STATEMENT OF RHYS PARRY

I **Rhys Parry** of Queensland Health at the Forensic and Scientific Services, 39 Kessels Road, Coopers Plaines, do solemnly and sincerely declare that:

Background

1. I have a Bachelor of Science and a Bachelor of Science (Honours) degree from the University of Queensland. I have a Post-graduate Certificate in Data Science from the University of New England. I have worked at Queensland Health Forensic and Scientific Services DNA Analysis unit since March of 2006. I have worked as a reporting scientist since August 2008. Prior to my current role, I have worked as a research assistant on numerous scientific projects and have lectured in anatomy, physiology, and basic experimental design.
2. I am currently employed at the Forensic and Science Services (FSS) as a scientist in Reporting Team 1 of the DNA Analysis Unit.
3. The duties of my current role are to analyse and review DNA profiles, write and review Statement of Witness (SOW) documents, and give expert testimony pertaining to the results of DNA analyses for the Queensland Police Service (QPS) and Queensland Courts.
4. Experimental statistics was a significant part of my post-graduate honours degree research thesis and I have completed post-graduate subjects with experimental design components.
5. My current supervisor is Sharon Johnstone.

Concerns following the 6 June 2022 decision

6. After 6 June 2022 my concerns were:
    a. that the DNA Analysis Unit maintained the process of analytical staff reviewing 'no DNA detected' and 'DNA insufficient for further processing' results without the reporting scientists seeing them, and
    b. that profiles with quantification values between 0.0011ng/µL and 0.0088ng/µL are not being processed through microcon. This is especially problematic, in my opinion, in situations where spermatozoa are observed in sexual assaults.
7. Prior to the decision being communicated to us on 6 June, I had no knowledge or input into the decision that was made. I thought the decision to return to amplification only

(without microcon) was problematic. I was concerned that the change in process could result in significantly lower probability of obtaining optimal DNA profiles from samples in the 0.0011ng/µL and 0.0088ng/µL range.

Concerns following the 19 August decision

8. On 19 August 2022 a further decision was made to microcon all samples within the above range to 35µL. I do not recall being involved in any official discussions prior the implementation of this new process. I thought the change was a clear step in the right direction scientifically. I had mixed feelings about not having the option to microcon to full, but I also, at this time, suspected that the QPS might no longer trust our processes and that they had requested that the process be limited to "microcon to 35µL" so as they could, if necessary, get testing done elsewhere. I thought, if that were the case, the decision was not unreasonable given the laboratory's poor performance in this range recently.

Microcon

9. On, or around, July 2017, Justin Howes asked me to review the calculations in a spreadsheet he had provided to me. He stated he was data mining the results of historical microcon processes but provided no other detail. In reviewing the calculations, I suspected he was trying to examine the probability of obtaining a result from low concentration DNA samples. I produced a model in R (statistical software) for the data as the simple percentage method he was using was not suitable for this sort of data. Annexed and marked RP-01 is a copy of this model. However, I checked the spreadsheet he had provided me and found no errors in the formulae used. I visited him in his office, on or around the 5th of August (I know this as this is the day after the date recorded on the 'record of analysis' file from my analysis) and stated I had checked the spreadsheet and found no errors in the formulae used. I handed him two A4 sheets of paper, one with a plot of the success probabilities and one with a table of the probabilities at various concentrations. I stated that the these were the results I thought he might be trying to achieve and stated that as the distribution of results was not uniform, percentage calculations weren't ideal. He asked that I leave the results with him, and he would look at them later. I did not hear anything more on the matter until in January 2018, when Amanda Reeves and Kylie Rika approached me to give an opinion on the analysis contained in a draft version of the options paper.

I used the same plot of the success probabilities and the same table of the probabilities that I had provided to Justin in August, as part of my written response that was submitted as part of the response from Amanda and Kylie. Annexed and marked RP-02 and RP-03 are copies of the table, and the written response, that I provided to Amanda and Kylie.

10. My analysis indicated that the mean expected success probability for microcon success at 0.008ng/μL was ~0.223 (0.19 – 0.25 95% confidence). That is, on average, between 19% and 25% of micro-concentrated samples produced usable profiles at that concentration. At 0.009ng/μL, which is just above the range of interest, between 23% and 30% of micro-concentrated samples produced usable profiles.
11. I do not think that a hard cut-off for processing of DNA samples should be based on the quant obtained, given that it is well known that the quant is not particularly accurate at low levels.
12. Two years prior to the Options Paper, Kylie Rika, Josie Entwhistle, Allison Lloyd, and Thomas Nurthen looked at the success rates of microcon for profiles in the 0.00214ng/µL and 0.088ng/µL range (Project 163). This project found that approximately 18% of profiles were informative across the full range. It, however, also made the error of not correcting for a non-uniform distribution of results, so this percentage is not a good indicator of the true result. The paper did correctly use quant ranges to assess the percentage success rates. As such, they were able to identify that there were numerous informative results at almost all but the lowest quant band.
13. Prior to the 2018 Options Paper, samples were automatically processed through microcon if the quantification value was between 0.0011ng/µL and 0.0088ng/µL because the likelihood of getting a reasonable profile from those samples without micro-concentration was considered low.
14. After 2018, anything above 0.0088ng/µL went straight to amplification. There was still an option to be processed through microcon after the amplification if required. Only P1 samples were automatically sent to microcon if they had a quant below 0.0088ng/µL.
15. Since the 2018 process was introduced, all decisions to microcon were done manually by case managers. Most samples in this range were reported as "DNA insufficient" or "No DNA" by a member of the analytical section staff, who are not trained as case managers.

16. Different reworking strategies lead to different outcomes in terms of how much concentration can be obtained from a microcon process. Ideally, the higher the final concentration, the more likely the chance of obtaining a useful profile.
17. The best outcome for a Microcon is approximately 5.7 times the original concentration. After the initial quant there is usually ~85µL of lysate remaining.
    a. 85µL sample straight to 'mcon to 35µL' is a concentration factor of approximately 2.42
    b. 85µL straight to 'mcon to full' (~15µL) is a concentration factor of approximately 5.7
    c. 85µL amped at 15µL then 'mcon to 35µL' is a concentration factor of approximately 2.
    d. 85µL amped at 15µL then 'mcon to full' is a concentration factor of approximately 4.7.
18. As can be seen above, the ability to microcon to full greatly increases the likelihood of obtaining a DNA profile compared to other strategies.
19. We know that if a sample is simply amplified in the 0.0011µL and 0.0088µL range, the probability of obtaining a useful DNA profile is not high. Based on current laboratory protocols, 0.033ng/µL is the optimal amount required for an amplification. Amplification below concentration risks a suboptimal result where not all information present is subsequently obtained. It is poor practice not to Microcon between 0.0011µL and 0.0088µL.
20. I was concerned that the June 2022 change in process to amplification without microcon would lead to sub-optimal results at the end of the process, which might be seen to reaffirm that the 2018 decision to move to optional processing was justified.

Reviewing 'No DNA detected' results

21. If a case does not have a nominated offender, then the DNA results are only reported to the QPS via the Forensic Register. In these situations, a SOW is not written.
22. Samples that have a quant between 0.0011ng/µL and 0.0088ng/µL are validated in bulk by members of the analytical team. These staff are not trained case managers. This is particularly problematic in sexual assault cases.
23. There are situations where low-quant sexual assault samples that have sperm present, and where no offender was identified, are not being reworked.

24. If the quant result is in the range given above, it is reported as 'no DNA detected' or 'DNA insufficient'. As such, where no offender is identified and no other samples produce quants outside this range, then it is quite possible for a case manager to never see these samples. Accordingly, there is a very high probability that these samples would not be reworked despite there being a reasonable probability of obtaining a DNA profile.

