Queensland Health

HealthSupport Queensland

Project Report #186 - Assessment of 3500*xl* A Genetic Analyzer for Processing Casework Powerplex® Samples

Adrian Pippia, Emma Caunt, Rhys Parry, Sharon Johnstone, Justin Howes and Cathie Allen

September 2019
Version 2.0

Queensland Government

Document Details

Contact for enquiries and proposed changes

If you have any questions regarding this document or if you have a suggestion for improvements, please contact:

Contact officer: Adrian Pippia
Title: Reporting Scientist – FRIT
Phone: +
Email: A

Version history

Version	Date	Changed by	Description
1.0	29/03/2018	Adrian Pippia	First issue
2.0	26/06/2019	Sharon Johnstone	Mgt Team Feedback and agreed approach

Document sign off

This document has been **approved** by:

Name	Position	Signature	Date
Cathie Allen	Managing Scientist		01/10/2019

The following officers have **endorsed** this document

Name	Position	Signature	Date
Justin Howes	Team Leader FRIT		20.09.2019
Paula Brisotto	Team Leader ER & Q		01-10-2019
Luke Ryan	Senior Scientist Analytical		20-09-2019
Allan McNevin	Senior Scientist ER		20.09.2019.
Kirsten Scott	Senior Scientist Q & P		30/09/2019
Allison Lloyd	A/Senior Scientist Intel		30/09/2019
Sharon Johnstone	Senior Scientist Reporting 1		20/09/2019
Kylie Rika	Senior Scientist Reporting 2		30/09/2019

Contents

Document Details ii

Contents iii

1 Background 4

2 Aims 5

3 Resources 6

4 Methods 6

4.1 Sample selection for baseline LOD & LOR for casework interpretation 6

4.2 Crime Scene (Casework) Samples 8

4.3 Sample selection for baseline LOD for extraction control interpretation 8

4.4 Profile Interpretation 1 9

4.5 Profile Interpretation 2 9

4.6 Statistical Assessment of Data 9

5 Results and Discussion 10

5.1 Assessment of Baseline, LOD and LOR for Casework samples 10

5.2 Casework sample Interpretation using LOD and LOR thresholds 11

5.3 Summary 16

5.4 Assessment of Baseline and LOD of Negative Controls 17

6 Conclusion 18

7 Recommendations 18

7.1 Recommendation 1 18

7.2 Recommendation 2 18

7.3 Recommendation 3 19

7.4 Recommendation 4 19

8 References 19

9 Appendix 21

9.1 Analysis of Baseline *xl* A 21

9.2 DNA profile interpretation spreadsheet – Sample Sets 25

9.3 Peak Height Comparison between 3130*xl* B and 3500*xl* A Genetic Analyzers 25

1 Background

This project is an accumulation of a number of different projects in an attempt to set conditions for the use of 3500*xl* for the analysis of casework samples amplified with PowerPlex™21 (PP21). The many attempts at this project have spanned over a large period of time and therefore the conditions with which samples were run have also varied. The intent of this project was to bring all previous work together to implement the use of the 3500*xl* with casework PP21 profiles.

A summary of the previous projects is as follows:

Project # 145 3500xl Validation – This project had separate experimental designs and final reports for Direct Amplification FTA, Extracted Reference and Casework 3500*xl* validations. At the completion of this project the 3500*xl* was successfully validated for the analysis of Direct Amplification FTA and Extracted Reference samples.
For Crime Scene samples, the review of the mixed DNA profiles noted excessive pull-up. The effect of this was that it was not possible to reliably determine the number of contributors to the mixtures analysed in this experiment and the level of pull-up would inhibit the ability of the scientist to distinguish between allelic and pull-up peaks. The pull-up peaks also caused excessive raised baseline which has the potential to mask allelic peaks. Spectral calibration was unable to reduce the pull-up observed enough to improve the ability to determine the number of contributors reliably. Given these findings it was decided that the 3500 *xl* was not suitable for implementation for the analysis of casework samples amplified with PowerPlex™21 at that time.

Project #177 3500 CW WEN - This project was intended to assess whether the updated PowerPlex®21 System and PowerPlex®5-Dye Matrix standards had improved the spectral separation issues seen in the initial 3500 *xl* Casework validation (Project# 145). The Internal Lane Standard (ILS) and the PowerPlex5-Dye Matrix standards were modified to include an updated dye (WEN) to replace the CC5 dye. The WEN dye has stronger emission and improved photostability when compared to CC5 and therefore it was thought that the improvement in the dye could extend to the improvement of the overall spectral separation.

The conditions of Project #177 were further changed due to simultaneous work being conducted for the implementation of STRmix 2.3.6 (Project # 170). Due to the change in the modelling of stutter in STRmix version 2.3.6, it was decided that work was required to reassess the stutter thresholds used and reassess the data used to formulate the stutter thresholds. It was proposed that some of the stutter thresholds were elevated due to the use of samples amplified using too much template.

Further work including varying the amount of WEN dye (0.5µL, 0.3 µL & 0.15 µL), 0.5uL being the preferred, reduction of the amount of template to produce the amplified product, quantification of all of the samples involved in triplicate, change to

the selection of samples and strategies to minimise the effect of pipetting variation were all attempted to assist in the reduction of pull-up present.

Ultimately, the amount of pull-up observed, unexpected AI, unexplainable peak shoulders, along with allelic dropout as large as 500RFU, caused the conclusion that the change in dye and other conditions had not improved the consistency of the DNA profiling and that due to the difficulty in interpretation routine implementation was not recommended.

Project # 186 Verfication of STRmix v2.0.6 for use with the 3500 - Concern was raised with the baseline assessment that had been previously run and a new baseline dataset was proposed. In order to mimic Crime Scene DNA profile conditions, the samples chosen included a range of DNA template concentrations from 0.050ng to 0.900ng with 0.500ng being the central value. These samples with a higher DNA template were chosen due to the observed increase in the artefacts, pull-up and noise, proportional to increasing amounts of input DNA.

The samples chosen were mixtures to be amplified five times. Mixtures were chosen due to the increased effect on baseline with the larger number of allelic peaks, along with the fact that more interpretation complications are observed when interpreting mixtures in comparison to interpreting single source DNA profiles. A total of 45 samples were used. Excess profiles and samples displaying preferentially amped patterns were removed from the data.

Previously run extracted negative controls and negative amplification controls (25 of each) were used to calculate the LOD for negative controls and the drop-in rate. Data from the Negative controls was compared to the data from the crime scene samples to determine that separate LOD's should be used specific to the sample type.

2 Aims

The purpose of this assessment is to:

- Determine appropriate analytical and reporting thresholds (LOD and LOR) for casework samples using the 3500 *xl* A Genetic Analyzer
- Apply the LOD and LOR along with -2rpt, -1rpt and +1rpt stutter thresholds, determined from '*Project #170 – Reassessment of in-house stutter thresholds and stutter files used in STRmix™' [9]*, to selected crime scene/ casework DNA profiles to assess allele detection capability and ease of profile interpretation
- Determine appropriate analytical and reporting thresholds (LOD and LOR) for Negative Extraction controls processed using the 3500*xl* A Genetic Analyzer
- Re-evaluation of the work completed thus far and a gap analysis of the remaining requirements necessary for routine implementation

3 Resources

Refer to Project Proposal #186 for the reagents, materials and equipment used.

