TMR (Yellow)	μ_{PK}	6.06	3.58	4.83
	σ_{PK}	4.15	2.43	3.63
	LOD	18.50	10.88	15.70
	LOR	**47.52**	**27.92**	**41.08**
CXR (Red)	μ_{PK}	2.87	2.10	2.49
	σ_{PK}	2.32	1.28	1.93
	LOD	9.84	5.94	8.27
	LOR	26.11	14.90	21.75
CC5 (Orange)	μ_{PK}	2.38	1.66	2.02
	σ_{PK}	2.31	1.87	2.14
	LOD	9.33	7.26	8.84
	LOR	25.53	20.33	23.40
Overall	μ_{PK}	3.94	2.54	3.32
	σ_{PK}	3.87	2.46	3.30
	LOD	15.56	9.91	13.21
	LOR	**42.68**	**27.10**	**36.28**

μ_{PK} = Average peak height, σ_{PK} = Standard Deviation, LOD = limit of detection, LOR = Limit of Reporting

[42] See Table 12, p29 of 2012 PowerPlex internal validation study, and Table 12 p29 of the 2013 PowerPlex internal validation study.

The table includes results from Genetic Analyzer B using the incorrect 3 second injection time. As expected, there is a clear difference between results from instruments A with the 5 second setting (see Table 6 column 1), and instrument B with the 3 second setting (see column 2), with B providing significantly lower results based on peak heights. This difference was noted in the 2012 study prior to discovery of the incorrect instrument setting "*It was noted on 3130xl A the baseline was raised more than expected compared to the other baseline runs on the same instrument and baseline runs on 3130xl B. This could be due to a prolonged period between spectral calibrations, ageing reagents and arrays and was taken into consideration when setting thresholds*[43]". Interestingly, the exact same wording is used in the 2013 validation study to describe the difference, despite QHFSS being aware the difference was actually due to the incorrect setting used for instrument B[44].

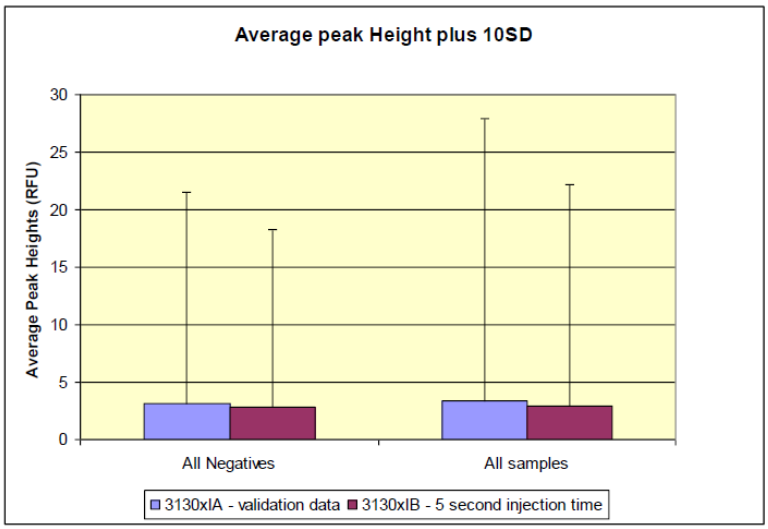
In fact, the 2013 validation does not highlight the original error at all or include the OQI which is deceptive. Without knowledge of OQI#34817, the reader cannot understand the importance of these results. This information would be vital to defence experts who typically obtain copies of internal validation studies when they review a DNA case file to check the suitability of key thresholds used in profile interpretation.

Critically, the results from both instruments were used to set the 2012 LOR by using the highest 'overall' LOR between both instruments (see column 3).[45] The highest 'overall' LOR is 42.27 RFU (see highlighted box in Table 6) and QHFSS rounded this to down to 40 RFU and used this as their LOR threshold since December 2012. However, given the results from instrument B should be excluded, the highest LOR is 48.60 RFU for instrument A. Rounded up, the correct LOR would be 50 RFU, 20% higher than the threshold used on crime scene samples.

Re-analysis using the 5 second injection time of samples initially examined on Genetic Analyzer B shows they are now similar to instrument A (see Figure 9)[46]. Incredibly, QHFSS decided not to change the LOR threshold. This demonstrates QHFSS knew the critical LOR threshold was wrong but chose not to change it. They failed to re-analyse previous crime scene samples with the correct LOR threshold, and persisted with the incorrect threshold. Therefore, all crime scene profiles from the BLACKBURN matter were analysed by QHFSS with the incorrect LOR threshold.

[43] QHFSS PowerPlex 21 amplification of extracted DNA validation (2012), p28, paragraph 1.
[44] QHFSS PowerPlex 21 amplification of extracted DNA validation, v2.0 (2013), p28, paragraph 1.
[45] This appears to be the average LOR across all results from both instruments for the 12.5ul amplification results, as they are higher than the LORs for the 25ul amplification results.
[46] Aguilera, M. et al. Summary report of baseline determination on 3130xl B after change in injection time. 2013, p3

Figure 9: Re-analysis of samples initially used in the 2012 validation with the incorrect 3 second injection time, now analysed with the 5 second injection time (3130xl B). As expected, the peak heights are now higher, and similar to samples analysed on 3130xl A with the 5 second injection time in the 2012 validation[47].



Failure to correct the initial error in the 2012 PowerPlex 21 internal validation is significant, as an incorrectly lowered LOR would result in labelled peaks that are unreliable (may not represent true alleles), which could lead to:

a) drop-in peaks interpreted as true alleles;
b) single contributor profiles reported as two-person mixtures (therefore indicating the presence of a non-existent person at a crime scene);
c) mixtures with an incorrect number of donors nominated (i.e., 3 instead of 2);
d) offenders and victims incorrectly nominated as contributors or non-contributors to crime scene evidence; and
e) a different likelihood ratio reported for DNA evidence.

Failure to provide the new LOR data in the 2013 PowerPlex 21 internal validation, and instead providing the old incorrect data and not highlighting the injection time error is misleading. The new LOR results should have been included in the 2013 PowerPlex 21 internal validation, rather than in a separate report that defence experts may not be aware of. This has prevented defence experts from properly evaluating QHFSS results, and detecting the potential errors listed above.

Another possible impact of the flawed validation data is its use in the STRMix 'Model Maker' function to calculate peak height and locus-specific amplification variance . Further investigation is needed to understand the impact of this.

[47] Aguilera, M. et al. Summary report of baseline determination on 3130xl B after change in injection time. 2013, p3.

4.3.2 Impact of incorrect thresholds on BLACKBURN DNA mixture interpretation

The possible errors arising from the incorrect LOR and drop-in settings could have profound consequences on the criminal justice system and requires further examination. These errors could have been significantly mitigated by the 'drop-in'[48] function used by STRMix to identify labelled peaks that are unlikely to be true alleles. These peaks are given a lower weighting by the software compared to true alleles when LRs are calculated, reducing the chance of offenders, suspects and victims being incorrectly interpreted as contributors or non-contributors to crime scene evidence.

There are four parameters used by the software to identify drop-in, including a user-defined threshold where a cap on the maximum allowed drop-in peak height is set. For example, the Office of the New York Medical Examiner's DNA laboratory's drop-in cap is 100 RFU (their LOR is 50 RFU).[49]

QHFSS have set their drop-in threshold to 40 RFU to be deliberately the same as their LOR threshold. "*Since our LOR was determined to be 40RFU, it seemed reasonable to set the drop-in level to 40RFU*"[50]. This is confirmed in the STRMix internal validation report: "*The maximum observed drop-in at a locus was 21 rfu, therefore we propose a value of 40 rfu (equal to the detection threshold) for the drop-in setting*[51]" As previously discussed, this threshold is incorrect, meaning drop-in peaks above 40 RFU may be given the same probabilistic weighting as true alleles. A drop-in peak 44 RFU high is present in the sample from BLACKBURN's left forearm (at D21) demonstrating a higher threshold is required. A drop-in threshold of up to 100 RFU should be considered.

The incorrect LOR threshold and incorrect drop-in threshold separately would be significant issues affecting accurate reporting of DNA profiles, however, the combined error has potentially dire consequences on accurately reporting DNA mixture evidence to the police and courts.

There are 19 two-person DNA mixtures from BLACKBURN's crime scene evidence that were analysed with STRMix. Seventeen have every labelled peak in a stutter position in the component interpretation table as a potential genotype for a second contributor[52]. Each of these

[48] Drop-in is the presence of 1 to 3 unexplained labelled peaks in an electropherogram, that when retested, are not reproducible. P. Gill et al, DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics*. (2012) 6(6): 679-688. O. Hansson & P. Gill. Characterisation of artifacts and drop-in events using STR-validator and single-cell analysis. *Forensic Science International: Genetics*. (2017) 57-65.

[49] Estimation of STRMix parameters for OCME New York Laboratory. (2016).

[50] QHFSS PowerPlex 21 amplification of extracted DNA validation (2012), p39, paragraph 6; QHFSS PowerPlex 21 amplification of extracted DNA validation (2013), p37, paragraph 6.

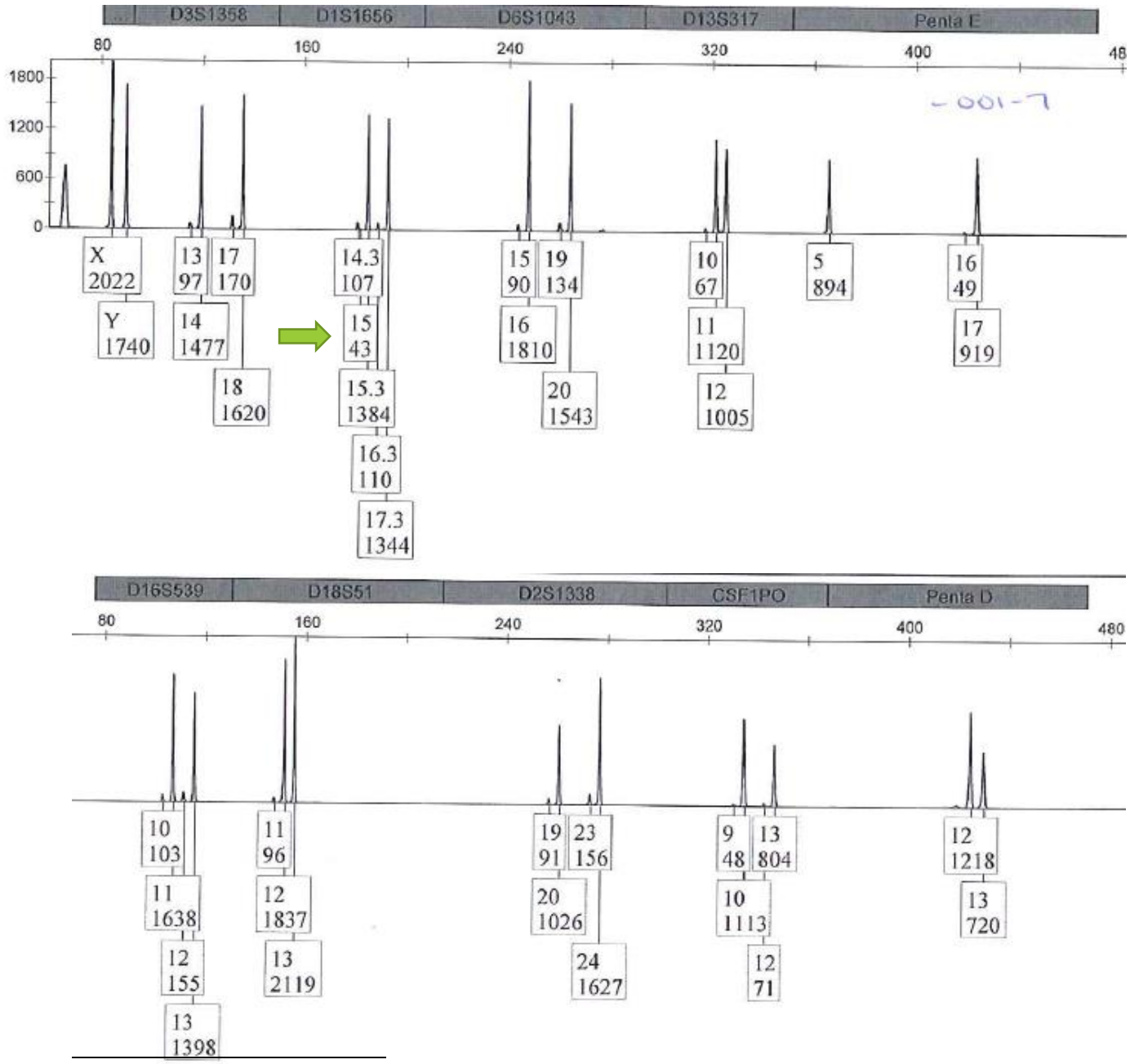
[51] QHFSS Verification of the DNA Profile Analysis module of STRMix using the Promega PowerPlex 21 system, December 2012. p22, paragraph 3.