25. I am aware of some significant examples of large numbers of sperm seen in samples with 'no DNA detected' or 'DNA insufficient' results (see table below). Spermatozoa are rated 0 (None observed); <1+ (Very Hard to Find); 1+(Hard to Find); 2+(Easy to Find); 3+(Very Easy to Find); 4+(Abundant). In my opinion, a 1+ (or higher) rated sample will generally produce a usable profile. Sperm counts of <1+ will often produce usable DNA profiles. I have seen a few samples with 2+ sperm reported as 'DNA Insufficient' that were subsequently reworked. I have never seen a sample with 3+ or 4+ reported as 'DNA Insufficient'.

Sample	Sperm	Initial Result	Final Result
[redacted]	1+	DNA Insufficient for further processing	2 Person mixture Support for suspect >100 billion. NCIDD upload.
[redacted]	1+	DNA Insufficient for further processing	3 Person Mixture.
[redacted]	1+	DNA Insufficient for further processing	2 Person mixture. NCIDD upload of unknown male
[redacted]	1+	DNA Insufficient for further processing	Single Source Support for suspect >100 billion. NCIDD Upload

26. A reporting scientist or reviewer might pick an issue up and decide to order a rework if they feel there is sufficient time to be able to get a result prior to a SOW being required for court. However, if the due date for court is too close, a reporting scientist may have to report the 'DNA not detected' or 'DNA insufficient' result even though it is less than ideal. Even if there are a couple of weeks available, hold ups in system or quality issues frequently cause reworks to take too long for a pending court date.

27. I am not aware of a time where the analytical team has rejected a request for a rework. I am aware that there are numerous instances where, samples have taken several weeks to pass through the analytical section due to processing and/or quality issues.

28. Seminal fluid and spermatozoa contain metallic salts of calcium, sodium, potassium, zinc and magnesium. Metallic salts can potentially inhibit DNA extraction and the ability to obtain a profile. Modern DNA kit chemistries are reasonably good at minimizing the effects of inhibitors, but there are plenty of examples of sperm positive samples not performing well at the quantification stage.

29. We have asked management that if sperm are observed, the analytical scientists don't review the 'no DNA' or 'DNA insufficient' results for sexual assault results, and it instead goes to a reporting scientist or for automatic rework. My understanding is that this request has been done both verbally and through email. Eventually, after some pressure, a spreadsheet was set up in November 2021 to record examples of where this was occurring. Numerous examples were recorded. This was a somewhat biased study as it only represented the 'no DNA / DNA insufficient' that were found as a result of routine casework and thus could be reworked and subsequently recorded in the spreadsheet. The majority of affected samples, in my opinion, would not have been discovered and therefore not reworked. Despite this evidence, there has been no change to the process.

Wording of statements

30. I have concerns about the wording of several matters in witness statements.

Multiple Unknown Profiles

31. If there is a result where there are multiple unknown DNA profiles, it is ordinarily reported without regard to the different unknown profiles. Some scientists will identify whether it is a male or a female unknown profile, but most scientists do not differentiate between unknown persons in a SOW, which in my opinion, has the potential to be misleading to stakeholders in the judicial process.

32. This means that if different unknown profiles are found across multiple samples in a case, the varying unknown profiles will not be reported in a statement beyond something similar to "This DNA profile did not match any the reference DNA profiles associated with this matter match the obtained DNA profiles and therefore is of unknown origin". Differentiation between unknown DNA profiles will, however, be available to the QPS in the Forensic Register as each unknown profile is identified using a unique number. Eg.

a. UKM1 - Unknown Male 1; UKM2 - Unknown Male 2, who is different to UKM1; UKM3 etc...
b. UKF1 - UK Female 1; UKF2 - Unknown Female 2, who is different to UKF1; UKF3 etc...
c. UKP1 - UK Person 1; UKP2 etc. This occurs when the sex indicator for the donor cannot be discerned).

33. I believe this same information should be reported in a SOW.

Three person mixtures that are potentially two person mixtures

34. The statistical modelling for STRmix requires a minimum number of contributors to model profiles. Whether or not a mixture is truly a third person mixture is often able to be determined by a scientist, using the electropherograms, at interpretation stage.
35. Under current processes, if there is uncertainty as to the number of contributors, it is common to add an extra contributor to the minimum it could possibly be. This is done because, mathematically, it is better to overestimate than underestimate the number of contributors, as the latter can lead to false exclusion or potential adventitious matches.
36. For example, a reported three-person mixture can potentially result from:
    a. three distinct contributors, with at least five clear alleles above the limit of reporting (LOR) present at one or more loci,
    b. two contributors, with no more than four alleles above the LOR at any locus but there are one or more high stutters or a very low-level sub-threshold piece of information that could potentially be additional DNA, or
    c. a single contributor with low peak heights, with numerous high stutter and/or potential subthreshold information.
37. Therefore, using the three-person example given in (b.) above, the result may become a three-person mixture for statistical purposes, when it really should be reported as a two-person mixture with some indication of a potential low-level third contributor.
38. While the statistical modelling may need to consider the possibility of a potential three-person mixture, this should be properly explained by the scientist in the SOW and in the more broadly in the SOW appendix.
39. This becomes particularly important in sexual assault cases, where a sample is reported as a three-person mixture with no further information. This may incorrectly

suggest to stakeholders that there was a third person's DNA present, when in fact, it is more of a mathematical construct.

40. I believe the appropriate report of a 'three-person mixture' result that actually looks like a two-person result would be something along the lines of:

    The mixed DNA profile obtained from this sample indicates at least 2 distinct contributions of DNA with a potential trace level third contribution. Therefore, it has been assumed for statistical purposes, that there are three contributions. A mixture of a least two people with a potential third person trace.

41. Or for a three-person mixture with a distinct major and two-person minor:

    The mixed DNA profile obtained from this sample indicates at a distinct contribution from a male donor and at least two lesser contributions of DNA. Therefore, it has been assumed for statistical purposes, that there are three contributions.

42. The exact wording should be developed in consultation with scientific experts.

Saliva testing

43. Currently DNA Analysis uses a very basic Phadebas method to detect saliva. I believe that our method of saliva testing is very outdated. The current method is suitable for the rapid screening of items, such as to localise where saliva may be located on an item of clothing. It is not a quantitative test and so provides no information on how much amylase (the active constituent of saliva that the test reacts to) is present. Amylase is present in sweat, vaginal secretions, and faecal matter. It is most concentrated in saliva but is also highly variable in concentration. Some people don't secrete amylase in their saliva. Amylase concentrations in saliva can also depend on the time of day and how recently someone has eaten.

44. The current method of saliva detection relies on a colorimetric change in the test (clear to blue). How distinct the blue coloration has to be to be recorded as a positive result is highly subjective and is also subject to other factors such as ambient light, or, for example, the observer's ability to perceive blue. The result of the current test is simply a record of the test turning blue and does provide information on how much amylase is present. As such, the potential for a false positive result is increased.

45. Irrespective of how intense the colour change is, the result is simply reported as 'positive for the possible presence of saliva'.

46. In my view, we should have moved to quantitative testing long ago. A quantitative test measures the amount of amylase in the sample. The amount of amylase would be extremely useful in providing reporting scientists information on the likelihood of the source being saliva or whether or not the detected concentration might potentially be the result of another bodily fluid. It would also decrease the chance of a false positive result being reported.
47. As a result, in my opinion, the appendix to the SOW overstates the value of the test.