4 Methods

4.1 Sample selection for baseline LOD & LOR for casework interpretation

4.1.1 Sample set 1

This sample set consisted of constructed mixtures prepared from blood samples from CTS cases. These CTS samples had been previously used in projects 141, 145 & 149.

Table 1 Sample set 1

Sample barcode	Sample ID	# Contribs	Mixt ratio	Total target input DNA (ng)	Quant (ng/µL) (Dec2016)
[redacted]	1	3	1:1:1	0.900	0.07920
	2	2	30:1	0.800	0.05405
	3	2	50:1	0.700	0.07065
	4	2	20:1	0.600	0.05525
	5	3	5:2:1	0.500	0.05570
	6	3	30:1:1	0.400	0.05120
	7	3	30:1:1	0.300	0.06265
	8	3	30:1:1	0.200	0.06760
	9	2	20:1	0.100	0.03250
	10	2	5:1	0.050	0.03645

These samples were used for the evaluation of Baseline, LOD and LOR for casework samples and ranged in total DNA template concentration from 0.050 ng to 0.900 ng. Each sample was amplified a total of 5 times, bringing the total number of individual samples to 50.

4.1.2 Sample set 2

The samples listed in Table 2 consisted of single source and mixed DNA samples sourced/constructed from Promega positive controls and NIST standards (2373 NIST A and 2391c NIST B)
The Promega positive control was used to create single source DNA profiles with DNA templates of 0.500 ng, 0.338 ng, 0.182 ng and 0.065 ng.

The mixed DNA samples were created using NIST standards and the Promega positive control and amplified in triplicate to target DNA templates 0.7 ng, 0.5 ng and 0.2 ng.

Table 2 Sample set 2

Sample Barcode	# Contributors	Mixture Ratio	Actual template (ng)
[redacted]	1	N/A	0.500
			0.338
			0.182
			0.065
[redacted]	2	1:1	0.690
			0.506
			0.198
[redacted]	2	5:1	0.690
			0.506
			0.198
[redacted]	3	5:2:1	0.638
			0.510
			0.208
[redacted]	3	30:1:1	0.683
			0.501
			0.200

These samples were used to assess the baseline, LOD and LOR thresholds calculated for casework sample interpretation.

4.1.3 Sample Set 3

One single source and 13 mixed DNA profiles were selected as detailed in Table 3.

Table 3 Sample set 3

Sample Id	Ratio	# amps	Known # Contribs
[redacted]	N/A (single source)	1	1
[redacted]	1:1	3	2

Sample Id	Ratio	# amps	Known # Contribs
[redacted]	5:1	3	2
	20:1	3	2
	20:1	3	2
	30:1	3	2
	50:1	3	2
	1:1:1	3	3
	5:2:1	3	3
	5:2:1	2	3
	30:1:1	3	3
	30:1:1	3	3
	30:1:1	3	3
	30:1:1	3	3

Note: * Denotes samples common to sample set 1, these samples were created for the assessment of the baseline, LOD and LOR thresholds calculated for casework sample interpretation.

4.2 Crime Scene (Casework) Samples

Casework samples were used to assess the LOD and LOR and stutter thresholds calculated from experiment 5.1 and 'Project # 170 - Reassessment of in-house stutter thresholds and stutter file used in STRmix™ [9], respectively.

These samples were selected from plate CCE20180206-04 (renamed VCE20180312-02_reprep) which consisted predominantly of Priority 3 (P3) casework samples.

4.3 Sample selection for baseline LOD for extraction control interpretation

25 Negative Extraction controls and 25 Negative Amplification controls run between July 2016 and July 2017, were selected randomly and dispersed amongst the mixture samples.

DNA Quantification and Amplification were performed according to the Standard Operating Procedures (SOPs) 33407v4 & 34052v2. DNA fragment analysis was performed using the conditions determined in Project #177 using 0.5 µL of the WEN ILS. Analysis was performed as per QIS 34131v1.

4.4 Profile Interpretation 1

All DNA profiles (for case work interpretation and negative control interpretation) were analysed with GeneMapper®ID-X v1.4 using the analysis panel outlined by Promega. Additional conditions were as follows:

- Samples analysed at 1 RFU
- All known alleles, -1, -2 and +1 repeat stutter of known alleles, known artefacts and spectral pull-up were removed using current thresholds. As defined by Promega artefact peaks in the N-2bp and/or N+2bp position at D1S1656, D6S1043, D13S317, vWA, D21S11, D7S820, D5S818, D12S391 and D19S433 loci and in the N-1bp position at Amelogenin were also removed. Any other artefacts observed by Forensic DNA Analysis were also excluded from baseline calculations.
- Any peaks determined to be carry over, cross talk or signal interference were also removed. Carry-over is defined as the physical transfer of DNA from one injection to the next. Cross-talk is defined as transfer of fluorescence between adjacent capillaries. Signal interference is defined as transfer of fluorescence from one capillary to any other capillary during an injection.

4.5 Profile Interpretation 2

DNA profiles for casework interpretation were analysed with GeneMapper®ID-X v1.4 using the analysis panel outlined by Promega. Additional conditions differing from Profile interpretation 1 were as follows:

- All known alleles, -1rpt and +1rpt stutter peaks were left on

4.6 Statistical Assessment of Data

4.6.1 Statistical Package

R statistical software ((3.3.2 (2016-10-31) – "Sincere Pumpkin Patch")) was employed to analyse data by way of 'Analysis of Variants' (ANOVA) procedure to calculate the following:

- The average peak height (µPk), Relative Fluorescent Unit (RFU)
- The standard deviation (σPK)
- Significant difference, as explained in each of the Experimental Designs, 5.1 and 5.2

A scientist experienced in the use of R statistical software completed the data analysis.

4.6.2 Manual Calculations

The Limit of Detection was calculated using Equation 1:

Equation 1
$LOD = \mu Pk + 3\sigma PK$

The Limit of Reporting was calculated using Equation 2:

Equation 2
$LOR = \mu Pk + 10\sigma PK$

5 Results and Discussion

5.1 Assessment of Baseline, LOD and LOR for Casework samples

The baseline calculations were performed on the data from 60-500bp (being the largest read region of the individual dye lanes (orange). It was noted however that for other dye lanes, specifically blue and yellow, this read region included non-specific peaks that influenced the final baseline figures (Refer to Table 4). Therefore, a more specific read region for each dye was chosen as detailed in Table 5.