[52] The other two profiles were weaker profiles. One had no labelled peaks in stutter position, the other had seven labelled peaks in stutter position.

seventeen profiles profile has a clear major contributor and only one or two extra low-level peaks in non-stutter positions. None of the 33 extra low-level peaks were reproducible (in subsequent electropherograms or reworked to confirm the extra peaks). Therefore, they are either very low-level two-person mixtures, or the extra peaks are drop-in, or capillary electrophoresis carryover[53], and the evidence is from a single donor. An example of the one of the low-level 'two-person mixtures' is shown below with only one extra peak 45 RFU high and no other sign of a mixture (there are 27 labelled peaks in stutter position).

Figure10: Electropherogram of F2 (Entrance floor) showing one extra peak at the D1 locus (labelled 15, at 43 RFU). See A and B for full electropherogram.

A.



[53] Capillary electrophoresis (CE) carryover is the physical transfer of DNA from one injection to the next within a plate batch analysed with Genetic Analyzer.

B.

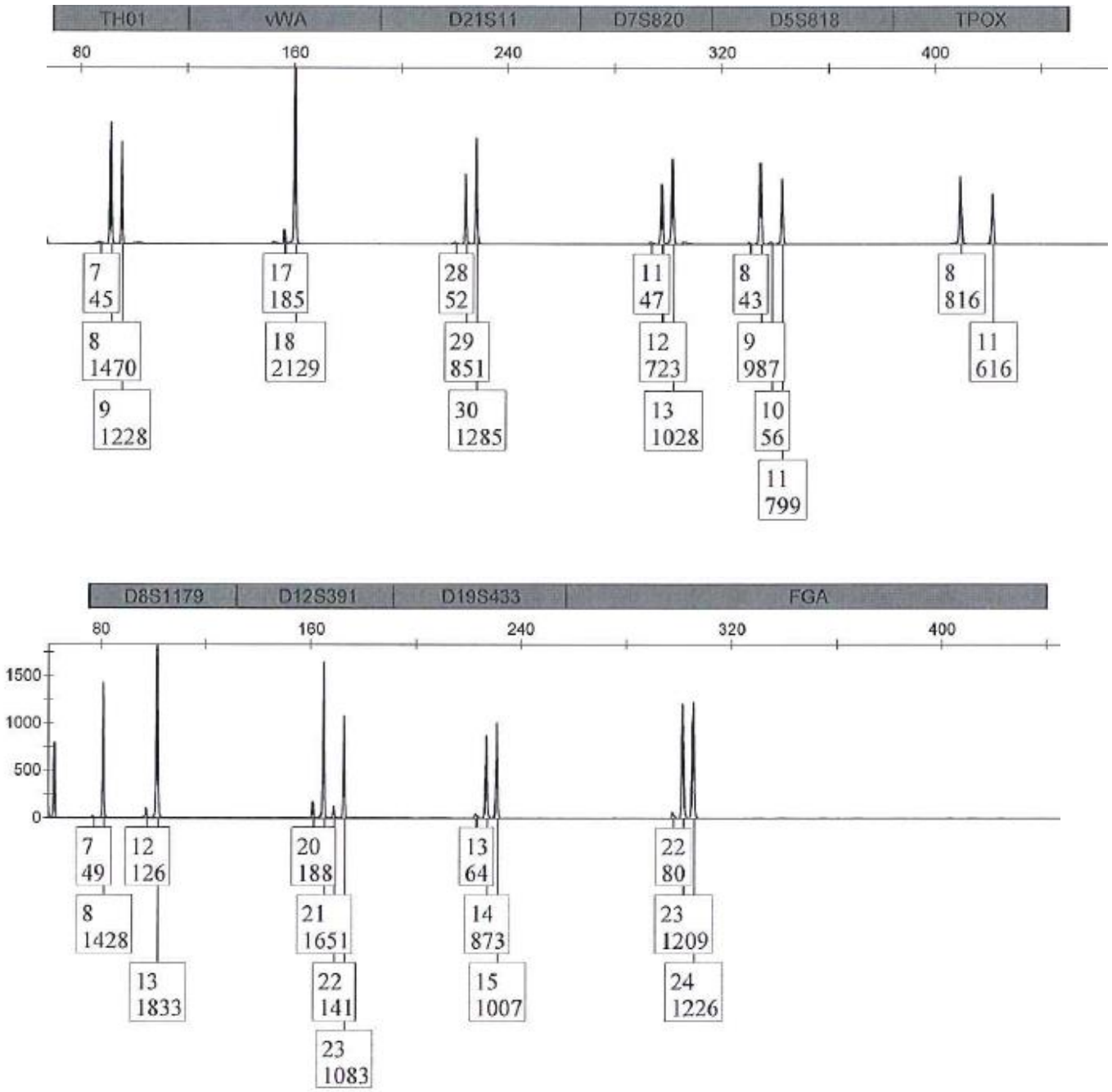

TH01
vWA
D21S11
D7S820
D5S818
TPOX
80
160
240
320
400
7
45
8
1470
9
1228
17
185
18
2129
28
52
29
851
30
1285
11
47
12
723
13
1028
8
43
9
987
10
56
11
799
8
816
11
616
D8S1179
D12S391
D19S433
FGA
80
160
240
320
400
1500
1000
500
0
7
49
8
1428
12
126
13
1833
20
188
21
1651
22
141
23
1083
13
64
14
873
15
1007
22
80
23
1209
24
1226

Table 7: Profiles reported as two-person mixtures by QHFSS. All peaks below 50 RFU should be excluded (due to the incorrect LOR), and all peaks below 100 RFU could provide incorrect LRs (due to the incorrect drop-in threshold).

Sample	No. of extra peaks and peak height (RFU)	QHFSS statement to court and LR for each person[54]	Change to QHFSS interpretation
Profiles with one extra peak below 50 RFU. Incorrectly reported as DNA mixtures because of incorrect LOR threshold.			
F2: Entrance Floor	1* (43)	Contributor: 2 Non-contributor: 2, 2, 2, 2, 2, 3, 4, 4, 5, 5, 10, 12	Single contributor
S1: Footpath	1* (45)	All excluded	Single contributor
L51: BLACKBURN's shoe	1* (43)	Non-contributor: 3, 3, 12, 16, 30	Single contributor
F8: Mobile phone	1* (44)	Non-contributor: 2	Single contributor
F4: Mobile phone	1* (49)	Contributor: 7,4,4,2,2,2 Non-contributor: 2, 2, 5, 11, 16, 18, 24	Single contributor
Profiles with 1 to 3 extra peaks below 50 RFU. LRs will be incorrect and other peaks may be incorrectly weighted as true alleles because of incorrect drop-in threshold.			
L41: BLACKBURN's pants	2* (40, 153)	All excluded	New LR
L46: BLACKBURN's pants	3* (42, 46, 53)	All excluded	New LR
L23: BLACKBURN's pants	3* (64, 41, 44)	All excluded	New LR
L43: BLACKBURN's pants	2* (42, 68)	All excluded	New LR
L45: BLACKBURN's pants	2* (40, 86)	Contributor: 8, 8, 4, 3 Non-contributor: 4	New LR
L44: BLACKBURN's pants	4* (74, 48,66, 41)	All excluded	New LR
F3: Mobile phone	2* (41, 70)	All excluded	New LR
Profiles with all extra peaks above 50 RFU. LRs may be incorrectly weighted as true alleles because of incorrect drop-in threshold.			
Swab of BLACKBURN's shin	1* (58)	Contributor: 6, 5, 3, 3 Non-contributor: 2, 2, 2, 3, 3, 3, 6, 15, 49	New LR
L22: BLACKBURN's pants	2* (65, 57)	All excluded	New LR
L45: BLACKBURN's pants	5* (51, 50, 109, 127, 81)	Contributor: 59	New LR

* Peak has not been reproduced in another electropherogram for that sample contained within the case file. All electropherograms used in the STRMix analysis were contained in the case file.

[54] LR = likelihood ratio. Seventy six reference profiles were compared against the profile QHFSS considered as a 'low-level' mixture. Where the likelihood ratio for contributors and/or non-contributors are listed, all other reference profiles are excluded.

The incorrect setting of the LOR and drop-in thresholds to 40 RFU has led to:

1. five profiles incorrectly reported as DNA mixtures, and seven people being incorrectly reported as contributors;
2. seven profiles reported with incorrect LRs, four people at risk of being incorrectly reported as contributors, and possibly numerous people incorrectly excluded; and
3. three profiles reported with incorrect LRs, possibly leading to numerous people incorrectly reported as contributors and non-contributors.

As well as reporting incorrect evidence, the police responded to the large number of 'DNA mixtures' reported by QHFSS and collected over 60 reference samples from friends, family, and work colleagues of BLACKBURN and other persons of interest in an attempt to identify the donor of the low-level 'second contributor'. When QHFSS analysis failed to provide informative results to identify the second contributor, police sent samples from the 'mixtures' to the Australian Federal Police (AFP) for Y-STR analysis.