Work system in laboratory

48. My understanding of the current work system in the laboratory is that it is possible, depending on the rostering of scientists, for separate scientists to perform the:
    a. extraction,
    b. quantification,
    c. 'no DNA' or 'DNA insufficient' culling,
    d. Micro-concentration (if relevant),
    e. amplification,
    f. CE analysis, and
    g. profile interpretation (by case managers).
49. The reporting and reviewing of cases and samples is done from a worklist on the Forensic Register (FR), where a case manager and reviewer will take the oldest matter on the list to do. This can mean that there is generally no contextual understanding of a sample except if the scientist examines other samples in the case. This is generally not done due to time constraints. A contextual framework is generally only obtained if the cases is assigned to a case manager. Case assignment represents only a small percentage of the cases examined.
50. Cases that require assigning are allocated by team leaders, such as Kylie Rika, Sharon Johnstone, and sometimes from Justin Howes.
51. I was the assigned case manager for the [redacted] matter. At that time my team leader was Amanda Reeves.
52. There is an issue in the FR whereby once an SOW had been released, further samples that are submitted are not automatically sent to the assigned case manager.
53. In 2019, further information had come in for the [redacted] case, which I was not aware of until someone told me some time later.

Validations

Quant Trio (Project 152)

54. This validation project, is in my view, is very poorly designed and contains multiple errors that have ramifications for other validations.

55. I do not believe the scientists undertaking this validation have the experimental, or statistical background required to understand the issues involved. They are capable at analysing DNA, but experimental design and analysis is a separate skill set.

56. This is by no means an exhaustive list, but reflects obvious errors within the report:

- The process consistently overestimates the quant value for single source DNA compared to the known DNA content (Table 6 & 7) with no explanation given as to why.
- The process consistently overestimates the quant value for mixed source DNA compared to the known DNA content (Table 6 & 7) with no explanation given as to why.
- The general overestimation of the quant is a likely reason why many reportedly high quant samples yield low rfu profiles. This should have been explored at the time of validation.
- While t-tests should not have been used at all, the incorrect type of t-test has been used throughout the experiment. The correct t-test should be a paired t-test whereas a t-test assuming different variances has been used.
- Using a paired t-test, there are three groups that are significantly different between plates A & B from table 14 at the $p \leq 0.05$ level and 1 group that is borderline at p=0.059.
- Using a paired t-test, there are two groups that are significantly different between plates A & C from table 16 at the $p \leq 0.05$ level and 1 group that is borderline at p=0.057.
- Using the correct t-test, there are two groups that are significantly different between plates C & B from table 16 at the $p \leq 0.05$ level.
- As stated previously, it is inappropriate to compare more than two groups using t-tests as it leads to an increased potential for a Type 1 error to occur (falsely rejecting the null hypothesis when in fact it is true). This is exactly what occurs in the experiment 4 data when using the correct t-test. The correct analysis for multiple groups in this scenario is an ANOVA.
- There is no exploration of the variation in the data. Quoting means without also quoting the standard deviation is meaningless as it indicates little about the estimated population distribution. For example, the mean of 51 & 49 is the same as the mean of 0.5 and 99.5, but the first example is far more preferable when investigating machine accuracy.

- Page 47 states that:
- "...the low t-test score at 0.01ng/uL is due to the low accuracy and the high variability at that DNA concentration level, therefore the t-test score of 0.00787 ($p \geq 0.05$) is not unexpected." A similar statement is found on p45-46. This sentence contains some errors:

    a) A t-test generates a probability 'p' (not a score). In essence, it is an estimate of the probability that any observed differences in the mean between groups are due to chance assuming that the null hypothesis is true. Thus, a p=0.01 implies that there is a 1% chance that observed differences between the means are not due some real effect.

    b) Therefore, from above it follows that "low accuracy and high variability" would have exactly the opposite effect, as it would mean there would be less chance the groups would be sufficiently separated to be significant.

    c) P needs to be smaller than the decided threshold (so $p \geq 0.05$ should be $p \leq 0.05$).

57. Additionally, in Section 7 it is stated that the limit of detection is 0.001ng/µL. It is unclear how this threshold was arrived at as there was no testing of concentrations less than 0.001ng/µL. Certainly, based on the results obtained, DNA could be detected fairly reliably at this concentration, albeit perhaps not accurately.

Repeatability and reproducibility

58. It is a requirement under NATA to have repeatability and reproducibility studies. For some considerable time in projects, it was the practice to, for example, put five samples on one plate, and five samples on another plate the next day, and consider this to be a repeatability and reproducibility analysis. I believe this is because the project officers fail to understand that they are testing the machine process, not testing the samples. When testing a machine run or a process, the experimental unit is the machine run or the process itself. The samples are only a variable by which the machine process is measured.

59. In relation to a machine validation, repeatability is the ability to get the same result consistently in a short series of runs of the machine process, and reproducibility is the ability to get the same result over time with different operators/conditions (it can also mean by different teams in different labs, but this does not apply to our studies). When testing a machine run or a process the experimental unit is the machine run or the process itself (represented in this case by a single plate). Thus, this study has

only four experimental units, which means it is a poor reflection of population of process runs it is trying to estimate. There are factor levels of "repeatability", "reproducibility", and DNA concentration level. Having large numbers of repeated DNA "samples" on a single run is generally meaningless (other than as an indicator of intra-sample and preparation variation but can be averaged to minimise variation from the quant process, pipetting errors, and other unwanted noise) and is an example of pseudo-replication. As such, sections 4a & 4b do not meet the generally accepted five repeats and five reproductions that are recommended by groups such as NATA and ENFSI.

60. I believe the validators do not understand the difference between the experimental units, factors, treatment levels, true replicates, pseudo-replicates and often confuse them. This is a mistake made in many validations.

61. I had some success with convincing Paula Brisotto that the lab was doing repeatability and reproducibility incorrectly and was able to get the process changed to running five plates across five days as a suitable means of investigating reproducibility. However, I noticed that recently this procedure is still not always followed (see Project #199 – where each machine was run only once with a large number of replicate samples and therefore the project has no repeatability or reproducibility analysis).

Response

62. Something, which I cannot recall, had flagged to me that there might be an issue with this validation, and I went back to it and found the above issues.

63. I have had discussions with Paula and Justin in the past about my concerns with our validations with only limited success in effecting change.

64. I sent an email to Justin on 8 March 2018 about the issues in the Quant Trio validation, but I never received a response. I did not raise it again as I had come to feel by this time that there was little point. Annexed and marked RP-04 is a copy of the email sent to Justin on 8 March 2018, and annexed and marked RP-05 is a copy of the attachment to that email.

65. I believe there is a history of poorly done validations. A piece of machinery is bought, and it is post-hoc justified / validated and then accepted as being fit for purpose even if the data obtained indicates otherwise.

Quant Studio5 (Project 185)

66. There are multiple issues with the validation, many the same as Quant Trio.

67. The design for this study is severely flawed. The proposal for Project 185 was also flawed, and annexed and marked RP-06 is a copy of the proposal with my hand written notes.

NIST Standards

68. A NIST standard is a small amount of precise, known quantity of DNA. It took me, and other staff, many years to convince management to use standards for quant studies. Eventually, the use of standards became routine.

69. This validation has 18 DNA concentration factors. There is a general tendency for the SAT recorded quant to be lower than the expected SAT quant. The standard itself is not 100% accurate and has a narrow range in which its concentration is guaranteed. So, some of the lower quant could be a result of the standard being at the lower end of the range. There is no evidence that this consistent underestimation was ever investigated, or even noticed.

70. It is unclear why NIST-A and NIST-B are considered as different. One is single-source male DNA and the other is single-source female DNA. There is no reason to believe that gender is likely to be a confounding variable. But in this experiment, it is treated as though it is for some reason.

71. I also found, in the Quant Studio5 validation folder, a document that outlines NIST standards being used out of date for the QuantStudio. I do not know who authored this document. This document is Annexed and Marked RP-07.

Percentage differences

72. The validation has taken only two measurements and averaged the differences between the results, measuring the percentage difference of results for repeatability and reproducibility.

73. The issue with measuring percentage differences between only two results is that depending on which result is obtained first, the percentage difference will change.