Table 4 LOD and LOR for casework samples

	Blue (RFU)		Green (RFU)		Yellow (RFU)		Red (RFU)		Orange (RFU)	
	LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR
60-500 bp	26.1	76.2	22.9	57.8	49.9	135.4	33.0	79.8	13.5	34.7
Specific Read region	11.9	30.0	22.7	57.3	44.2	116.6	34.7	84.7	13.5	34.7

Table 5 Dye specific read regions

Dye lane	Read region (bp)
Blue	76-470
Green	70-467
Yellow	63-450
Red	65-430
Orange	60-500

The LOD and LOR from the specific read regions generated very different results. It was decided that colour specific LOD's and LOR's would be used in the subsequent part of the project; assessing the ease of profile interpretation. Table 6 has the final dye specific LOD and LOR thresholds used, rounded up to the nearest 5 RFU.

Table 6 Thresholds used for crime scene sample assessment

Blue (RFU)		Green (RFU)		Yellow (RFU)		Red (RFU)		Orange (RFU)	
LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR
15.0	30.0	25.0	60.0	45.0	120.0	35.0	90.0	15.0	35.0

5.2 Casework sample Interpretation using LOD and LOR thresholds

DNA profiles were assessed qualitatively by an experienced case manager to determine whether the DNA profiles were able to be interpreted using the specific dye lane LODs and LORs calculated from experiment 5.1

The following criteria were used:

- Assessment of number of contributors
- Were the DNA profiles concordant
- Were any additional alleles present
- Was pull-up present and what effect it had on the interpretation
- Whether the LOR for each dye lane was suitable given the level of the baseline with the profile

5.2.1 Sample set 1

The information obtained from the assessment of these 50 mixtures was deemed inappropriate given the mixed DNA samples selected were the same samples used to calculate the Baseline, LOD and LOR thresholds.

5.2.2 Sample set 2

5.2.2.1 Number of contributors

Assessing the number of contributors of the individual DNA profiles was conducted by:

- Counting the maximum number of alleles at any one locus, taking into consideration the presence of stutter (-2rpt, -1rpt and +1rpt)
- Assessing allelic balance between the potential genotype combinations of the individual contributions
- Consideration of the consistency of the ratios between the contributions present across the entire DNA profile

To evaluate the influence of stutter (-2rpt, -1rpt and +1rpt) on estimating number of contributors, two approaches were considered:

Approach One – Any instance of high stutter that did not coincide with an expected allele would increase the number of contributors by one.

Approach Two - The presence of high stutter peaks alone would not increase the number of contributors established. An increase in the number of contributors would rely on the presence of an allelic peak outside stutter positions and not associated to the intuitive allelic combinations.

Stutter thresholds used for -1rpt were allele specific stutter thresholds.

5.2.2.2 Number of contributors – Approach One

The number of contributors was assessed incorrectly for 14 out of the 16 DNA profiles which equates to 87.5% of the sample set. For all but one of the DNA profiles, the number of contributors was higher than expected due to -1rpt or +1rpt stutters, being above the allele specific thresholds recommended. The remaining DNA profile was assessed as having a higher number of contributors due to the observation of an extra peak outside of the expected alleles (129549144, 0.7 ng = '22' peak (34RFU – LOD=30) at D6).

5.2.2.3 Number of contributors – Approach Two

The number of contributors was assessed incorrectly for 1 out of the 16 DNA profiles due to the observation of an extra peak outside of the expected alleles (129549144, 0.7 ng = '22' peak (34RFU) at D6), the same as for the first approach.

5.2.2.4 Concordance

With the exception of the above mentioned extra (drop-in) peak, all samples were concordant with the expected results. There were nine occasions of allelic drop out, two of which were visible above LOD but below LOR.

5.2.2.5 Pull-up

Peak heights were observed to range from 200 RFU to 5800 RFU for DNA profiles with a DNA template of approximately 0.2 ng, 170 RFU to 17000 RFU for DNA profiles with a DNA template of approximately 0.5 ng and 100 RFU through to 15000 RFU for DNA profiles with a DNA template of 0.7 ng. Pull-up was present in all profiles and was proportional to the peak heights. The presence of pull-up peaks increased as the DNA template input and in turn peak heights increased. Pull-up peaks with heights above the LOR were able to be distinguished from true allelic peaks and did not adversely affect profile interpretation. Interpretation was more difficult in the shorter fragment length STR's where pull-up was more prevalent, requiring the use of GeneMapper®ID-X v1.4 for interpretation post reading. See Figure 1 for an example of the raised baseline in the green and yellow dyes that caused interpretation difficulty in the smaller fragments.