4.3.3 Evidence from BLACKBURN's pants ([redacted], LR 8)

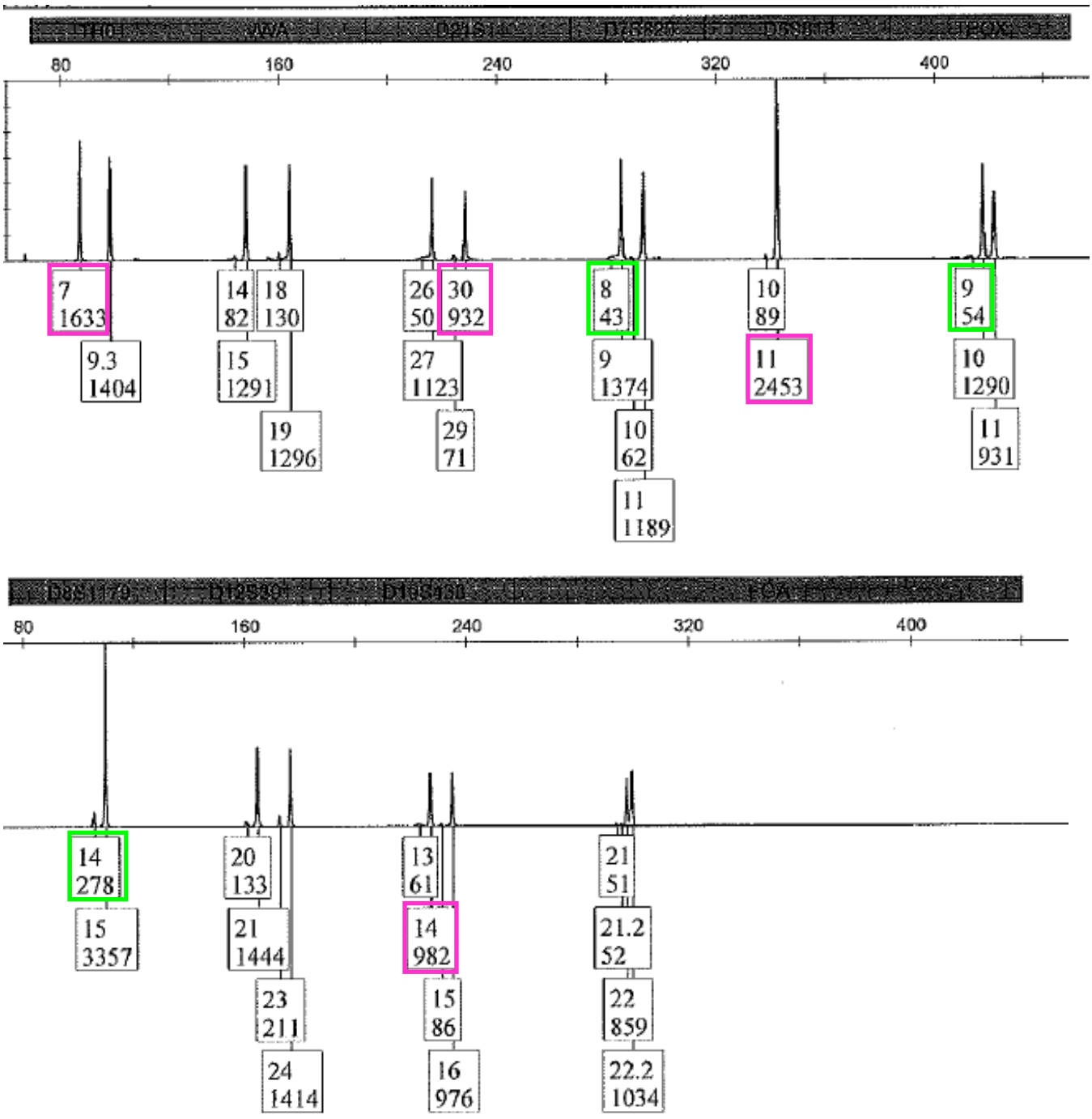
Undoubtedly the most important DNA evidence presented by QHFSS in [redacted]' trial was L45 from BLACKBURN's pants. QHFSS reported it as a two-person mixture conditioned on BLACKBURN[55], with [redacted] and three other people as possible contributors (LR 8, 8, 4, and 3)[56]. Mr EBERHARDT stated in his closing speech "*there is a forensic link between him [[redacted]] and the deceased's clothing*"[57], and "*a partial mixed DNA profile that's six times more likely*[58] *to come from him than a random member of the Aboriginal population was found on Shandee's pants*"[59]. In the Shandee's Story podcast Greven Breadsell states "*there was Aboriginal DNA all over her*"[60], and Hedley Thomas reports about a third of the trial transcript is about [redacted] being the possible offender. The absence of physical evidence linking [redacted] to the crime left a vacuum for any physical evidence presented by defence favouring an alternative offender to be amplified.

The L45 electropherogram shows 28 labelled peaks in stutter positions and two low-level peaks in non-stutter position that were all used in the STRMix component interpretation. The extra peaks were not confirmed by any re-analysis which should have occurred given drop-in is not

[55] 'Conditioned' means BLACKBURN's DNA was expected to be found because of the bloodstains on her pants, and her DNA was accounted for as a 'known contributor' in the 'DNA mixture'.
[56] File 3, p0778)
[57] R vs [redacted] [2017], p64, paragraph 45
[58] The calculation of LR 8 used the Caucasian allele frequency database, LR 6 used the Aboriginal allele frequency database.
[59] R vs [redacted], p 71, paragraph 30.
[60] Hedley Thomas interview with Greven Breadsell (30:55). Electronic file 'HT_Greven_Breadsell_box0077'

reproducible. BLACKBURN is the major contributor. The low-level 'second contributor' is therefore potentially probative and required accurate interpretation.

Figure 11: Electropherogram of L45 (BLACKBURN's pants). Green boxes are peaks labelled in stutter position consistent with [redacted]'s DNA. Pink boxes are true alleles consistent between BLACKBURN and [redacted]. Dashed blue boxes are the two extra peaks, both consistent with [redacted].



Coincidentally, [redacted] has eight alleles consistent with peaks in stutter position of BLACKBURN's true alleles, and eleven alleles that are the same as BLACKBURN's. By chance, at 15 out of 20 loci [redacted] has either one or two alleles that match peaks in stutter position or true alleles from BLACKBURN. The two extra low-level peaks are consistent with [redacted]. This information was used in the component analysis in STRMix.

Table 8: [redacted]'s DNA profile. Bold numbers are alleles consistent with peaks in stutter position or true alleles from BLACKBURN's DNA.

D3	D1	D6	D13	Penta E	D16	D18	D2	CSF1PO	Penta D
16,17*****	**11,15**	10,**18**	9,9	**16,17**	**11,11**	**14**,16*****	**19**,21	**10,10**	**10**,11

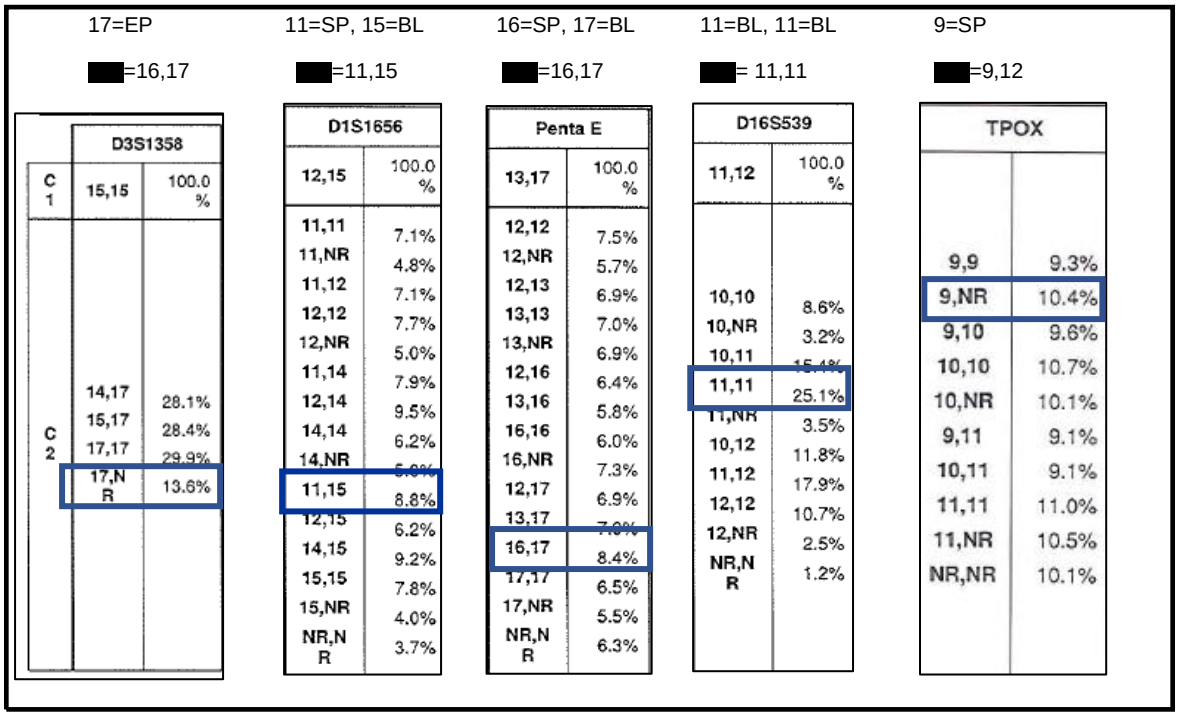
THO1	vWA	D21	D7	D5	TPOX	D8	D12	D19	FGA
6,**7**	17,21	**30**,36.2	**8**,12	**11**,13	**9**,12	10,**14**	19,19	**14**,15.2	19,20

***** Extra peak in L45.

The STRMix component interpretation has weighted the probability of genotypes for each locus in the L45 profile. For loci where peaks in stutter position and BLACKBURN's true alleles are consistent with [redacted] the weighting is high compared to other possible genotype combinations for 'contributor two' (C2).

If the LOR was set to 50 RFU, one of the two extra peaks (40 RFU) would not be labelled and therefore excluded from analysis. If the STRMix drop-in threshold was not incorrectly set at 40 RFU by QHFSS, and instead set at a conservative 100 RFU, it is possible the second extra peak (86 RFU) would have been evaluated as drop-in by the software and given a lower weighting. The LR for this locus was the highest across all loci, therefore had a significant impact on the overall LR and [redacted] being nominated as a contributor. One or both changes could lead to [redacted] being reported as a non-contributor.

Figure 12: STRMix interpretation component analysis of L45. The genotype consistent with [redacted] is highlighted in each rectangle. The percentage in each right column represents the weighted probability of each genotype.



EP= extra peak; SP= Peak in stutter position; BL= true allele from BLACKBURN; WD= [redacted]'s genotype; NR= No reportable data (STRMix then considers any second allele from a suspect could accompany the first allele).

QHFSS were aware due to their interpretation of this profile, [redacted] was the main defence contention[61]. Records show the DNA statement was requested "*due to the defence line of questioning during the committal court process in which they presented DNA intelligence reports and wanted an explanation about other persons outlined in one of them (namely* [redacted]*).*"[62] Prior to the DNA statement being released, QPS highlighted to QHFSS that defence were trying to understand the meaning of the LR for [redacted]. QPS asked if QHFSS could "*provide some sort of explanation around the stat's meaning*"[63], however, the reporting scientists stated "*that it would be best to request an SOW*[64] *on the whole case rather than potentially do it piecemeal and that it would be about 4 weeks from request to release*".