74. For example, if the first result is 100 and the second is 150, it is a percentage difference of 50% increase, but if your first result is 150 and your second is 100, the percentage difference is 33.33% decrease. Clearly, these are not the same despite the obtained measurements being the same.

75. At 5.3, experiment 3a, it states that repeatability "*is an assessment of whether the QS5 produces the same results when one sample set is processed in duplicate by one user under the same conditions*", and "*Plates 1 and 2 from the Sensitivity and LOD experiment will be used for this experiment*". This is again not repeatability because there is only one repeat (ie. two runs of the machine process of interest), which is not considered sufficient. When reproduced, it is only run one other time. Again, there are two repeats, but the authors think they have twelve because of the pseudo-replication on each plate. This, in my opinion, severely compromises the results and conclusions that can be drawn.
76. The average errors calculated at the bottom of Table 3 are incorrect. Negative and positive "errors" have been added together and have effectively cancelled each other out.
77. The t-tests performed in Table 4 are between all of the NIST A result on each of the machines. The NIST A results are a serial dilution and are therefore not from the same population and thus violate the assumptions of IID for t-tests. But even assuming the sample population was correct, a t-test is not the proper test for this type of comparison as an ANOVA is required.
78. The authors have compared QS5-A and QS5-B with the 7500 using t-tests. The issue with measuring this way is that the difference between QS5-A and the 7500 may not be great, and the difference between QS5-B and the 7500 may not be great, but the difference between QS5-A and QS5-B may be very different.
79. The authors have not tested whether the QS5 machines are the same as each other, they have just assumed they are. In not comparing them to each other, the validation does not consider whether one machine falls above, and one falls below (for example as they do in Figure 2).
80. At the bottom of page 8, p-values are used for comparing variation. "*Variability in quantification result repeatability for both QS5s across targets and NIST standards is apparent as can be seen from the P-values in Table 5....*". This demonstatrates a lack of understanding of what p-values represent.

Project Report tables

81. On page 12 of the Project Report, Figure 4 shows that at 0.09ng/µL there is a 125% change between the two results, which is odd. There is no mention of looking into this further. Similarly, at 0.009ng/µL. In Figure 5 on page 13, a similar issue appears

with the 0.002ng/µL with a nearly 150% change. Radical departures from the expected are not mentioned or examined further

82. An additional issue with percentage-based analyses is that they don't scale, so a 25% error when speaking about 5ng is very large absolute error, but 25% at 0.005% is not significant. At the lower concentrations, the greater inaccuracy is probably not an issue, however where high concentrations have ±25-30% errors, it goes unnoticed and unmentioned.

83. In any event, the methodology does not properly examine the repeatability and reproducibility of the QS5 machine process. As it stands, the experiment is misconceived.

NIST OQI

84. OQI 56218, annexed and marked RP-08 relates to NIST standards being used outside of their use-by date. The standards in question expired on 31/12/2017, and in Project 185, these expired NIST standards were used in serial dilutions to compare the 7500 and QS5 in terms of sensitivity, limit of detection and accuracy. The OQI document states: "... *the accuracy of the concentration of each serial dilution is not the critical element of this experiment as the QS5/7500 were assessed at several concentrations above and one concentration below the LOD. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments at these reducing concentrations. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of this experiment showed comparable performance between the 7500-A, QS5-A and QS5-B and recommended the LOD remain at 0.001ng/µL.*"

85. These assertions are problematic. The 7500 was an old machine that was not likely performing optimally, whereas the QS5 should have been performing optimally. Merely determining if the QS5 was as good as the 7500 is, in my opinion, a poor criterion for acceptance. In that regard not being able to determine if the QS5 machines were achieving expected outcomes is a major short-coming. It is also an example of how major issues are written off as insignificant or not relevant.

86. The document states that quantification is an estimation and has shown to have variation of +/-30% in successive internal validations (Quant Trio Validation – see point 51 above). If this is the case, it is unclear why the experiment relies on so few

samples in each group, instead of running five or ten to attempt to control for this variation.

87. The Limit of Detection (LOD) value of 0.001ng/µL is based on the Quant Trio experiment, which did not examine DNA concentrations below this level. It also contains problematic maths and experimental design. However, the 0.001ng/µL threshold has become a hallmark of our standard operating procedures.

Validation of QIAsymphony SP for Bone Extraction (Project 192)

88. This validation has similar issues to the previous two validations.

89. This validation looked at the extraction process for bones. Extraction used to be done organically, but the chemicals used are potentially hazardous, so there was a desire to move away from organic extraction towards bone extraction using the robotic platforms.

90. Table 1 on page 5 shows the ten case work samples that had come in for identification historically. The normal process is to get four sub-samples (aliquots) of each bone and submit them all separately and ideally the aliquots should all come back with similar quants and the same DNA profile. Each was quanted historically, and their range (of the four aliquots) is found in the 'Original Quant Range' column. These original samples were quanted after an organic extraction.

91. If you compare Table 1 to the actual results obtained on pages 6 (organic extraction), 8 (overnight extraction), 9 (5-hour extraction), the results do not compare well. Table 2 shows the result for Sample 2 as 1.883, where the expected result was 10-20. Sample 3 gives twice as much as expected, and all but one result is markedly different from the expected result. The negative control also had an allele count of 9 where it should have been 0. The contamination event is discussed but accepted and the experiment is not redone.

92. These results are highly variable. The main problem is that the authors have taken these results and just accepted them. There is no investigation into why there might have been such marked differences between the expected and the obtained. It is equally unclear why only a single quant measurement (as per the Project Proposal Methodology) was taken for each sample. As such, there is no way to account for inter-aliquot variation and inter-quant variation.

93. I believe that the person who has done this extraction was not skilled at organic bone extraction. Organic has always been considered to be one of the best methods for

extracting DNA from bone, but it has to be done by someone who is skilled at it. The evidence for this comes from the results obtained. For example, Sample 6 and other Supplementary project results where the organic results are not consistent with expectations.

94. Again, this study lacks repeatability and reproducibility and fails to understand what constitutes an experimental unit.

95. One of the bone samples had a known quant value of 0.00 (and so should not have been included in any study) except as a potential negative control.

96. Furthermore, it is not valid compare the bone that has 50ng/µL to the bone that has 2ng/µL, because they are such different concentrations and the variability of each are very different. However, the validation has compared each bone on page 12 and 13 as if they are from similar populations.

97. There is no exploration of why three aliquots (2, 4, and 7) with a high concentration of DNA only produced a partial profile (Experiment 1 Table 2). This is evidence that some other aspect of the process has failed but it is not examined (or repeated).

98. There are five treatment groups, with eight dilution levels (Sample 9 doesn't count and Sample 4 was removed due to contamination) in each and only one aliquot for each bone per concentration level. As such, this is an example an n=1 study. It is not possible to extrapolate any information from this study because, as it stands, it is meaningless.

99. I complained about this validation verbally to Paula and explained the issues with regard to its design. She took notes and said it would be fixed. I was not consulted again. Further validation work was undertaken in supplementary Project 192. This was to correct the repeatability and reproducibility issues as well as increasing 'n'. It is unclear if many of the aspects discussed above were carried out correctly as the methodology in the supplementary project is extremely vague.

100. I have some concerns with results as presented in the Supplementary study. The repeatability and reproducibility tests should be similar because the same aliquots are being used for both. Thus, the relationship between the result obtained from each bone for the organic extraction and the two robotic extractions should be similar. That is, the Bone 1 repeatability graph should look very similar to the Bone 1 reproducibility graph. Similarly for Bone 2 etc. However, in this validation, repeatability results are not consistent with the reproducibility results. Irrespective, the data has been found in favour of the robotic platform without investigation into these inconsistencies.

[redacted]

Annexed and marked RP-09 is a copy of the Project Report #192 for Validation of Qiasymphony SP for Bone Extraction which contains data from the supplementary validation.