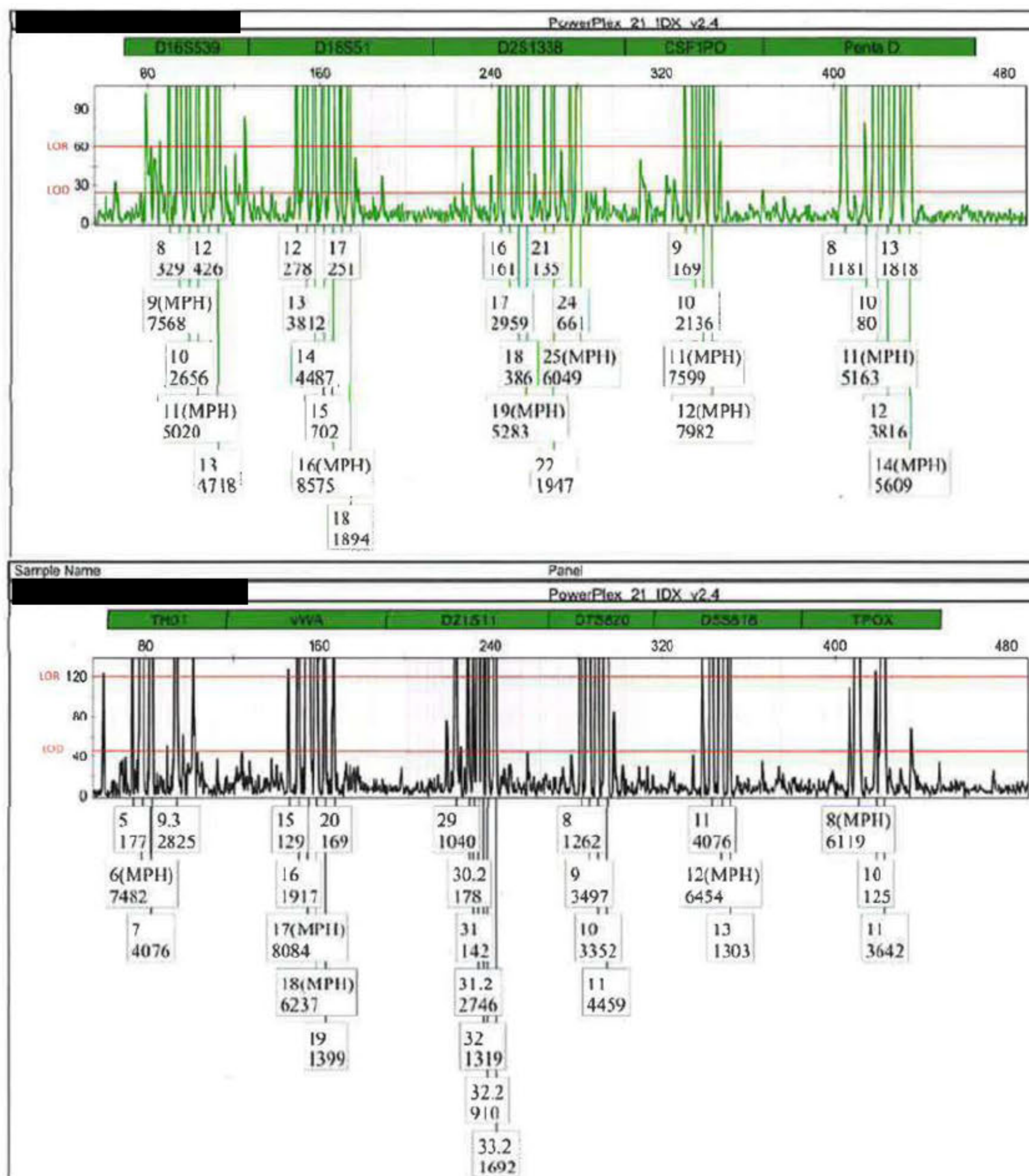

Figure 1 Raised baseline and pullup peaks in green and yellow dye lane

5.2.2.6 Dye specific LOR Suitability

LOR calculated for each dye lane was above the baseline and considered suitable for profile interpretation purposes. The pull-up peaks above the LOR were distinguishable from true peaks based on their irregular peak morphology or positioning in the allelic bin and were removed at the GeneMapper®ID-X v1.4 phase of DNA analysis.

5.2.3 Sample set 3

5.2.3.1 Number of contributors

The same method as that described for 'Sample Set 2', was used to assess the number of contributors to these samples with the following differences:

- The use of 'reproducibility' calculation [12,13] was assessed for each DNA profile for those samples where three amplifications were available, and reproducibility was an appropriate method to use.

The correct number of contributors was assigned to eight of the thirteen mixed DNA profiles. Where the assessment of the number of contributors was incorrect, one additional contributor was assigned due to the presence of high stutters (-2rpt and -1rpt).

One sample (579901284) had an elevated number of contributors due to the addition of a contributor due to poor reproducibility. In this example the mixture ratio was 50:1 and therefore the inconsistent minor component is not unexpected. The reproducibility guidelines may not be an appropriate method for assessing the number of contributors to a mixture generated using the 3500 *xl* A Genetic Analyzer and therefore they will need to be reassessed.

5.2.3.2 Concordance

Of the 39 DNA profiles (individual amplifications) interpreted, 8 profiles were observed to have an extra allele outside the expected DNA profile. Four of these were in the form of elevated stutter, 2 of which could be elevated due to pull-up. Drop-in peaks:

- Sample [redacted], F01 = allele '8' at D6 (44 RFU)
- Sample [redacted], G01= allele '16' at D1 (43 RFU); coincided with stutter position (24%)
- Sample [redacted], F02 = allele '12' at D16 (63 RFU); coincided with stutter position (23%)
- Sample [redacted], G02 = allele '22' at D2 (80 RFU)
- Sample [redacted], C01 = allele '9' at D7 (64 RFU); coincided with stutter position (19.6%)
- Sample [redacted], E01 = allele '14" at CSF (175 RFU)
- Sample [redacted], F05 = allele '20' at D12 (155 RFU); coincided with stutter position (19.2%)
- Sample [redacted], D03 = allele '10' at D6 (41 RFU)
- Sample [redacted], C04 = allele '17' at D2 (109 RFU); coincided with stutter position (24%)

There were 23 occasions of allelic drop out, which were visible above LOD but below LOR.

5.2.3.3 Pull-up

Peak heights were observed to range from 100 RFU to 20000 RFU. The effect of Pull-up peaks on interpretation was the same as seen for Sample set 2, with there being some difficulty in the interpretation of the smaller STR's due to an elevated baseline, see example below.