These events are alarming for the following reasons:

1) QHFSS knew this profile was now the key piece of evidence in the [redacted] trial, particularly because there was no DNA linking [redacted]; and
2) There was time for QHFSS to re-analyse the sample to confirm if the two extra weak peaks were in fact true alleles, drop-in, or from capillary electrophoresis carryover.
3) There was time for QHFSS to review the STRMix outputs and see that one extra peak caused [redacted] to be nominated as a contributor.
4) QHFSS failed to assist police and defence when they requested an explanation about what the statistic relating to [redacted] actually meant. Defence may not have emphasised the importance of this evidence if QHFSS explained clearly that an LR of 8 was uninformative in terms of human identification and that approximately 12% of the population would have the same pieces of DNA.

When the reporting scientist presented this evidence in the trial under cross-examination, he stated the LR for [redacted] was "*so close to one as to be meaningless*"[65] The judge directed the jury to leave the court and commented about the QHFSS scientist "*he set about minimising the significance of the evidence in a way that was odd*[66]" and "*it's not appropriate for a witness like this to categorise such evidence as meaningless*[67]. When the jury returned, they were advised the QHFSS scientist could not evaluate the evidence as meaningless.

It is unknown if this evidence affected the outcome of the trial, but it certainly provided physical evidence to support an alternative offender put forward by defence.

[61] File 1, p1936, paragraph 4
[62] File 1, p1874 QPRIME task, paragraph 1.
[63] File 1, p 1934, paragraph 3.
[64] SOW=statement of work
[65] R vs [redacted], Day 6, p30, paragraph 9.
[66] R vs [redacted], Day 6, p37, paragraph 8.
[67] R vs [redacted], Day 6, p39, paragraph 2.

4.3.4 QHFSS STRMix internal validation report

There are signs in the QHFSS STRMix internal validation report[68] indicating the incorrect user-defined settings and use of half-volume reactions may have impacted on the accuracy of results.

1. The document reports an instance where the profile of a single contributor was incorrectly interpreted, which they state: "*this would lead to the incorrect genotype being uploaded to NCIDD*[69]", and another instance where this would have only been just avoided.

2. For mixture deconvolution experiments the report states: "*in a number of samples where the small contributors had low template levels (especially for the half volume amplifications), the smaller contributor was excluded by STRMix, despite them being known contributors*[70]".

   False exclusion of a known contributor was observed at least four times in the half volume reactions; twice in three person mixtures, and twice in two-person mixtures. Alarmingly, this occurred both times in two-person mixtures with a template of 0.5ng (the optimal PowerPlex 21 DNA quantity) for 20:1 and 5:1 ratio mixtures[71] which would be expected to provide reliable results. False exclusion was not observed in full volume reactions.

3. QHFSS acknowledge DNA mixtures with low-level contributors were not reliably interpreted during the validation study. "*The PowerPlex 21 Amplification of Extracted DNA Samples Validation document discusses the stochastic effects observed with low template samples. This verification backs up the observation that DNA profiles derived from samples where the input template reaches the levels often described as 'low copy number' (100-150pg) might not be reliably interpreted (especially with respect to mixtures)*."[72] Low-level DNA is often associated with second and third contributors in DNA mixtures, despite the overall concentration reaching analysis thresholds, and is therefore expected.

QHFSS observed obvious signs STRMix was not operating at its intended potential, however, proceeded with implementation and using it on crime scene evidence rather than troubleshooting. It must have been clear after OQI#34817 was reported that data used in the

[68] Verification of the DNA Profile Analysis module of STRMix using the Promega PowerPlex 21 system. December 2012.
[69] p15, paragraph 4. NCIDD = National Criminal Investigation DNA Database
[70] p16, paragraph 4.
[71] Tables A1 and A2, p20-21.
[72] p17, paragraph 2.

PowerPlex 21 and STRMix internal validations was unreliable, or some user-defined settings were wrong, but there is no indication that re-analysis of the STRMix internal validation occurred.

Despite noticing clear limitations for correct profile interpretation of low-level second contributors in mixtures, they failed to place either limitations on reporting these results or highlight the risk of error when they were reported. L45, a sample from BLACKBURN's pants ([redacted], LR 8), is an example where the 'second contributor' of this sample is in very low proportions to the major contributor. STRMix estimated the mixture proportions for contributor 1 was 96%, and contributor 2 at only 4%[73] (about a 50:1 ratio mixture). According to concerns raised in the validation report, QHFSS knew this evidence was at risk of being incorrect, though no warnings were included in the DNA statement or articulated during court testimony in the [redacted] trial.

4.4 Summary

There are at least two key settings informed by the 2012 QHFSS internal validation of PowerPlex 21 that are wrong due to the incorrect setting on Genetic Analyzer B. The LOR threshold should be set at least 20% higher and the drop-in threshold should be increased (possibly up to 100 RFU). These settings are crucial for accurate reporting of crime scene evidence, in particular two and three-person DNA mixtures with low-level contributors. The ramifications of the incorrect LOR and drop-in thresholds for over several months would have been clear to QHFSS management. It should have led to re-analysis of all low-level DNA mixtures by STRMix, and where needed evidentiary statements re-issued and the police and courts notified of potential errors in DNA evidence already presented. A change in LR, however, may not lead to a change in the verbal scale used by QHFSS to describe the LR.

The errors and flawed data in the 2012 PowerPlex 21 validation were not disclosed in the re-issued 2013 validation report. Re-analysis of the affected samples in a separate report prove QHFSS were aware of the incorrect thresholds though failed to change them. All BLACKBURN crime scene evidence was therefore analysed with incorrect thresholds, and incorrect results presented to courts without any appropriate warning.

The profile from BLACKBURN's pants QHFSS reported as a two-person mixture with [redacted] as a contributor (LR 8) was the most critical piece of DNA evidence presented in [redacted]'s trial is incorrect. It is possible after re-analysis with correct thresholds, [redacted] will not be reported as a contributor.

[73] File 3, 0773 STRMix output.

The possible ramifications of the flawed validation studies include:

1. incorrect interpretation and reporting of DNA evidence;
2. incorrectly reporting a victim, offender or innocent person as a contributor or non-contributor to crime scene evidence;
3. incorrectly classifying a single contributor profile as a DNA mixture or adding an extra person to a genuine mixture. The incorrect DNA information may coincidentally match to a suspect, or reported as an unknown person introducing doubt into prosecution and defence scenarios;
4. incorrect profiles uploaded to the National Criminal Investigation DNA Database;
5. reporting of incorrect likelihood ratios;
6. incorrect DNA intelligence provided to police; and
7. incorrect DNA evidence presented in statements and court testimony.

Recommendation 5: All DNA mixtures reported by QHFSS for the BLACKBURN matter require re-analysis by an independent expert after the user-defined settings are corrected.

Recommendation 6: All DNA mixtures relating to other matters require evaluation and where required, re-analysis after the user-defined settings are corrected.

Recommendation 7: An independent investigation is needed of the initial PowerPlex 21 and STRMix internal validation studies and the amended PowerPlex 21 internal validation to determine if any other issues and errors exist. Potentially re-validation is required. A course of action is required to re-analyse affected samples.

5.0 Inaccurate reporting of DNA evidence

5.1 Background

QHFSS scientists are relied upon to accurately interpret and report DNA evidence to the police and courts. Limitations and uncertainties of DNA evidence must be disclosed to enable reliable decisions on admissibility and allow the jury to properly weight the evidence. There is evidence in the BLACKBURN case file that QHFSS are not accurately reporting DNA results in key areas.

1. Incorrect classification of single contributor partial profiles as complex mixtures that cannot be interpreted.
2. Incorrectly using unlabelled sub-threshold information to exclude possible contributors and non-contributors from DNA mixtures.
3. Incorrectly reporting a mixture as unsuitable for 'meaningful interpretation' when it could be analysed with STRMix.

5.2 Incorrect classification of single contributor partial profiles as complex mixtures that cannot be interpreted

Four crime scene samples which are clearly single contributor partial profiles, were reported to the police and court in the BLACKBURN DNA statement as:

'*The mixed DNA profiles obtained from these samples indicate the presence of DNA from an indeterminate number of contributors. Given the uncertainty as to the number of contributors, the results of these samples are, in my opinion, unsuitable for meaningful interpretation*[74].'

Three of these samples were retrieved from [redacted]' car.

1. V13: tapelift front driver's seat belt (File 5, p0159)
2. V24: Mount Franklin water bottle (File 5, p0150)
3. V41: Coke bottle from rear passenger's side footwell (File 5, p0141)
4. Tapelift of BLACKBURN's left forearm (File 2, p1114)

The interpretations are incorrect, and so is the evidence presented to courts. Is it possible these samples were affected by the defective Proteinase K, so QHFSS used what they thought was a neutral reporting option?

[74] DNA Statement dated 29 September 2016, p29.

Figure 13: V13 electropherogram (front driver's seat belt).

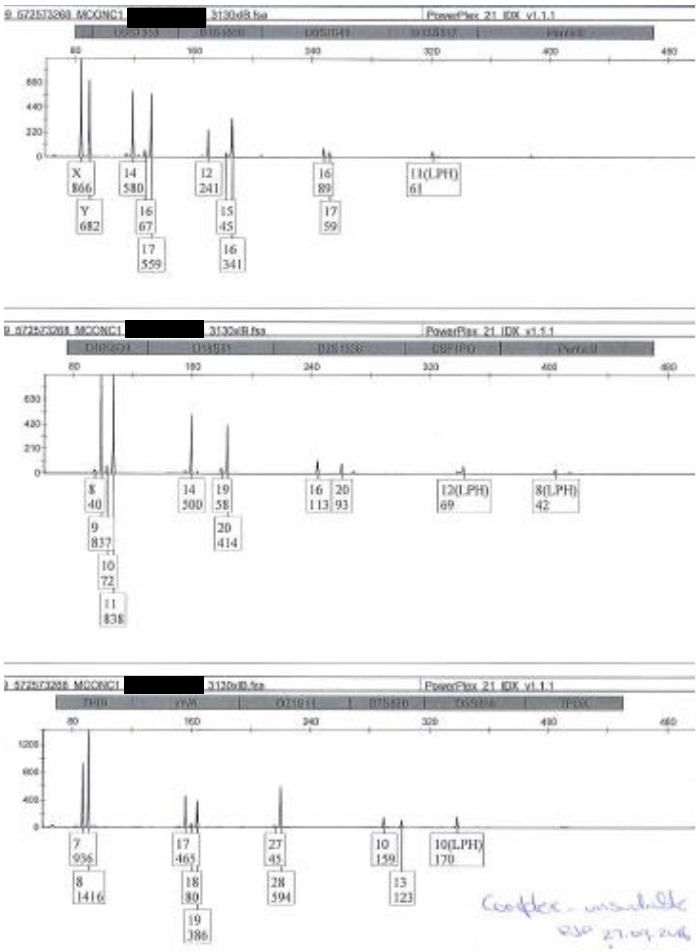


Table 9: History of V13 DNA results (front driver's seat belt)

Date	Status
6 May 2014	Sample received by QHFSS
21 May 2014	Electropherogram dated 21 May 2014 (File 5, p0159). Clear partial profile matching [redacted]. Coded as having undergone Microcon concentration.
11 July 2014	Reported in QPRIME as 'undergoing rework'.[75]
21 July 2014	Electropherogram dated 21 July 2014 (File 5, p0155). Clear partial profile matching [redacted]. No update of QPRIME based on this result.
5 Sep 2014	E-mail from QPS to Justin Howes requesting update on this sample
24 Sep 2014	Reported on QPRIME as a two-person mixture
1 Oct 2014	Reported on QPRIME as linking to [redacted] (LR = >100 billion)
28 Sep 2016	Reported on QPRIME as 'complex mixed profile unsuitable for interpretation or comparison'.
29 Sep 2016	DNA statement issued (complex mixed profile unsuitable for interpretation or comparison).