101. It has been observed that there are a lot of mixed DNA profiles from recent bone samples. It seems to coincide with commencement of the QS5 process and I am unsure whether it is this process or whether it is something in the sampling (as there were untested procedural sampling changes that occurred).

102. I am concerned that a proper investigation into which method was best (organic vs robotic) was not successfully conducted.

Risks associated with validation issues

103. While the risk of the below occurring are very low, the potential effect of an occurrence is potential extremely damaging. I outlined these risks in my email to Justin Howes on 8 March 2018 about the issues in the Quant Trio validation. As previously stated, I did not get a response. The risks included:

    a. Defence asking for copies of validation studies and seeking expert advice on the results.
    b. The rejection of DNA evidence due to inappropriate validation/verification of equipment.
    c. Potentially having to rework hundreds or thousands of samples.
    d. Losing scientific respect nationally by other DNA labs.
    e. Losing the confidence and respect of the community because any successful defence challenge will be in the public arena.
    f. Having to contend with an ongoing defence challenge and corresponding s95 reports as the lab's underlying science will be viewed as weak.

104. With design improvements, many of these experiments could have been done to a much greater degree of scientific validity with minimal extra cost or in many cases lower cost.

105. The low quality of the validations / verifications means that the lab has a poor understanding of the variation expected from various pieces of equipment. This potentially leads to unnecessary re-amplification and ReGS in order to obtain consistent EPG results, especially given that the Quant Trio system appears prone to over-estimating the quant, which could lead to under-amplification.

106. There is a potential major cost of having to redo thousands of samples deemed inadequate due to insufficient validation quality in the event of a successful defence challenge.

107. There is also a concern where the results of one study are used as a foundation for subsequent studies. This compounds the error, even if the subsequent studies were to be conducted correctly.

108. The issues listed with the above projects are not limited to these few examples given. In my opinion, the types of errors listed above are repeated extensively through many if not most of our validations and projects.

Professional development

109. Staff at FSS are routinely denied the ability to obtain new skills. Secondment or temporary release to work elsewhere is not an option. Several valuable staff had to quit their positions in order to work elsewhere on short-term contracts (eg. Robert Morgan, Julie Connell). The experience and skills they would have gained from these positions would have been extremely useful to FSS.

110. In my 2014 Performance and Development Plan, I requested to undertake training in statistics in order to refresh skills that I had not used for many years, and to learn new techniques for statistical analysis that had become routine due to improvements in computing. I was not actively supported to do so other than being allowed to use PDL (Professional Development Leave) to attend exams. Upon request, I was directed to SSDU, who stated that as the course was not considered essential, it was not likely to be covered by any form of support. At the time, I felt that given the onerous paperwork required for the application process, and given the low probability of success, I did not proceed further with it and funded all my study myself. At one point I requested permission from Justin Howes to photocopy some notes, but this was denied.

111. I am the only person in the laboratory with a higher-level statistics qualification (to my knowledge). Since gaining this qualification, I feel I have been actively excluded from input into project design and analysis. I have heard from some staff that they had been told specifically not to seek advice from me. I believe that this is because the way I want to analyse results or run projects often leads to outcomes that are at odds with the outcomes desired by the Decision-Making Group (DMG) and requires an

understanding of statistical methods and experimental design with which the DMG are unfamiliar.

112. It is my belief that Emma requested my assistance for part of the Verifiler stutter analysis. It is my understanding that Emma asked management if she could bring me in on the analysis and was told no by Justin Howes. She then asked Kirsten Scott (Project Leader), who allowed it as there was no-one else capable of running the analyses required.

113. After performing a number of analyses for the Verifiler Team, Emma Caunt, Cassandra James, Angela Adamson and I co-wrote a feedback document, and a copy of this document is annexed and marked RP-10. However, Kirsten responded in email that the choice of authors made her "uncomfortable". A copy of this document is annexed and marked RP-11. I understood this to mean that management team members who had not contributed to the analysis had been left out as authors and that I should not have been included, despite being a major contributor to the work and the final document. This, in my opinion, is a clear example of the professional exclusion that occurs within DNA analysis.

PowerPlex21 (PP21) and STRmix

114. The [redacted] case was, from memory, the first big case we did using PP21 and STRmix.

115. I have never had any issues with the introduction of STRmix itself; it is based on sound statistical methodologies.

116. It was a national agreement to use PP21, but after we implemented it, the other states decided not to use it. PP21 has a few issues: it is highly stochastic at low levels of DNA and even from amplification to amplification in samples with good levels of DNA. This means that it is common to see quite significant fluctuations in peaks heights from one amp to another (where they should be reasonably consistent). STRmix however, tends to be quite good at handling this variation.

117. The use of PP21 is more an issue because the model that STRmix is based on assumes certain essential patterns in the way DNA behaves, and as PP21 doesn't always reflect those patterns, you can get results that are acceptable, but probably not as ideal as they could be. This is a result of the aforementioned stochastic effects observed in PP21. We have recently looked at Verifiler to replace it, but it suffered from similar issues and STRmix did not model it as well as it could have.

118. I did some of the early STRmix validations but wasn't involved in any of the PP21 validations.

Sperm microscopy

119. Sperm microscopy is the process of visual observing spermatozoa cells in samples. Traditionally, there are two ways of doing this; smears and suspensions. A smear is when the sample (eg. swab) is wiped across the surface of a microscope slide thereby leaving a smear which can then be stained and visually searched for sperm. A suspension is where the sample is soaked in water and some of the sperm/cellular material present becomes suspended in the water. A small amount of that water is added to a microscope slide, dried, stained, and searched for sperm.
120. The DNA Analysis Unit traditionally used smears but switched to suspensions, which worked okay until about 2014, when, I believe, Alan McNevin decided to change some aspect of the suspension sampling process. I believe Amanda Reeves discovered that there was a disparity in the numbers of sperm observed at the evidence recovery stage compared to the numbers of sperm being observed on the differential lysis slides (a slide made from a suspension used as a back-up control for the differential lysis process).
121. If a DNA profile was not obtained from the sample but sperm had been observed at the evidence recovery in the beginning, we could go to the differential lysis slide to see if the differential lysis process had been successful.
122. It is my understanding that Amanda raised this issue in order to get the process changed and that this led to a protracted series of events (of which I have no firsthand knowledge) that was finally resolved when Matt Hunt re-examined the issue and developed the current process, several years after the issue was first raised.
123. It is remarkable to me that it took literal years to resolve a simple technical issue that could have been resolved in a matter of weeks. In all that time, there was potential for evidence to have been missed and/or samples to have not been processed optimally.

Culture

124. The success of raising issues depends on who raises the issue.

125. In my experience, the burden of scientific evidence required for acceptance is far, far greater for a project that finds adverse to management objectives than one that aligns with them (for example see Verifiler project vs PP21 project).
126. I have had some limited success with raising issues with Paula Brisotto, but management never come and ask how an experiment could be designed or how best to analyse the results. As such, many of our validations are invalid.
127. I do not think I have all the answers, but as it stands, I believe I have a better understanding of many of the issues faced in experimental design and experimental analysis than most other staff but have been actively excluded in using and developing these skills. I believe it would be very advantageous be able to freely consult with external experts in experimental design, statistics and/or validation to improve our skills in this area and to provide feedback on experimental plans/methods.
128. I believe there needs to be a separate project team that is independent from the management team. This team would be responsible for data mining, design and analysis of validations, and other similar projects. Ideally, it would consist of maybe two permanent scientists and additional staff could be rotated in as needed.
129. I feel that despite the gender balance of the management team, the laboratory culture is quite misogynistic. It is my perception that female staff that require more work flexibility due to familial commitments often have difficulty obtaining it.
130. There was once a situation where Amanda Reeves and I were in on the weekend to get a Priority 1 sample completed and uploaded to NCIDD. There was something unusual about how the DNA profile would need to be reported and we could not contact any managers, so we made an executive decision to do it in a certain way just to get it onto NCIDD so that QPS would have access to the intel. I thought it was a slight variation of process, but a reasonable decision. I am aware that Amanda got reprimanded for this, but I never heard further about it.
131. I believe that management have highly prioritised turnaround times, QPS requirements, and cost saving over result quality.
132. There are three categories of Quality notification in the lab. In descending order of severity these are: Opportunities for Quality Improvement (OQI), Adverse Events, and Notifications. I believe that over time there has been a gradual dilution of quality systems. Issues that historically would have led to an OQI now are considered

Adverse Events; historical Adverse Events now tend to be reported as Notifications and many things that would have been notifications historically are now overlooked.