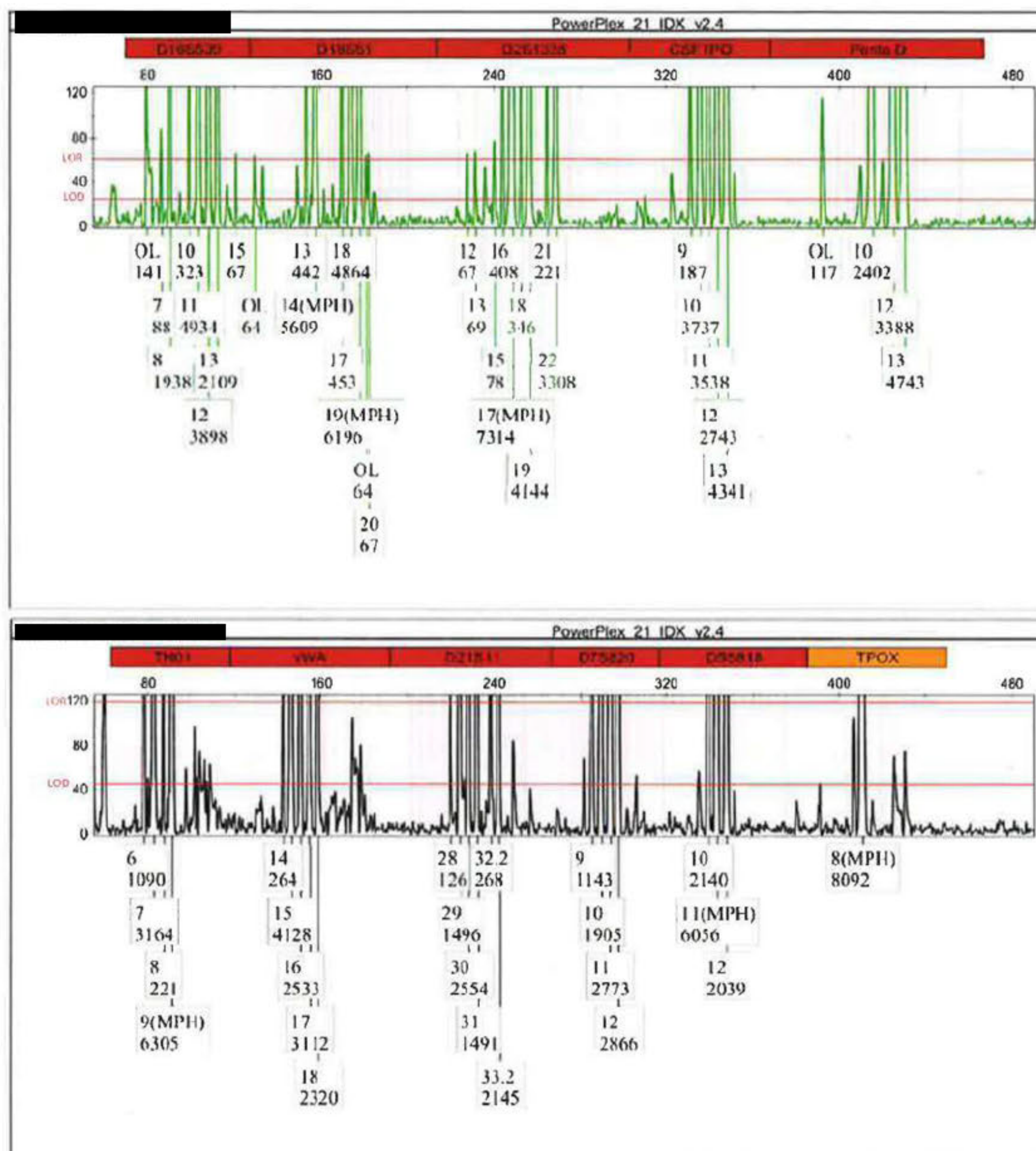

Figure 2 Raised baseline and pullup peaks in green and yellow dye lanes

5.2.3.4 Dye specific LOR Suitability

LOR calculated for each dye lane was above the baseline and considered suitable for profile interpretation purposes. The pull-up peaks above the LOR were distinguishable from true peaks based on their irregular peak morphology or positioning in the allelic bin and were removed at the GeneMapper®ID-X v1.4 phase of DNA analysis.

5.2.4 Crime Scene Sample set

The assessment of these samples could not be as rigorous given that their source of the DNA is not known. Eleven volume crime samples were re-run on the 3500*xl* A

Genetic Analyser, nine of which were amplified at optimal total DNA input of 0.5ng and the other two samples were 0.18 ng and 0.255 ng of total DNA input.

The number of contributors was able to be assessed for all samples, however high stutter peaks were observed. There was difficulty encountered in the interpretation of some profiles due to the observation of some extremely high peak heights that exceeded 30000 RFU. These peaks heights presented problems as these large peaks had an effect of raising the baseline in the low molecular weight STR's, as well as exceeding the maximum height that could be modelled by STRmix. In particular, artefacts at TH01 made interpretation difficult as demonstrated in the below example (Figure 3).

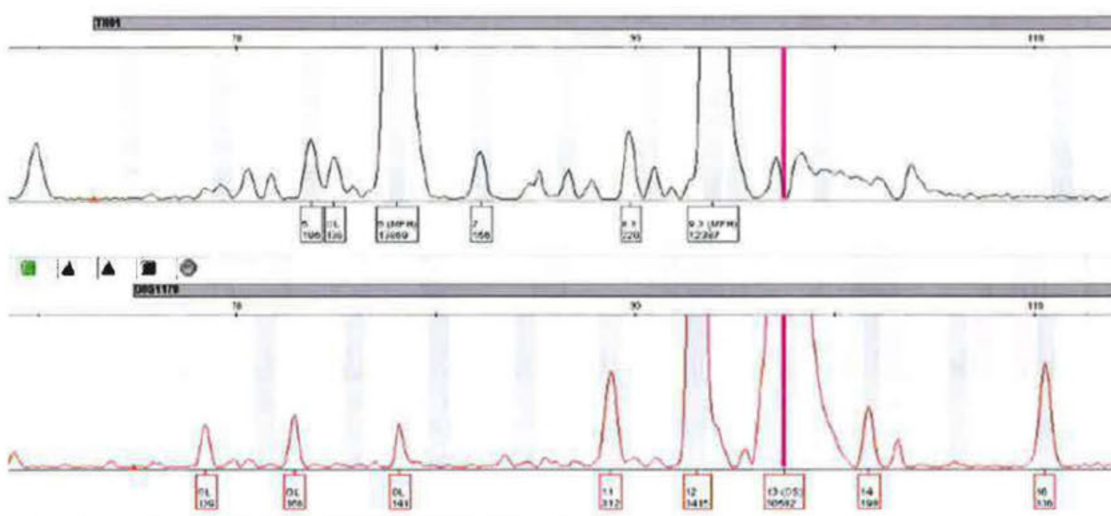

Figure 3 TH01 artefacts

5.3 Summary

The dye specific LOD and LOR thresholds calculated for use with the 3500*xl* A Genetic Analyzer coupled with the PowerPlex®21 amplification system are suitable for application of DNA profile interpretation. Instances of raised baseline were noted by all profile assessors and coincided with high peak heights of the profiles. Raised baseline and high pull-up peaks as a result of the high peak heights, created issues when interpreting the DNA profiles, specifically the inability to decide whether a peak was a true allelic peak or artefact. Although this alone did not lead to the incorrect number of contributors being assessed, the ambiguity surrounding potential allelic peaks associated to regions of raised baseline/ pull-up in the small molecular weight STRs proved troublesome. Therefore the interpretation of mixed crime scene samples and ability to achieve consistency in interpretation remains difficult to obtain.

Sub threshold peaks were observed in all DNA profiles assessed. In total 23 sub thresholds peaks were observed, the majority within Sample Set 3, and in all instances were used to inform the number of contributors of the profile. An assessment of the way sub threshold information can be used in determining the

number of contributors needs to be further assessed if the 3500*xl* A Genetic Analyzer is implemented for routine work.

Similarly, stutter peaks were observed in the profiles of all sample sets and in some instances, these peaks were above the corresponding stutter thresholds, resulting in an incorrect assessment of the number of contributors. A review of the way stutter information can be used in determining the number of contributors needs to be assessed if the 3500 *xl* A Genetic Analyzer is implemented for routine work.

5.4 Assessment of Baseline and LOD of Negative Controls

Fifty negative controls (25 extraction negatives and 25 amplification negatives) were used to assess the baseline and calculate the LOD and LOR for samples without DNA.

As for the assessment of LOD and LOR in casework samples, a more specific read region for each dye was chosen as detailed in Table 6.

Table 7 LOD and LOR for negative controls

	Blue (RFU)		Green (RFU)		Yellow (RFU)		Red (RFU)		Orange (RFU)	
	LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR	LOD	LOR
60-500 bp	23.5	69.5	19.3	47.8	36.0	95.6	23.8	52.6	9.3	22.3
Specific read region	8.8	21.3	17.8	43.2	21.8	49.5	23.8	52.4	9.3	22.3

Figure 4 (below) and Table 7 (above) show the difference in the LOD and LOR between samples that contain DNA and negative controls. This data supports the implementation of different thresholds for casework samples and negative controls. There is a risk of not identifying potential contaminating allelic peaks in negative controls by using the higher thresholds for casework LOD and LOR.