[75] The QHFSS entry on QPRIME states: *'This is not a final result, sample/s are currently undergoing rework. Rework can mean that part of the process to obtain a DNA profile is repeated or additional testing to improve the DNA profile is being undertaken. This rework could be due to: instrument failure, requiring the sample to be re-processed; interpretation difficulties, requiring the sample to be re-run to resolve any issues. Final results are pending'.*

Figure 14: V24 electropherogram (Mount Franklin water bottle)

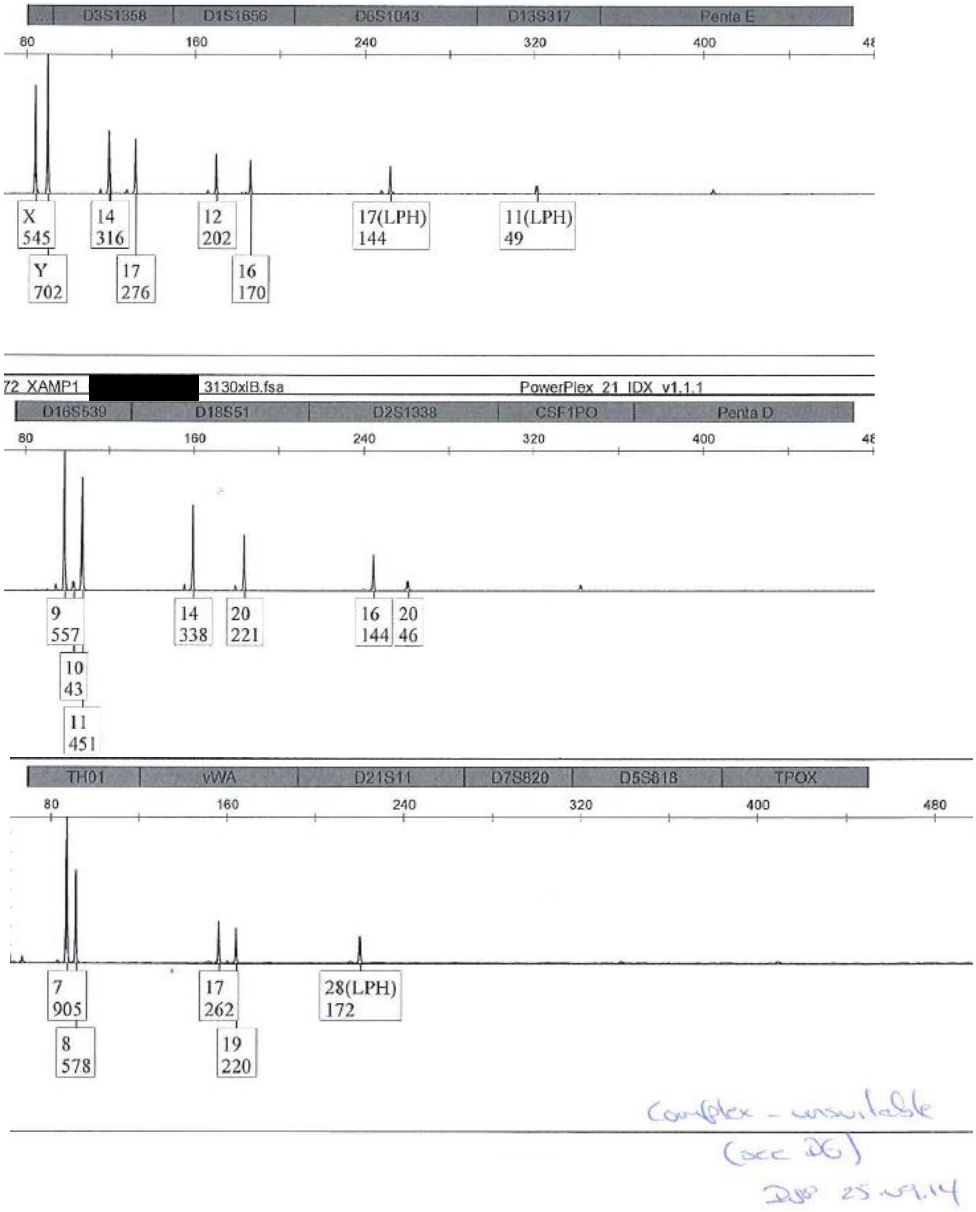


Table 10: History of V24 DNA results (Mount Franklin water bottle)

Date	Status
6 May 2014	Sample received by QHFSS
22 May 2014	Electropherogram dated 22 May 2014 (File 5, p0146). Clear partial profile matching [redacted]. Coded as having undergone Microcon concentration.
30 June 2014	Reported on QPRIME as 'undergoing rework'.
10 July 2014	Electropherogram dated 10 July 2014 (File 5, p0150). Clear partial profile matching [redacted].
5 Sep 2014	E-mail from QPS to Justin Howes requesting update on this sample
26 Sep 14	Reported on QPRIME as 'complex mixed profile unsuitable for interpretation or comparison'.
29 Sep 2016	DNA statement issued (complex mixed profile unsuitable for interpretation or comparison).

Figure 15: V41 electropherogram (Coke bottle from rear passenger's side footwell)

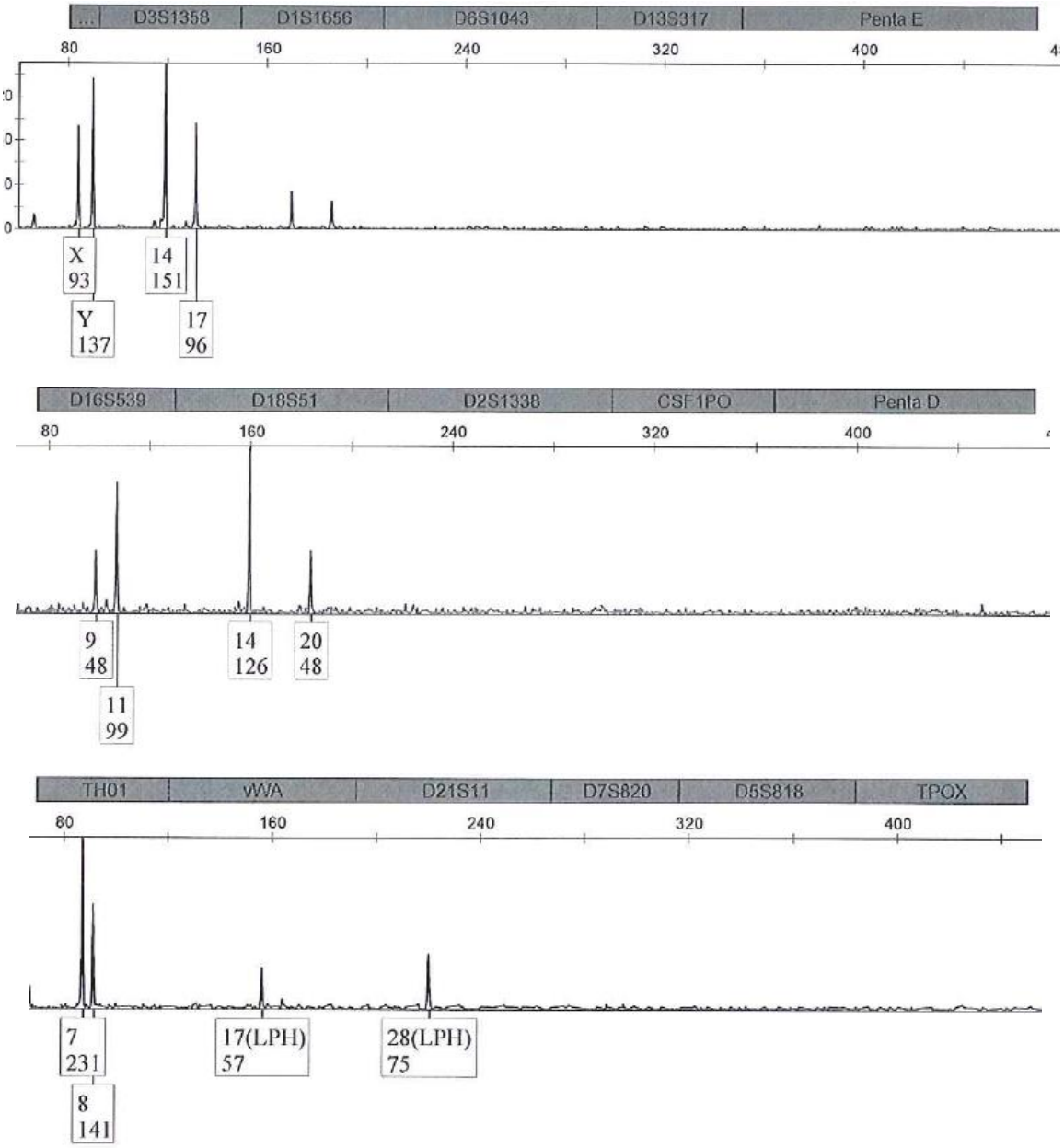


Table 11: History of V41 DNA results (Coke bottle from rear passenger's side footwell)

Date	Status
6 May 2014	Sample received by QHFSS
21 May 2014	Electropherogram dated 21 May 2014 (File 5, p0141). Clear partial profile matching [redacted]. Coded as having undergone Microcon concentration.
30 May 2014	Electropherogram dated 30 May 2014 (File 5, p0137). Clear partial profile matching [redacted].
3 July 14	Reported on QPRIME as 'complex mixed profile unsuitable for interpretation or comparison'.
29 Sep 2016	DNA statement issued (complex mixed profile unsuitable for interpretation or comparison).