All the facts and circumstances declared in my statement are within my own knowledge and belief except for the facts and circumstances declared from information only, and where applicable, my means of knowledge and sources of information are contained in this statement.

TAKEN AND DECLARED before me at Brisbane in the State of Queensland this 28th day of September 2022.

[redacted] .... [redacted] .....

Rhys Parry Witness

Schedule of Exhibits

RP-01	Model titled "Mconc_Cubic_Model_Only.R" dated Friday 4 August 2017
RP-02	Graph titled "Plot of Cubic Function of Mean Quant vs Probability of Mcon Success"
RP-03	Rhys Parry - Response to Project #184 Proposal
RP-04	Email from Rhys Parry to Justin Howes on 8 March 2018 "Quant Trio validation"
RP-05	Attachment to email from Rhys Parry to Justin Howes on 8 March 2018 titled "Quant Trio Issues Report.doc" "
RP-06	Project Report #185 Validation of two QuantStudio 5 Real-Time PCR Systems dated June 2017, with handwritten notes by Rhys Parry
RP-07	Documents outlining NIST standard issues
RP-08	Report for OQI 56218 Use of NIST Standard in Project#185
RP-09	Experiments of concern conducted by Rhys Parry, and Project Report #192 titled "Validation of QIAsymphony SP for Bone Extraction – Supplementary Repeatability and Reproducibility"
RP-10	Feedback document tittled "Analysis of data for stutter threshold selection for VFP" authored by Emma Caunt, Cassandra James, Angela Adamson and Rhys Parry
RP-11	Email chain between Kirsten Scott and Cassandra James dated 23 June 2021 titled "Query authorship RE: Verifiler Stutter" and "Verifiler Stutter"

[redacted]..... .....

Rhys Parry Witness

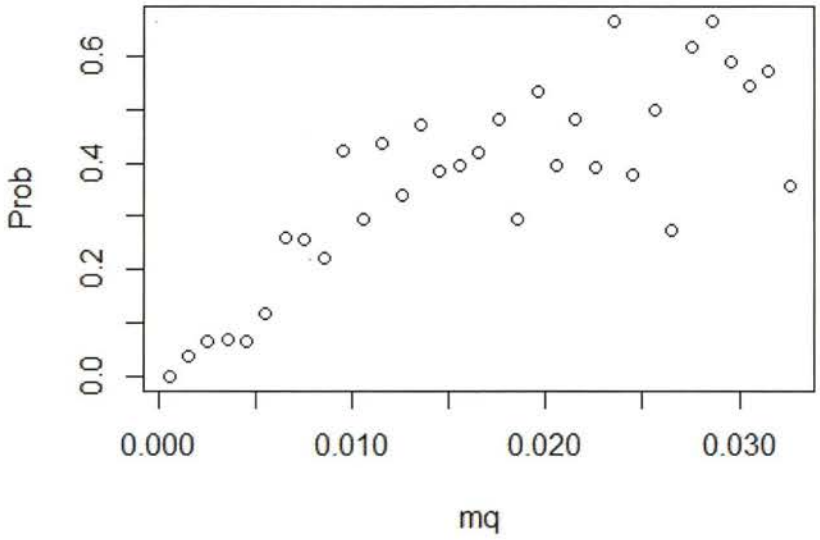
RP-01

Mconc_Cubic_Model_Only.R

ParryR

Fri Aug 04 16:12:57 2017

```
################ MCONC.R ###################
#Predict probability of success of microconcetration proceedure based on
quant of sample.

# set the number of significant figures for output
options(digits=3, show.signif.stars=T)
par(mfrow=c(1,1))
rm(list = setdiff(ls(), lsf.str()))


# Read data
library(readr)

## Warning: package 'readr' was built under R version 3.3.3

mconc.df <- read_csv("G:/RJP HP4 FRIT/Projects/Mconc Project/Mconc
Data/MCONCDATACSV.csv")

## Parsed with column specification:
## cols(
##   Barcode = col_integer(),
##   Quant1 = col_double(),
##   exh = col_character(),
##   Quant2 = col_double(),
##   probsucc = col_double(),
##   Quant3 = col_double(),
##   logmult = col_double(),
##   Total = col_integer(),
##   Prob = col_double(),
##   mq = col_double(),
##   Scnt = col_integer()
## )

mconc.df$exh <- factor(mconc.df$exh)

# Plot the data
plot(Prob~mq, data = mconc.df)
```




```
mconc1.lm <- glm(Prob ~ mq, data=mconc.df, family="binomial", weights=Total)
mconc2.lm <- glm(Prob ~ mq+I(mq^2), data=mconc.df, family="binomial",
weights=Total)
mconc3.lm <- glm(Prob ~ mq+I(mq^2)+I(mq^3), data=mconc.df, family="binomial",
weights=Total)


anova(mconc1.lm, mconc2.lm, mconc3.lm, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Prob ~ mq
## Model 2: Prob ~ mq + I(mq^2)
## Model 3: Prob ~ mq + I(mq^2) + I(mq^3)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        31        125
## 2        30         54  1     71.2  < 2e-16 ***
## 3        29         38  1     16.0  6.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Summarise the fitted model.
print(anova(mconc3.lm))
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
```

```
##
## Response: Prob
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                        32        407
## mq        1    281.7        31        125
## I(mq^2)   1     71.2        30         54
## I(mq^3)   1     16.0        29         38
```

```
1-pchisq(38,29)
```

```
## [1] 0.122
```

```
# Produce the Analysis of Variance table
print(summary(mconc3.lm))
```

```
##
## Call:
## glm(formula = Prob ~ mq + I(mq^2) + I(mq^3), family = "binomial",
##     data = mconc.df, weights = Total)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.906  -0.851   0.156   0.752   2.103
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.18e+00   2.73e-01  -15.28  < 2e-16 ***
## mq           5.66e+02   7.06e+01    8.01  1.1e-15 ***
## I(mq^2)     -2.64e+04   5.05e+03   -5.23  1.7e-07 ***
## I(mq^3)      4.08e+05   1.04e+05    3.95  8.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 406.971  on 32  degrees of freedom
## Residual deviance:  38.046  on 29  degrees of freedom
##   (1698 observations deleted due to missingness)
## AIC: 171
##
## Number of Fisher Scoring iterations: 4
```