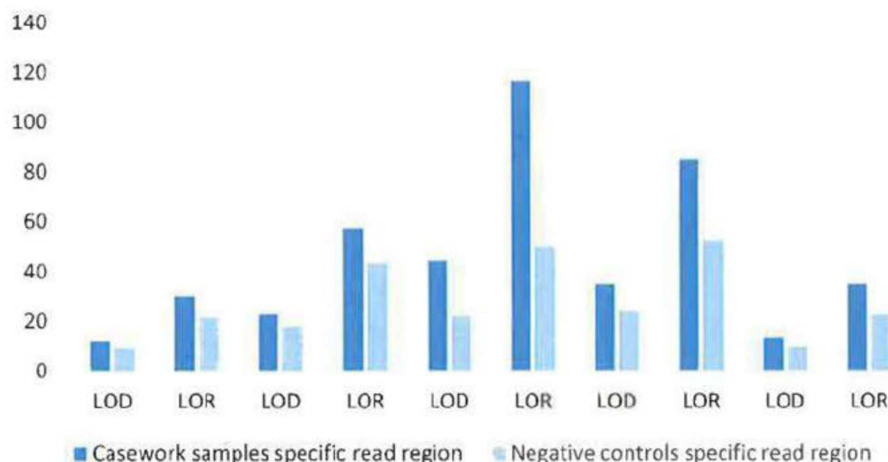

Figure 4 LOD and LOR comparison between casework samples and negative controls

6 Conclusion

At completion of this project it was determined that the implementation of the 3500*xl* A Genetic Analyzer coupled with the PowerPlex®21 amplification kit was not preferred for routine case work. The Forensic DNA Analysis laboratory has been informed that the use of the 3130*xl* Genetic Analyzer will not be supported by the supplier Life Technologies Applied Biosystems from 2020. In the absence of another validated amplification kit implemented for routine case sample use, the 3500*xl* A Genetic Analyzer coupled with PowerPlex®21 amplification could be implemented provided supporting documentation in the appropriate interpretation, including determining the number of contributors and relevant reworking strategies. Further data analysis using STRmix v2.6.2 may be required to enable such documentation.

The high peak heights noted during profile interpretation resulted in elevated baseline and pull-up which created ambiguity in the assessment of true allelic peaks versus peak like artefacts. These interpretation difficulties, although not disproportionate to the higher peak heights observed with the 3500*xl* instrument in comparison to the 3130*xl*, would still require an interpretation and rework strategy to mediate peak heights and assist with consistent interpretation.

Creation and use of dye lane specific LODs and LORs for assessment of negative controls and casework samples were considered more suitable than using all information within the 60-500 bp read region. Despite the calculated LOD and LOR thresholds appearing suitable the elevated baseline and pull-up proved a hinderance during DNA profile interpretation.

Use of stutter and sub threshold information to determine number of contributors also needs to be assessed as part of an implementation plan.

7 Recommendations

7.1 Recommendation 1

It is recommended that the 3500 *xl* A Genetic Analyzer should not be implemented for routine casework sample processing at this time due to the interpretation difficulties associated to high peak heights and resultant elevated baseline, artefacts and pull-up particularly in the low molecular weight regions.

7.2 Recommendation 2

It is recommended that, if for business continuity reasons, the 3500 *xl* A Genetic Analyzer is required to be implemented in combination with the PowerPlex®21 amplification system for routine case work, that mixture samples be re-assessed using the loci specific stutter thresholds developed in project #170 in conjunction with the use of the current STRmix version.

7.3 Recommendation 3

It is recommended that the use of stutter and sub threshold information in determining number of contributors needs to be assessed as part of supporting documentation for the implementation plan.

7.4 Recommendation 4

It is recommended that an interpretation and rework strategy be documented to assist with consistent interpretation.

8 References

[1] Coble, Michael D; Validation and Thresholds – DNA Interpretation Workshop 1; U.S. National Institute of Standards and Technology (NIST); ISFG Pre-Conference Workshop, Melbourne, Australia September 2-3, 2013

[2] Gilder et al. Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing. J Forensic Sci, January 2007, Vol. 52, No. 1, 97:101

[3] Applied Biosystems Invitrogen; Validation Summary of Methods for Use of Identifiler with the ABI Prism 310 Genetic Analyzer; section 8.1.3.1(b)

[4] Butts et al. NIST validation studies on the 3500 Genetic Analyzer. Forensic Science International: Genetics Supplement Series 3 (2011) e184-e185

[5] Personal Communication - from Dr Duncan Taylor; Forensic Science South Australia; 25 January, 2017

[6] Bregu et al. Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis. J Forensic Sci, January 2013, Vol. 58, No. 1

[7] Acedo, P. Caunt, E. Morgan, R. McNevin, A. Ryan, L. Allen, C. Powerplex®21 *Optimisation Program Phase 1 Part b, 2014*

[8] Mathieson, M. Ryan, L. Micic, B. Allen, C. Verification of 3500*xl* B, 2016

[9] Caunt, E. Nurthen, T. Pippia, A. Howes, J. Allen, C. Project #170 Reassessment of in-house stutter thresholds and stutter files used in STRmix™, 2018

[10] SWGDAM Guidelines for Validation of Probablistic Genotyping Systems, section 4.1.12

[11] ESR, *STRmix™ V2.5 Operation Manual*. 2017, Institute of Environmental Science and Research Limited.

[12] Caunt, E. Morgan, R. Howes, J. Allen, C. Development of guidelines for the determination of number of contributors to a Powerplex®21 profile – Version 2, 2015

[13] E. Caunt, R. Morgan, Assessment of the number of contributors for mixed PowerPlex 21 DNA profiles within Forensic DNA Analysis, Feb 2015

[14] Applied Biosystems. User bulletin applied biosystems® 3500/3500*xl* genetic analyzer. Foster City, CA; 2011

[15] Bright, J et al. Variability of mixed DNA profiles separated on a 3130 and 3500 capillary electrophoresis instrument. Australian Journal of Forensic Sciences, 2014, 46:3, 304-312

[16] Promega, *PowerPlex 21 System Technical Manual*. 2017

9 Appendix

9.1 Analysis of Baseline 3500*xl* A

Baseline data was supplied as a GMIDx output from 100 baseline samples read at 1 rfu. This data was converted to an Excel spreadsheet and saved within the project folder structure at I:\Change Management\Proposal#186 - Assessment of 3500*xl* Analysis of Casework Powerplex21 samples\Baseline\3500 baseline project186_0.5WEN\.

The Excel spreadsheet was converted to a long format such that the dye colours formed a single variable and the peak heights formed a single variable. This file was saved as I:\Change Management\Proposal#186 - Assessment of 3500*xl* Analysis of Casework Powerplex21 samples\Baseline\3500 baseline project186_0.5WEN\Significant Difference Testing – RJP\ Stripped_Data_Baseline_3500_2017.xls.

The data was imported into R-studio version 1.0.136, running R version 3.3.2 (2016-10-31 "Sincere Pumpkin Patch"; Copyright (C) 2016 The R Foundation for Statistical Computing), using standard methods.