Figure 16: Electropherogram, tapelift BLACKBURN's left forearm

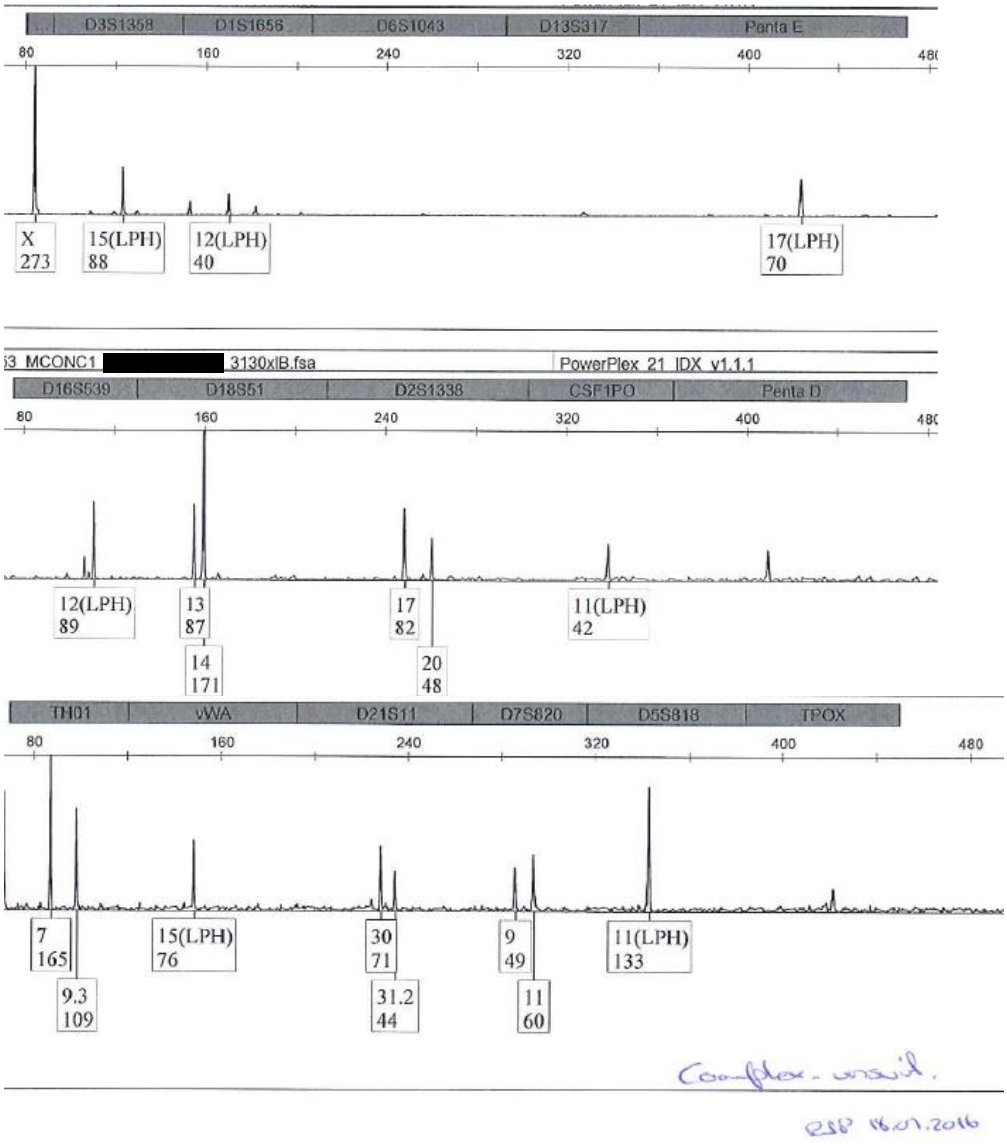


Table 12: History of results, tapelift from BLACKBURN's left forearm

Date	Status
22 Feb 2013	Sample received by QHFSS
28 Feb 2013	Reported on QPRIME as 'no DNA detected'.
1 May 2014	QPS request rework of the sample (File 1, p1869).
8 May 2014	Electropherogram dated 8 May 2014 (File 2, p1114). Clear partial profile. All alleles but one match to BLACKBURN[76]. Coded as having undergone Microcon concentration.
8July 2014	Reported on QPRIME as 'complex mixed profile unsuitable for interpretation or comparison'.
29 Sep 2016	DNA statement issued (complex mixed profile unsuitable for interpretation or comparison).

[76] At D21 the genotype is 30, 31.2 (70RFU, 44RFU). BLACKBURN's reference is 27,30. There is no sign of the 27 allele. The 31.2 appears to be drop-in and suggests the QHFSS drop-in threshold is therefore too low.

To report DNA results as complex mixtures that cannot be interpreted when they are clearly single contributor profiles is completely inaccurate and misleading to the police and courts, and disregards potentially informative DNA results. It is not clear why QHFSS have undertaken this practice. These results were peer reviewed by a second scientist and passed through their quality management system. The 'rework' reporting phrase from QHFSS in QPRIME states: "*This rework could be due to: instrument failure, requiring the sample to be re-processed; interpretation difficulties, requiring the sample to be re-run to resolve any issues. Final results are pending.*[77]*"* Is it possible the profiles are deemed unreliable due to an instrument failure so QHFSS reported them as complex mixtures that cannot be interpreted to conceal this? Were these samples affected by the defective Proteinase K. If so, was probative information lost that could identify the offender, or link BLACKBURN to [redacted]' vehicle? Alternatively, are the QHFSS scientists so poorly trained in basic aspects of profile interpretation they cannot distinguish the difference between a single contributor profile and a basic mixture?

5.3 Incorrect use of unlabelled sub-threshold information for mixture interpretation

During internal validation of PowerPlex 21, QHFSS set the LOR at 40 RFU. All labelled peaks in an electropherogram are included in STRMix analysis after examination by a scientist to remove any obvious artifact peaks. STRMix uses the information from the peaks and various parameters to deconvolute mixtures, then uses a reference comparison database (that includes suspects, victim and elimination profiles for a specific case) to probabilistically determine whether a person could be considered a 'contributor' or a 'non-contributor' to the DNA mixture, and then assign a likelihood ratio.

For six crime scene samples, QHFSS have followed this process and reported results to QPRIME for contributors and non-contributors in 2013 and 2014. In 2016 the reporting scientist has then noted for each mixture below, that unlabelled sub-threshold peaks in corresponding electropherograms were used to exclude 78 people previously reported as either contributors or non-contributors[78]. This was performed post-STRMix analysis, the unlabelled sub-threshold peaks were not included in the STRMix analysis. The court statement does not include the people removed using this incorrect practice and therefore the evidence is inaccurate.

1. S1: Swab of blood from footpath (File 2, p1317);
2. F5: Swab from mobile phone, BLACKBURN (File 4, p0728);
3. F8: Swab from mobile phone, BLACKBURN (File 5, 0276);
4. L22: Tapelift from right hand side buttock area of pants, BLACKBURN (File 4, p0670);

[77] '[redacted] QPRIME Records', p253

[78] See various handwritten entries in STRMix reports for corresponding electropherograms.

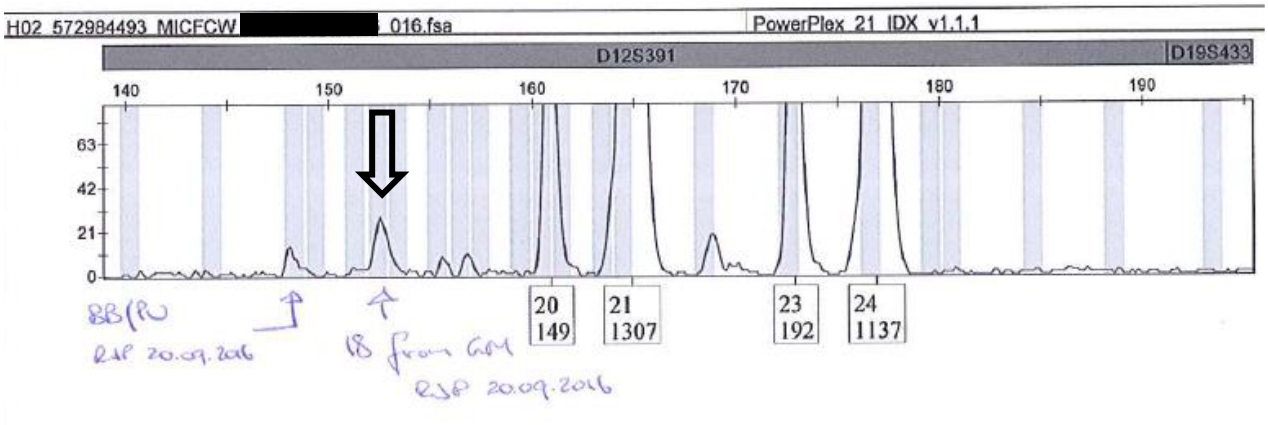
5. L43: Tapelift from rear right hand side lower leg area of pants, BLACKBURN (File 3, p0808; and
6. L51: Swab from front upper label of left-hand side shoe, BLACKBURN (File 3, p0962).

It is unclear why QHFSS would perform this extra step post-STRMix analysis that was not included in the original analysis two years prior. This extra step is not a reliable method given unlabelled sub-threshold peaks are at high risk of being artifacts or drop-in, and therefore risks falsely excluding offenders or victims from DNA mixtures.

Table 13: Samples where unlabelled sub-threshold peaks were used to exclude people from being contributors and non-contributors post-STRMix analysis.

Sample	Locus, and aprox. sub-threshold peak height (RFU)	No. of people removed
S1	D18: ~25 RFU	1
F5	D1: ~20 RFU; D2: ~ 20 RFU; D8: ~ 30 RFU	20
F8	D12: ~25 RFU	6
L22	D3: ~20 RFU; Penta E[79] ~ 40 RFU; D12: ~50 RFU	20
L43	CSF1PO: ~25 RFU	6
L51	D12: ~25 RFU	25
	Total	**78**

Figure 17: Electropherogram from L51 demonstrating the scientist's allocation of the 'sub-threshold peak' as an 18 allele at D12[80]. BB/PU is 'bad base line' and 'pull-up'.



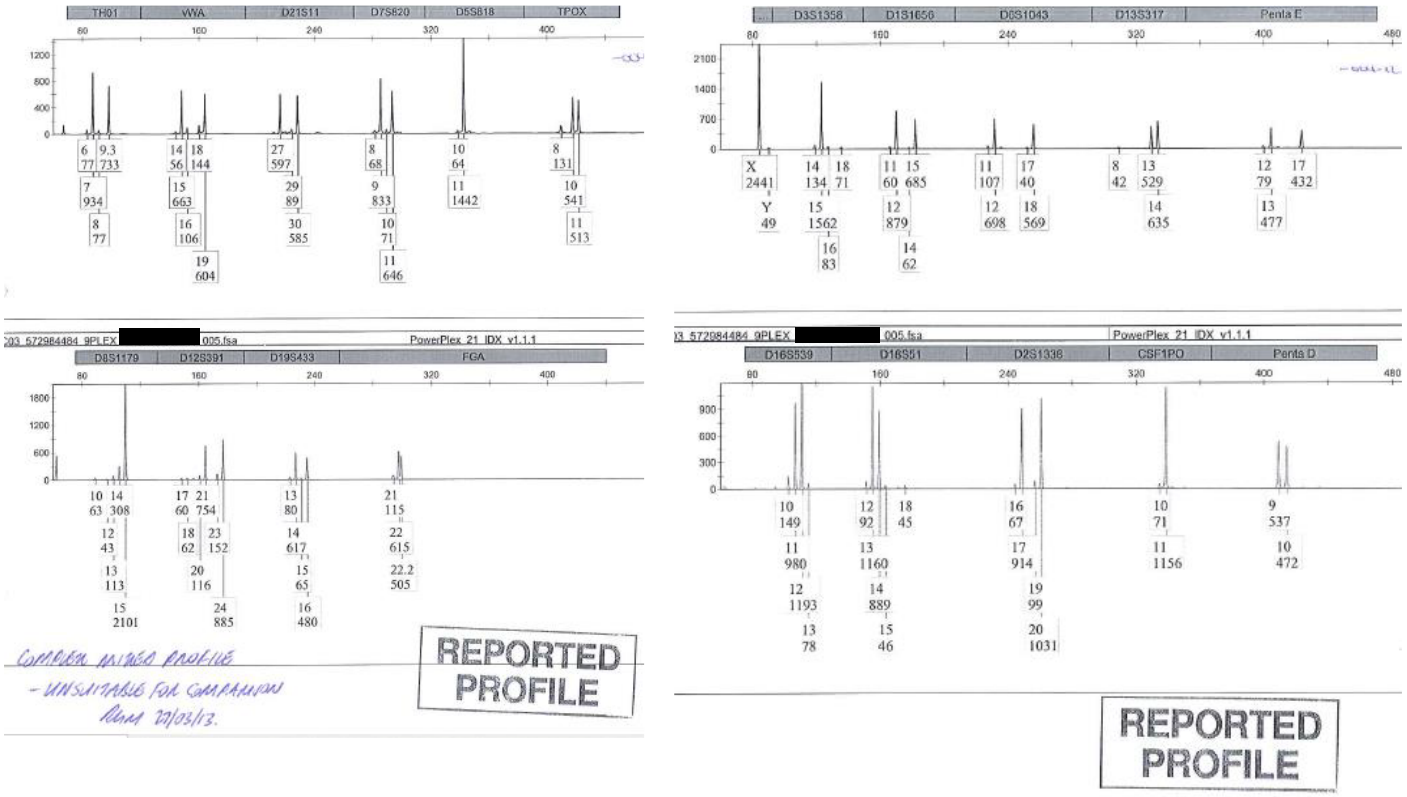
[79] Handwritten note on p0661 states CSF1PO locus contained a sub-threshold peak, yet electropherogram on p0670 shows a circled peak at Penta E.
[80] File 3, p0962