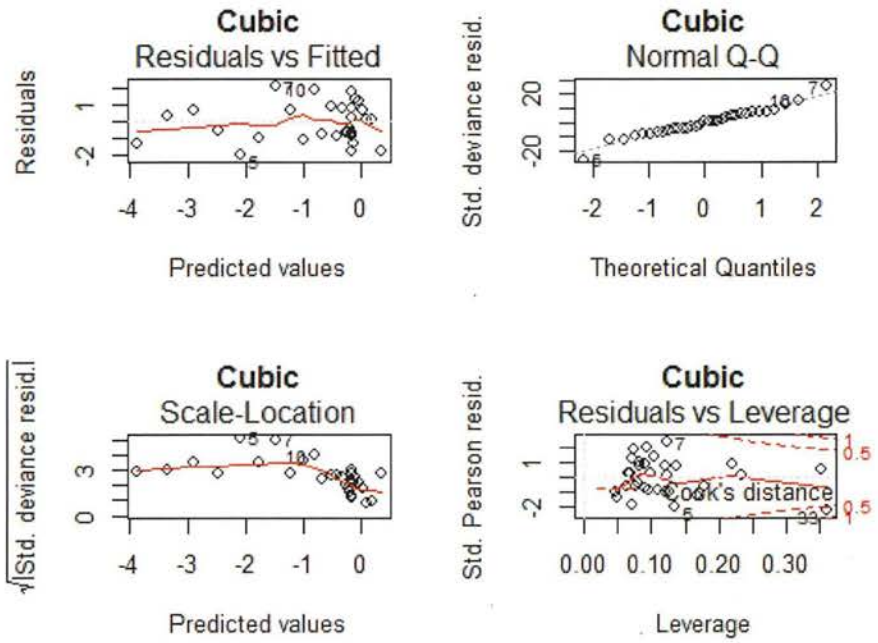
```
# Confidence intervals for regression parameters
print(confint(mconc3.lm, level=0.95))
```

```
## Waiting for profiling to be done...
```

```
##                   2.5 %    97.5 %
## (Intercept)      -4.73     -3.66
## mq              429.79    706.93
## I(mq^2)      -36445.71 -16631.87
## I(mq^3)      207091.07 613341.55
```

```
# exp(cbind(OR = coef(mconc3.lm), confint(mconc3.lm)))
```

```
par(mfrow=c(2,2))
plot(mconc3.lm, main="Cubic")
```




```
par(mfrow=c(1,1))
```

```
plot(Prob~mq, data = mconc.df, ylab="Predicted Probability of Successful
Mcon", xlab="Mean Quant", Main="Cubic")
```

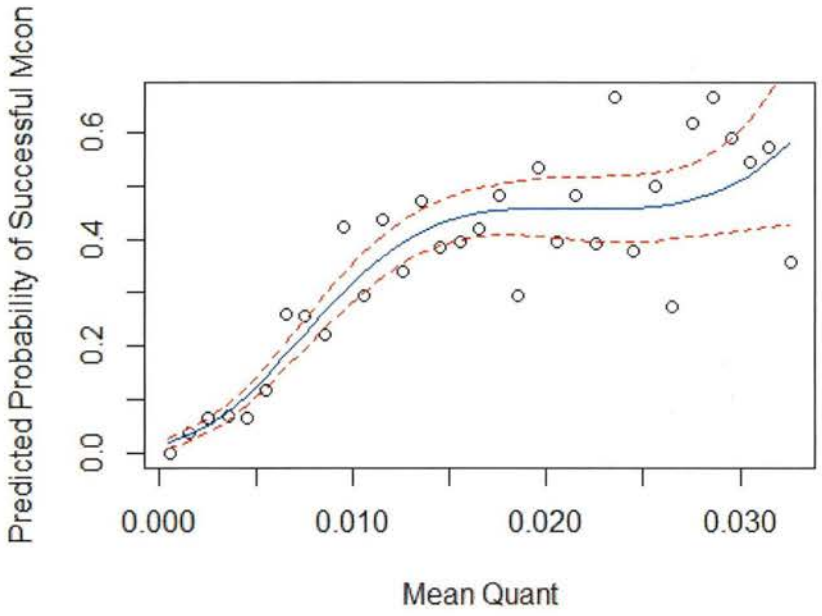
```
## Warning in plot.window(...): "Main" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "Main" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "Main" is not
## a graphical parameter
```

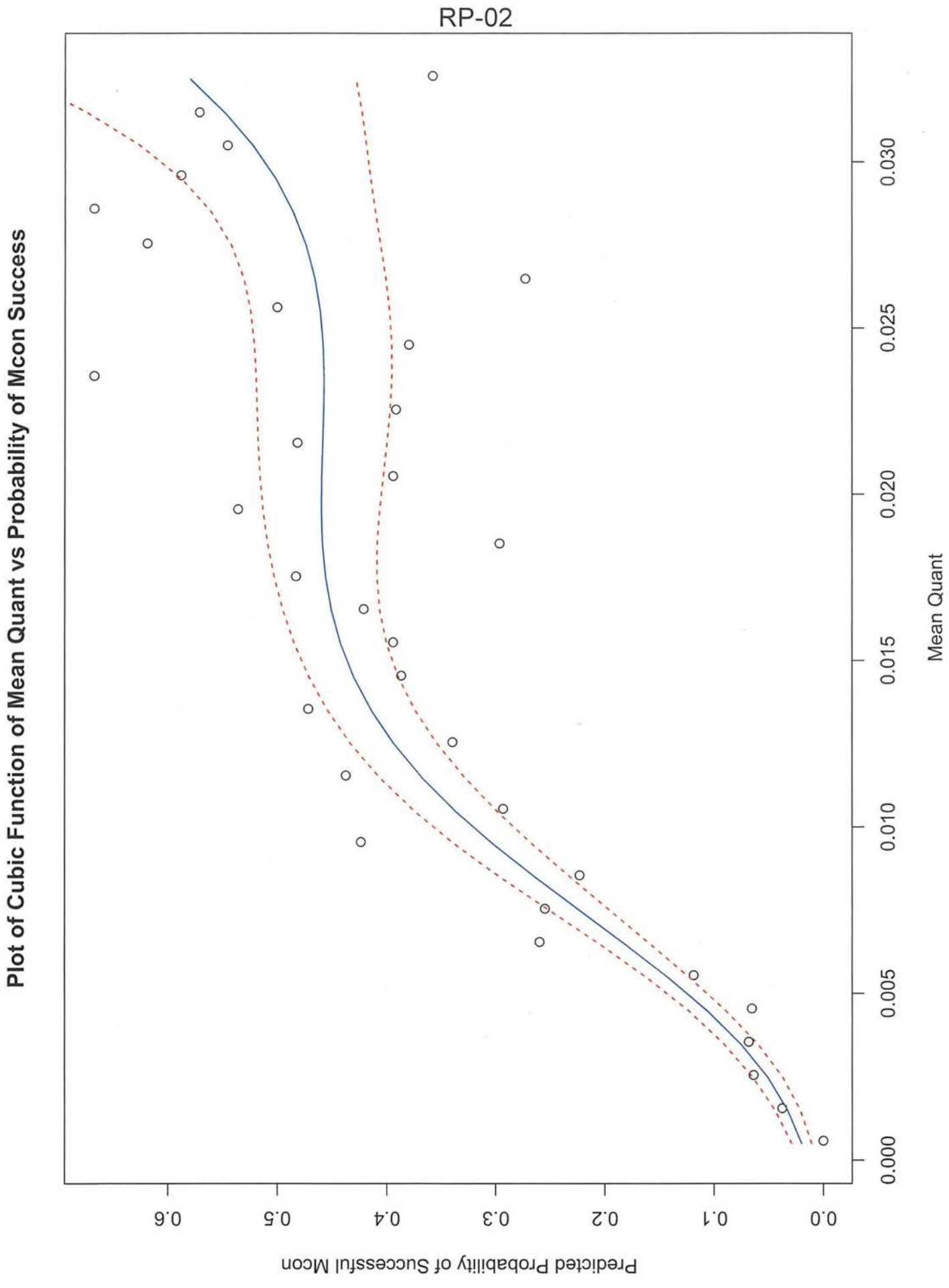
```
## Warning in axis(side = side, at = at, labels = labels, ...): "Main" is not
## a graphical parameter
```

```
## Warning in box(...): "Main" is not a graphical parameter

## Warning in title(...): "Main" is not a graphical parameter

newdata2 <- with(mconc.df, data.frame(mq=seq(0.0005, 0.033, by=0.001)))

preds <- predict(mconc3.lm, newdata2, type="response", se.fit=TRUE)
mqnew <- seq(0.0005, 0.033, by=0.001)


predf <- preds$fit # predicted
lower <- preds$fit - (1.96*preds$se.fit) # lower bounds
upper <- preds$fit + (1.96*preds$se.fit) # upper bounds

lines(seq(0.0005, 0.033, by=0.001), predf, type="l", bty="n", col="blue")
lines(seq(0.0005, 0.033, by=0.001), lower, lty=2, col="red")
lines(seq(0.0005, 0.033, by=0.001), upper, lty=2, col="red")
```




```
mqnew <- mqnew+0.0005
lower <- lower*100
upper <- upper*100
predf <- predf*100

predtable <- data.frame(mqnew,predf, lower,upper)
Results <- (predtable[,c(1,3,2,4)])
print (Results)
```

```
##    mqnew lower predf upper
## 1  0.001  1.06  1.98  2.91
## 2  0.002  2.11  3.28  4.44
## 3  0.003  3.75  5.12  6.49
## 4  0.004  6.04  7.57  9.11
## 5  0.005  8.94 10.65 12.35
## 6  0.006 12.28 14.24 16.21
## 7  0.007 15.87 18.21 20.55
## 8  0.008 19.55 22.34 25.12
## 9  0.009 23.21 26.42 29.63
## 10 0.010 26.71 30.26 33.81
## 11 0.011 29.96 33.74 37.52
## 12 0.012 32.86 36.77 40.68
## 13 0.013 35.35 39.32 43.29
## 14 0.014 37.38 41.39 45.40
## 15 0.015 38.92 43.00 47.09
## 16 0.016 39.99 44.20 48.42
## 17 0.017 40.63 45.04 49.46
## 18 0.018 40.89 45.58 50.27
## 19 0.019 40.87 45.87 50.87
## 20 0.020 40.65 45.97 51.29
## 21 0.021 40.32 45.95 51.57
## 22 0.022 39.98 45.86 51.74
## 23 0.023 39.69 45.76 51.84
## 24 0.024 39.52 45.73 51.93
## 25 0.025 39.53 45.81 52.09
## 26 0.026 39.72 46.07 52.42
## 27 0.027 40.07 46.57 53.06
## 28 0.028 40.54 47.37 54.20
## 29 0.029 41.02 48.53 56.05
## 30 0.030 41.46 50.13 58.80
## 31 0.031 41.84 52.20 62.57
## 32 0.032 42.24 54.82 67.39
## 33 0.033 42.79 58.00 73.20
```