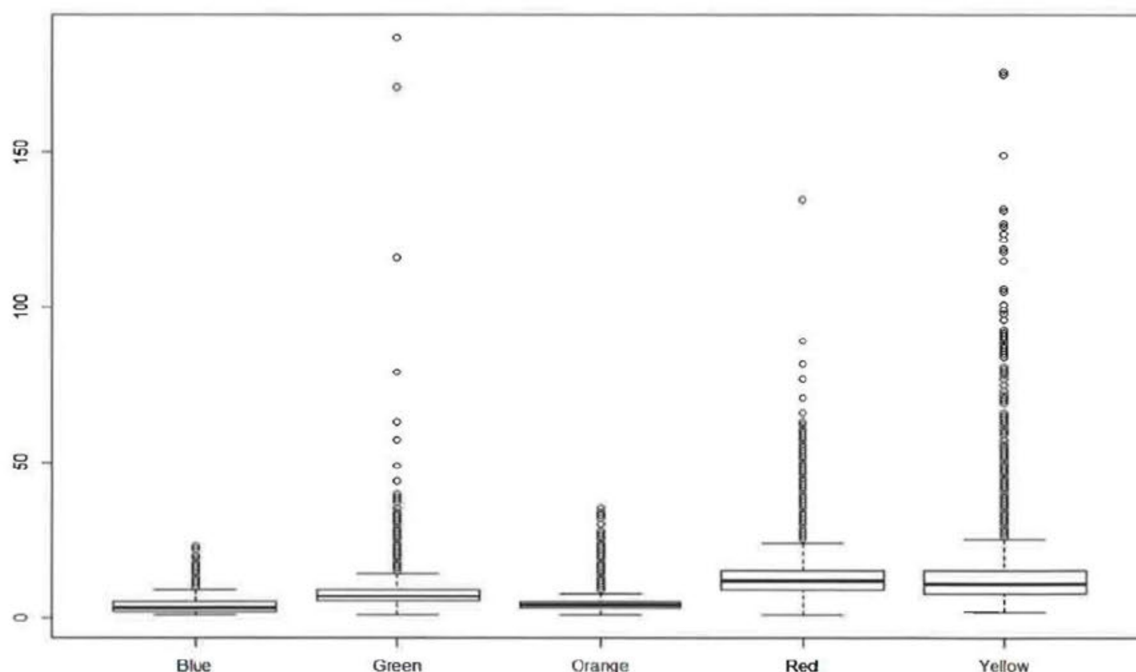

Figure 5 Boxplot of the distribution baseline peak heights of the five dye colours

Figure 5 shows that the distribution of the blue and orange dye peaks is similar, as is the distribution of the red and yellow dye peaks. The main difference between these pairings is the spread and quantity of the outlier peak heights. The median values (indicated by the solid black bar) are reasonably evenly distributed between the 1st

and 3rd quartiles indicating that the vast bulk of the data is concentrated in a small band, but the skew values in table 1 suggest that the outliers contribute to heavy right tails (see Figure 6 below), particularly in the green, red and yellow dyes. It should be noted that the maximum outliers are over 120rfu indicating an abundance of theoretically rare events or the more-likely possibility that these outliers are influenced by unseen drop-in peaks.

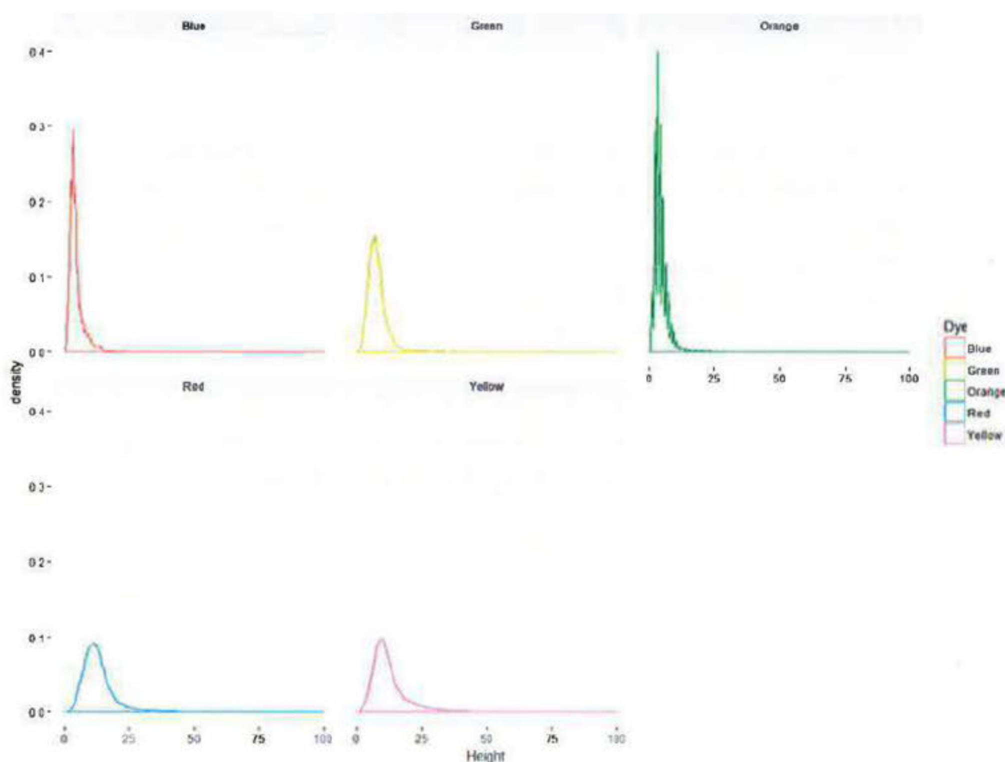

Figure 6 Density plots of the distribution baseline peak heights of the five dye colours

Table 8 below outlines the summary statistics for baseline peak heights in the different dye groups. The mean baseline height of the Blue Dye is 4.1rfu with a standard deviation of 2.6. The Green Dye has a mean of 7.9rfu (sd=4.94); the Orange Dye has a mean of 4.39rfu (sd=2.99); the Red Dye has a mean of 13.3rfu (sd=7.14); and the Yellow Dye has a mean of 13.2rfu (sd=10.3). It should be noted that the number of samples in each group exceeds 7000 (n-value Table1).

Table 8 Edited R Output of summary statistics for baseline comparison

```
Descriptive statistics by group
Dye: Blue
       vars    n  mean   sd median trimmed  mad min max range skew kurtosis   se
Height    2 7147   4.1  2.6      3    3.65 1.48   1  23    22 2.04      5.8 0.03
---------------------------------------------------------------------------------
Dye: Green
       vars    n  mean   sd median trimmed  mad min max range skew kurtosis   se
Height    2 8187   7.9 4.94      7    7.36 2.97   1 187   186 13.4      397 0.05
---------------------------------------------------------------------------------
Dye: Orange
       vars     n  mean   sd median trimmed  mad min max range skew kurtosis   se
Height    2 10726 4.39 2.99      4    3.92 1.48   1  35    34  3.1     15.5 0.03
---------------------------------------------------------------------------------
Dye: Red
       vars    n  mean   sd median trimmed  mad min max range skew kurtosis   se
Height    2 9135  13.3 7.14     12    12.3 4.45   1 135   134 3.08     20.5 0.07
---------------------------------------------------------------------------------
Dye: Yellow
       vars    n  mean   sd median trimmed  mad min max range skew kurtosis   se
Height    2 8431  13.2 10.3     11    11.5 4.45   2 176   174 5.56     50.8 0.11
```

A linear model was developed to compare the means of the different dye colours and assess them for statistical differences.

$$E(y) = \beta_0 + \beta_1 \text{Green} + \beta_2 \text{Yellow} + \beta_4 \text{Red} + \beta_5 \text{Orange}$$

Where β_0 is the mean of the blue dye and $\beta_{1\text{-}5}$ are the differences between the mean of the blue dye and the mean of the other dyes of interest. Green = 1 if true, 0 otherwise; Yellow = 1 if true, 0 otherwise; Red = 1 if true, 0 otherwise; Orange = 1 if true, 0 otherwise. Comparisons are made against the Blue dye (β_0) singly. That is, when Green is compared to Blue, Green takes the value 1 and all other colours take the value 0.