5.4 Incorrectly reporting a mixture as unsuitable for 'meaningful interpretation'

STRMix software was implemented by QHFSS to enable interpretation of DNA profiles considered too complex to interpret using the previous binary method. This provides an opportunity to deconvolute DNA mixtures and compare components to a nominated reference profile set, with subsequent generation of likelihood ratios. QHFSS use this process for DNA mixtures containing two and three contributors, but DNA profiles indicating four or more contributors were reported as complex mixtures and no interpretation was made. QHFSS attempted to validate STRMix for four-person mixtures but failed because they used computers with the wrong specifications. The scientist evaluates all DNA mixtures, and those indicating two or three contributors based on information in the most informative loci, will be analysed with STRMix.

A DNA mixture from BLACKBURN's shoes was reported as: '*[these samples] indicate the presence of DNA from an indeterminate number of contributors. Given the uncertainty as to the number of contributors, the results for these samples are, in my opinion, unsuitable for meaningful interpretation*'[81]. This result was first reported in QPRIME on 2 April 2013.

Figure 18: Electropherogram from L50, BLACKBURN's shoes[82].



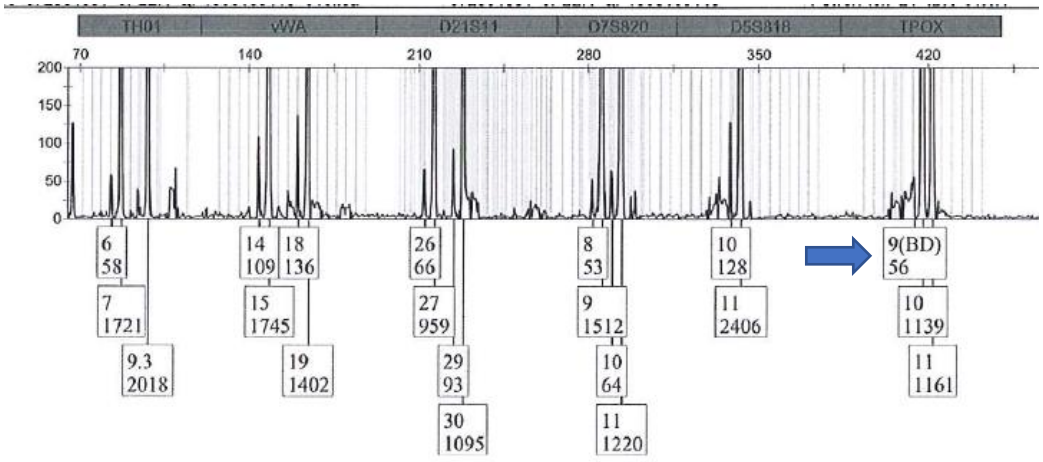
[81] L50: tapelift from BLACKBURN's shoes. DNA statement p24.
[82] Electropherogram dated 12 March 2013, File4, p0515 to p0516.

It is unclear why this profile was not analysed with STRMix as it appears likely to be a three-person mixture. Even accounting for possible stutter, there a no more than six peaks labelled at the most informative locus, with peak heights suggesting the most likely number of contributors as three. It is unknown whether this is a single incident where a DNA mixture was not analysed, or whether this is part of a systemic issue that should be investigated further. Though, a partial profile from V38 ([redacted]' vehicle) is also reported as a complex mixture with an indeterminate number of contributors that could not be interpreted but appears to be a three-person low-level mixture[83]. Potentially, this could be ignoring valuable evidence that could link a victim or offender to a crime scene.

5.5 Inclusion of an artifact in mixture interpretation

One DNA mixture (L40) contained a peak GeneMapper software labelled as 'BD' (broad), meaning it exceeded the maximum peak width defined in the analysis method and may be unreliable. The analyst is required to manually remove labels from artefacts prior to STRMix analysis to avoid incorrect outputs[84]. In this instance, despite the peak being labelled with a poor-quality flag, the analyst ignored the quality warning, and the information was used in the STRMix component interpretation. This is a concerning practice that could lead to incorrectly nominating victims and suspects as contributors or non-contributors of a DNA sample.

Figure 19: Use of a poor-quality information in STRMix two-person mixture analysis. It is clear this sample has a bad baseline[85] and should be repeated.



[83] (File 5, p0132)
[84] QHFSS Interpretation and Statistical Analysis of NDA profiles Using the STRMix Expert System (v1), p4, paragraph 8.
[85] Bad baseline is when the horizontal axis has 'messy peaks'. It represents too much background noise and risks incorrect results.

5.6 Systemic QHFSS errors previously affecting criminal cases

QHFSS have reported incorrect evidence for DNA mixtures over many years. In March 2015 The Courier Mail reported errors in mixture interpretation affected 60 matters, with replacement statements needed for 24[86]. QHFSS blamed 'a minor mis-code' in a version of STRMix used between July 2014 to January 2015 for the incorrect results. However, STRMix developers state the errors were caused by QHFSS incorrectly using the software due to not purchasing an updated user manual, and stated:

"*When we looked at the circumstances needed to cause this, we thought it was almost impossible. We can't replicate it. The question would be, have they followed recommended processes? Are they following the manual? They haven't bought the manual. Every other lab has bought it*."

In January 2019, QPS reported they undertook a review of all cases involving DNA mixtures of three or more contributors between 2013 and 2018[87]. Results released by QHFSS were recalled "*due to the uncertainty as to the number of contributors within a mixture of DNA*". The QPS review identified 138 matters where they believed the DNA evidence was potentially significant to the prosecution and reported "*as a result of the review the integrity of all cases has been confirmed*[88]". There were no explanations from QHFSS about what caused the systemic errors over so many years. It is also concerning that their quality system did not detect the errors prior to evidence being provided to the police and courts, demonstrating multiple levels of failure. There is a high risk that other incorrect DNA evidence has been released by QHFSS, though has gone undetected.

5.7 Summary

QHFSS have incorrectly interpreted twelve crime scene profiles for the BLACKBURN matter, and incorrectly nominated 78 people as either contributors or non-contributors to evidence due to flawed practices. Inaccurately reporting single contributor partial profiles as complex mixtures that are not suitable for interpretation and using unlabelled sub-threshold peaks post-STRMix analysis demonstrates even basic profile interpretation principles are not being followed. These

[86] 'Queensland authorities confirm 'mis-code' affects DNA evidence in criminal cases.' Courier Mail, 20 March 2015. https://www.couriermail.com.au/news/queensland/queensland-authorities-confirm-miscode-affects-dna-evidence-in-criminal-cases/news-story/833c580d3f1c59039efd1a2ef55af92b

[87] https://mypolice.qld.gov.au/news/2019/01/31/improvement-to-dna-analysis-introduced/ Accessed 25 July 2022.

[88] 'More than 130 cases of mixed DNA samples recalled in Queensland.' Brisbane Times, 31 January 2019. https://www.brisbanetimes.com.au/national/queensland/more-than-130-cases-of-mixed-dna-samples-recalled-in-five-years-20190131-p50uw0.html Accessed 25 July 2022.

issues evaded peer review and the QHFSS quality assurance system and genuinely risks offenders evading identification and incorrect judicial outcomes.

There is previous evidence of systemic mixture interpretation errors affecting criminal matters. Errors detected between July 2014 to January 2015 caused by what QHFSS claim was a ‘minor mis-code’, and more recently in 2019 where police conducted a review of mixtures released between 2013 and 2018 potentially affecting 138 matters. It is clear there are serious and systemic issues inhibiting QHFSS from accurately reporting DNA evidence to the police and courts.

Recommendation 8: The competency of QHFSS scientists to correctly interpret DNA mixtures needs to be independently assessed, and where needed, further training provided. QHFSS standard operational procedures need to be independently reviewed.

6.0 No record of suspects' DNA being compared to BLACKBURN crime scene evidence

[redacted] and [redacted] were nominated by police as among the top six suspects in BLACKBURN's murder[89]. Their reference DNA samples were sent to the AFP for Y-STR analysis and comparison against two batches of crime scene samples that also underwent Y-STR analysis (these were a selection of samples QHFSS reported as DNA mixtures). The results of these comparisons were reported by the AFP on 30 October 2013 and 21 February 2014. [redacted] and [redacted] were also included as suspects in the 2019 Coronial inquest.

The police running log, the Forensic Register, QPRIME (including all intelligence reports) and the DNA statement does not contain any record of [redacted] and [redacted]'s DNA being compared against any crime scene samples. While their DNA may have previously been on NCIDD, the only two DNA profiles uploaded to NCIDD was from a cigarette butt recovered from a drain at the intersection of Juliet and Harriet St, and from a pile of cigarette butts near the Girl Guide's hut. None of the DNA profiles from BLACKBURN's body, clothing, or mobile phone were uploaded to NCIDD.

There were no receipts for [redacted] and [redacted]'s reference DNA in the BLACKBURN case file (among the other receipts, see File 1 p1733 onwards). An intelligence report dated 26 July 2014 lists the barcodes of all reference DNA profiles compared to crime scene samples at that time[90]. None of the barcodes correspond to [redacted] or [redacted], and there are no barcodes that could not be attributed to an unamed person. 'STRMix Database Comparison' reports conducted in 2014 (e.g., File 3, p 0956) are generated for each DNA mixture analysed by STRMix. The 'database' comprises of all the reference and unknown single contributor crime scene profiles which were compared against each crime scene DNA profile for the BLACKBURN case (a total of 76). The report lists the identification number of each reference profile being compared against each crime scene profile. The names of each donor were reconciled against each reference sample identification number. [redacted] and [redacted] are not in the STRMix database. Each STRMix Database Comparison report generated in July 2016 for each DNA mixture across the five files (about a dozen) were checked and [redacted] and [redacted] are not listed in any of those STRMix reports either.

It appears that [redacted] and [redacted]'s DNA have not been matched against any crime scene profiles from the BLACKBURN case despite being nominated among the top six suspects.

[89] Police Running Log, p718.
[90] File 1, p1815

It is unclear whether this error originated in QHFSS or was an error by the QPS. Given only the DNA profiles of two cigarette butts were uploaded to NCIDD, if [redacted] and [redacted]'s DNA was on any of the other crime scene samples, their DNA would not have been matched.

The failure of QHFSS to compare [redacted] and [redacted]'s DNA to crime scene DNA not uploaded to NCIDD is confirmed in an e-mail exchange between Cathie Allen (FSS Managing Scientist) and Inspector David Neville on 20 December 2021, the day after the Australian released the article 'Shandee's Story Podcast: Did lab check DNA of high priority targets'.[91]

"Hi David,

Our Library Services have provided the below article regarding the Blackburn case. We've reviewed the information within the article and wish to advise you of the following:

- [redacted] *nor* [redacted] *are listed within the Statement of Witness (not unexpected if their samples were supplied to FSS as 'Suspect Checks')*
- *Neither* [redacted] *nor* [redacted] *are mentioned by name within the Casefile-this provides additional information to suggest that information is being provided from non-FSS source/s (as we had raised with you previously)*
- *There are 4 Intel samples associated with the casefile that do not have names attached to them or were supplied to FSS (not unexpected for an intelligence sample) and FSS are in a position where we can neither confirm nor deny that these belong to either* [redacted] *or* [redacted]. *These barcodes are* [redacted] *if the QPS wished to check.*
- *The 4 barcodes listed above were contained within an Intelligence Report issued to the QPS on the 26th of September 2014. This Intelligence Report details results of the statistical analyses of the casework samples against available reference DNA profiles, Suspect Checks and full Unknown DNA profiles.*
- *If* [redacted]*'s and* [redacted]*'s samples were meant to be delivered to FSS and haven't been, are you able to follow-up on this please?"*[92]

Inspector Neville confirms in his response to Cathie Allen on 21 December 2021:

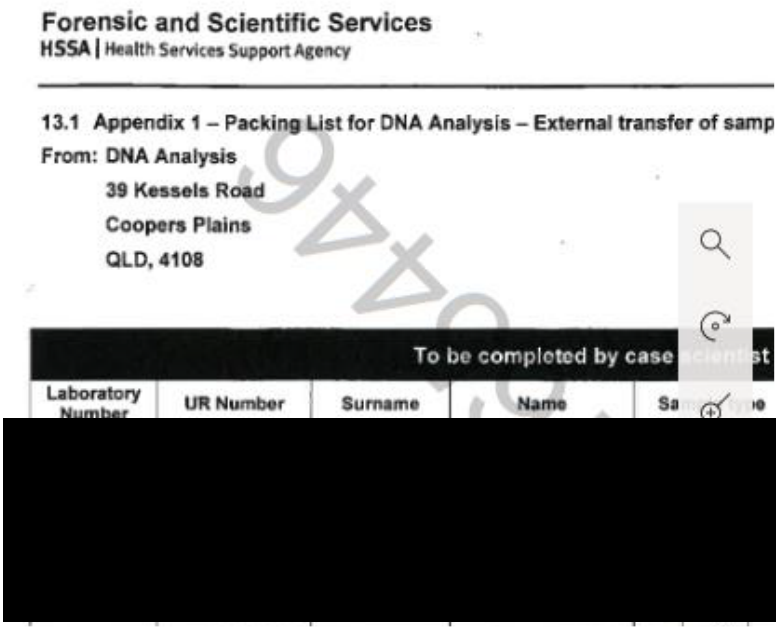
> "..*there are no further reference samples that require testing or that remain undelivered. All comparisons/eliminations have been carried out using reference samples or profiles recorded on NCIDD.*[93]"

The four barcodes listed above relate to [redacted], and [redacted] according to QPRIME records, therefore confirming that [redacted] and [redacted]'s DNA was not compared with DNA mixtures (19) from key evidence not uploaded to NCIDD. There is a record of [redacted] and [redacted]'s reference samples being sent to the AFP by QHFSS with

[91] The Australian, 19 December 2021.
[92] Page 2264 of BLACKBURN DNA case file.
[93] Page 2264 of BLACKBURN DNA case file.

the other four suspect samples on 1 July 2013[94]. QHFSS therefore had possession of the samples but a communication fault either within QPS or QHFSS prevented the samples from being matched.

Figure 20: QHFSS packing list including the six suspect samples for transfer to the AFP. [redacted] and [redacted] are the top two samples on the list.



Forensic and Scientific Services
HSSA | Health Services Support Agency

13.1 Appendix 1 – Packing List for DNA Analysis – External transfer of samp

From: DNA Analysis
39 Kessels Road
Coopers Plains
QLD, 4108

To be completed by case

Laboratory Number	UR Number	Surname	Name	Sa
[redacted]				

Recommendation 9: [redacted] and [redacted]'s DNA profiles need to be compared against the BLACKBURN crime scene profiles by an independent expert.

Recommendation 10: The failure to compare suspects' DNA to crime scene evidence should be examined by an independent expert to confirm if it is a systemic error either in QPS or QHFSS processes. If so, an investigation is needed to determine how many other cases are impacted. Where required, further DNA comparisons of crime scene DNA samples against suspect samples should be conducted by an independent expert to ensure no offenders have been falsely excluded.

[94] Page 1905 BLACKBURN DNA case file.

7.0 Conclusion

Serious and systemic flaws exist in the analysis of DNA evidence by QHFSS which has affected multiple key processes, and there have been clear breaches of public trust. These issues appear to have been present over at least a decade. How did these serious issues persist for so long undetected? When challenged about their competency, QHFSS typically state they are accredited by the National Association of Testing Authorities (NATA), and they use their accreditation status as comprehensive proof to deny issues when raised. This requires explanation to understand there are critical gaps in the current forensic science quality framework, and how the serious issues at QHFSS persisted for so long without detection.

QHFSS has been accredited by NATA under the international standard ISO/IEC 17025 since 1998. NATA conducts regular audits of QHFSS typically every four years, and as a result, they have maintained accreditation. NATA states its role is to:

> "*serve the national and public interest, by ensuring that organisations (accredited facilities) comply with relevant international and Australian standards and so are competent to provide consistently reliable outputs and data to government, industry and the wider community. NATA accreditation provides an assurance of the competence, impartiality and integrity of facilities*."

NATA plays a vital role in the national forensic science quality framework, however, is limited by what it can assess under the ISO/IEC 17025 standard. For example, professional conduct, outputs, success rates and the technical suitability of standard operational procedures and internal validation studies are not assessed. Given the range of serious and systemic issues in the QHFSS DNA analysis process and how long they have persisted undetected, it is clear the national forensic science quality framework needs improvement. There are growing concerns among the forensic community in Australia that current accreditation standards are not sufficient to ensure quality outputs are being delivered to the police and courts. A 2022 paper by leading expert Alastair Ross, '*ISO-accreditation- is that all there is for forensic science?*'[95] outlines these concerns. The gap in professional and quality oversight therefore provides an environment for poor quality and corrupt conduct to become entrenched in an organisation's culture.

Recommendation 11: The national forensic science quality framework requires additional measures to ensure the accuracy and reliability of forensic DNA analysis.

Recommendation 12: A set of QHFSS data that quantitates DNA profiling success should be collected and reported to QPS and key justice stakeholders annually.

[95] Alastair Ross & Wim Neuteboom (2022). ISO-accreditation-is that all there is for forensic science? Australian Journal of Forensic Sciences, 54:1, 2-14, DOI:10.1080/00450618.2020.1819414

Dr Tony Raymond led the investigation into the scientific issues affecting the Chamberlain case as part of the 1987 Royal Commission of Inquiry. His findings led to the establishment of the National Institute of Forensic Science, the Senior Managers of Australian and New Zealand Forensic Laboratories, the Specialist Advisory Groups, and a range of enduring national quality and training advancements that were world leading initiatives. While the QHFSS issues are likely to be distressing for the national and international forensic science community, it should be taken as a similar opportunity for significant and lasting improvement.

Appendix A: List of Recommendations

Recommendation 1: A root cause analysis needs to be conducted by an independent quality expert to confirm the period of the faulty dishwasher use and determine how many laboratory processes and samples across all cases were affected. Those crime scene samples should be evaluated for further testing and where needed addendum statements released.

Recommendation 2: An independent investigation is required to examine the failure of semen samples, including determining when the issues first started, the cause of the failure, and identify all samples affected.

Recommendation 3: The success rate of QHFSS DNA analysis requires close examination to uncover the real scope of the flawed testing and identify samples that require re-testing. The following analysis of QHFSS data needs to be undertaken by an independent expert:

a) determine the success rate of each sample type (item and collection method as per Krosch paper);

b) determine the success rate of samples submitted for DNA analysis presumptively positive for blood;

c) determine the success rate of samples submitted for DNA analysis presumptively positive for semen;

d) determine the success rate of samples submitted for DNA analysis confirmed positive for semen (confirmed through microscopy);

e) determine the success rate of samples taken from obvious stains of biological fluid (presumptively positive);

f) how many samples were reported to QPS as 'INCORRECT RESULT' by QHFSS; and

g) conduct a trend analysis on profile success rates (by item and collection method) from 2010 onwards

Recommendation 4: An independent review is required of QHFSS quantitation thresholds in relation to the PowerPlex 21 validation data and any other internal research performed. Ideally the QHFSS PowerPlex 21 method requires optimisation to reduce stochastic effects, and a new quantitation threshold set. Evaluation is needed of critical crime scene samples that were previously not fully tested to determine if they should be further analysed.

Recommendation 5: All DNA mixtures reported by QHFSS for the BLACKBURN matter require re-analysis by an independent expert after the user-defined settings are corrected.

Recommendation 6: All DNA mixtures relating to other matters require evaluation and where required, re-analysis after the user-defined settings are corrected.

Recommendation 7: An independent investigation is needed of the initial PowerPlex 21 and STRMix internal validation studies and the amended PowerPlex 21 internal validation to determine if any other issues and errors exist. Potentially re-validation is required. A course of action is required to re-analyse affected samples.

Recommendation 8: The competency of QHFSS scientists to correctly interpret DNA mixtures needs to be independently assessed, and where needed, further training provided. QHFSS standard operational procedures need to be independently reviewed.

Recommendation 9: [redacted] and [redacted]'s DNA profiles need to be compared against the BLACKBURN crime scene profiles by an independent expert.

Recommendation 10: The failure to compare suspects' DNA to crime scene evidence should be examined by an independent expert to confirm if it is a systemic error either in QPS or QHFSS processes. If so, an investigation is needed to determine how many other cases are impacted. Where required, further DNA comparisons of crime scene DNA samples against suspect samples should be conducted by an independent expert to ensure no offenders have been falsely excluded.

Recommendation 11: The national forensic science quality framework requires additional measures to ensure the accuracy and reliability of forensic DNA analysis.

Recommendation 12: A set of QHFSS data that quantitates DNA profiling success should be collected and reported to QPS and key justice stakeholders annually.