Table 9 Edited R output of model coefficients.

```
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## DyeBlue      4.0958     0.0741    55.3   <2e-16
***
## DyeGreen     7.8959     0.0692   114.1   <2e-16
***
## DyeOrange    4.3868     0.0605    72.6   <2e-16
***
## DyeRed      13.2993     0.0655   203.0   <2e-16
***
## DyeYellow  13.1984     0.0682   193.6   <2e-16
***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
```

It can be seen from Table 9 that the mean baseline peak height of the Blue dye is highly significantly different from the mean peak height of all other dyes ($p \approx 0$). Releveling of the data to make comparisons between all dye colours (Table 12) shows that the mean baseline peak height is highly significantly different for all groups ($p \approx 0$). As such, it can be concluded that the baseline of each dye is representative of a distinct population of peaks and for this reason, it would be inadvisable to combine dye data.

Table 10 R Output of 95% confidence intervals for dye baseline mean.

```
##             Estimate Std. Error 2.5 % 97.5 %
## DyeBlue         4.10     0.0741  3.95   4.24
## DyeGreen        7.90     0.0692  7.76   8.03
## DyeOrange       4.39     0.0605  4.27   4.51
## DyeRed         13.30     0.0655 13.17  13.43
## DyeYellow      13.20     0.0682 13.06  13.33
```

The 95% confidence intervals for the mean baseline height for each dye are given in table 10. Therefore, there is 95% confidence that if the study was rerun on different samples, that the mean would fall within the ranges provided. As can be seen, the ranges are quite small and would not have a meaningful effect on the calculation of the LOD/LOR if the upper limit was used instead. As such, the mean was used and rounded to the nearest higher multiple of 5 rfu (Tables 11 and 12).

Table 11 Recommended LOD and LOR values for individual dyes (mean)

	Blue	Green	Yellow	Red	Orange
LOD	12	23	45	35	14
LOR	31	58	116	85	35

Table 12 Recommended LOD and LOR values for individual dyes (rounded to nearest greater 5rfu)

	Blue	Green	Yellow	Red	Orange
LOD	15	25	45	35	15
LOR	35	60	120	85	35

Table 13 Edited R output of inter-dye significance comparison

```
                            Estimate Std. Error t value Pr(>|t|)
relevel(Dye, "Green")Green    7.8959     0.0692   114.1   <2e-16 ***
relevel(Dye, "Green")Blue     4.0958     0.0741    55.3   <2e-16 ***
relevel(Dye, "Green")Orange   4.3868     0.0605    72.6   <2e-16 ***
relevel(Dye, "Green")Red     13.2993     0.0655   203.0   <2e-16 ***
relevel(Dye, "Green")Yellow  13.1984     0.0682   193.6   <2e-16 ***
                             Estimate Std. Error t value Pr(>|t|)
relevel(Dye, "Yellow")Yellow  13.1984     0.0682   193.6   <2e-16 ***
relevel(Dye, "Yellow")Blue     4.0958     0.0741    55.3   <2e-16 ***
relevel(Dye, "Yellow")Green    7.8959     0.0692   114.1   <2e-16 ***
relevel(Dye, "Yellow")Orange   4.3868     0.0605    72.6   <2e-16 ***
relevel(Dye, "Yellow")Red     13.2993     0.0655   203.0   <2e-16 ***
                             Estimate Std. Error t value Pr(>|t|)
relevel(Dye, "Orange")Orange   4.3868     0.0605    72.6   <2e-16 ***
relevel(Dye, "Orange")Blue     4.0958     0.0741    55.3   <2e-16 ***
relevel(Dye, "Orange")Green    7.8959     0.0692   114.1   <2e-16 ***
relevel(Dye, "Orange")Red     13.2993     0.0655   203.0   <2e-16 ***
relevel(Dye, "Orange")Yellow  13.1984     0.0682   193.6   <2e-16 ***
```

It could be argued that the skew discussed above undermines the theoretical assumptions of the linear model (ie. that the residuals are normally distributed around the expected response variable). However, an analysis was run to determine the most informative distribution for the data and this was found to be a log-normal distribution as it gave the lowest AIC value (Table 14).

Table 14 AIC values for various possible distributions representing the data

	Weibull	Gamma	Log-normal	Normal
AIC	267640	264100	261427	299177

Calculating the means for the baseline using a log-normal model and then transforming these back to real rfu values produced results very similar to those given in table 2 above (see Tables 15 & 16). The back-transformed means (Table 16) however, resulted in slightly lower LOR and LOD values than those presented above. As such, it is recommended that the original recommended values be kept.

Table 15 R output of log-normal means of baseline peak heights.

```
                  Coefficients:
          Estimate Std. Error t value Pr(>|t|)
DyeBlue    1.24844   0.00606     206   <2e-16 ***
DyeGreen   1.96350   0.00567     346   <2e-16 ***
DyeOrange  1.31172   0.00495     265   <2e-16 ***
DyeRed     2.47964   0.00536     462   <2e-16 ***
DyeYellow  2.42093   0.00558     434   <2e-16 ***
```

Note: To convert these to real peak heights the exponent of these values must be calculated.

Table 16 Back-transformed log-normal results for baseline means by dye.

```
DyeBlue  DyeGreen DyeOrange    DyeRed DyeYellow
 3.48      7.12      3.71      11.94     11.26
```

9.2 DNA profile interpretation spreadsheet – Sample Sets

The information from the interpretation of Sample Sets 2, 3 and "Crime Scene Samples" are located "I:\Change Management\Proposal#186 - Assessment of 3500xL Analysis of Casework Powerplex21 samples using 3500xL A\Results - Threshold Assessment" as the following excel spreadsheets:

- "Sample Assessment of Sample Set 2 using New Thresholds (Expected Vs Observed)_AAP"
- "Sample Assessment of Sample Set 3 using New Thresholds (Expected Vs Observed)_EJC"
- "Crime scene sample Assessment Using New Thresholds"

9.3 Peak Height Comparison between 3130xL B and 3500xL A Genetic Analyzers

Information from the peak height comparison of the 'Crime Scene Samples' is located "I:\Change Management\Proposal#186 - Assessment of 3500xL Analysis of Casework Powerplex21 samples using 3500xL A\Results - Threshold Assessment" as the following excel spreadsheet:

"Project 186 3130 Vs 3500 – Non Probative Samples"

FSS.0001.0006.4199