RP-02
Plot of Cubic Function of Mean Quant vs Probability of Mcon Success
Predicted Probability of Successful Mcon
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Mean Quant
0.000
0.005
0.010
0.015
0.020
0.025
0.030

RP-03

RJP – Response to Project #184 proposal

- I will leave issues of what constitutes "meaningful information" and the issue of NCIDD "interaction" to others, though it think the former it needs to be specifically stated in the abstract (even a single peak can be informative if it excludes) and the rationale for the latter as a criterion needs to be stated more clearly in the introduction.
- It is good that source and substrate data have been added, though it would have been ideal to also gather sample source (blood/semen etc), substrate(swab/tapelift etc) for all samples as this could have been factored into the analysis using standard linear modelling techniques.
- It should be "n" not "N" for a sample size ("N" refers to a population size)
- Pg 14. It is unclear if the n=2201 is before or after the exclusion of unsuitable samples.
- My main concern with this proposal is the use of percentages and non-normalized data to draw conclusions from the data that are not valid.
    - By not normalizing the very low quant (<0.0088ng/uL; n=1449) data which represents the bulk of the samples(n_{total}=1731), percentages derived from data combined with the above very low quant samples (eg. Figure 8 and figure 9) are artificially skewed by the large number of close-to-zero quant values. Thus, it would not be expected for there to be an insignificant increase in the percentage of successful microcons as presented in figures 8 & 9). Even if 100% of the microcons in the 0.015-0.020 range were successful (n=94), this would have little effect on the mean success rate of the n=1492 samples that have lower quants (94/1492 = 6.4%) at maximum.
    - The data needs to be normalized by obtaining the probability for the mean quant using a frequency distribution for a range of quant values.
    - My own analysis of the data shows that the data can be best modelled by a third order regression of the success/fail probability against the quant. I developed the data as a frequency distribution based on divisions of 0.001 ng/uL. The probability of success was calculated based on the outcome of all samples within a single division, thus normalizing the data. This reduced the data to 33 points. The data was analysed as a binomial distribution as is appropriate with binomial data and the 95% confidence intervals calculated.
    - These outcomes are presented in graphical and tabular form in the attached pages suggests a very different set of conclusions.
    - As can be seen from the results there is a mean success rate of approximately 30% at 0.010ng/uL up to approximately 43% at 0.015ng/uL. This is at odds with the conclusions drawn in section 7.2 of the project and with the justification for the use of 0.015ng/uL in the introduction to Experiment 2 (pg 8).

- As such, I conclude that setting the cut-off for no processing at 0.0088ng/uL is probably too high.
- Additionally, conclusion drawn from percentage values derived from non-normalized data cannot be trusted as the data is clearly skewed towards very low-level quants.

Table 1. 95% confidence intervals for the microcon success probabilities for all quant ranges. (eg. Line 6 represents the probability of success for all samples with a quant between 0.0055 and 0.0064.)

	Mean Quant for range	lower	Estimated Prob of Success	upper
1	0.001	0.061921	1.984695	2.907470
2	0.002	2.111484	3.275817	4.440151
3	0.003	3.746543	5.116828	6.487114
4	0.004	6.038001	7.574229	9.110456
5	0.005	8.936327	10.645507	12.354687
6	0.006	12.277503	14.244627	16.211752
7	0.007	15.868023	18.210662	20.553300
8	0.008	19.552401	22.337853	25.123304
9	0.009	23.205051	26.415076	29.625101
10	0.010	26.709850	30.259965	33.810081
11	0.011	29.959510	33.738579	37.517648
12	0.012	32.862823	36.769795	40.676767
13	0.013	35.350065	39.319138	43.288211
14	0.014	37.375481	41.387961	45.400441
15	0.015	38.919212	43.002380	47.085547
16	0.016	39.989907	44.204209	48.418510
17	0.017	40.625908	45.044506	49.463105
18	0.018	40.891674	45.579421	50.267168
19	0.019	40.869451	45.867744	50.866037
20	0.020	40.649724	45.969556	51.289388
21	0.021	40.323576	45.945520	51.567465
22	0.022	39.977440	45.856505	51.735570
23	0.023	39.689097	45.763385	51.837673
24	0.024	39.523421	45.726976	51.930532
25	0.025	39.526412	45.808084	52.089757
26	0.026	39.716517	46.067684	52.418852
27	0.027	40.074323	46.567177	53.060032
28	0.028	40.538169	47.368584	54.198998
29	0.029	41.021312	48.534376	56.047440
30	0.030	41.456547	50.126451	58.796354
31	0.031	41.839757	52.203470	62.567183
32	0.032	42.240691	54.815589	67.390487
33	0.033	42.793029	57.995491	73.197953

RP-04

Rhys Parry

From: Rhys Parry
Sent: Thursday, 8 March 2018 4:40 PM
To: Justin Howes
Subject: Quant Trio validation
Attachments: Quant Trio Issues Report.doc

Hey Justin

As requested.

Thanks

Rhys Parry
Reporting Scientist

Forensic & Scientific Services,
Health Support Queensland, **Department of Health**

p | [redacted]
a | 39 Kessels Rd, Coopers Plains, QLD 4138
w Queensland Health e | [redacted]

HSQ's vision | Delivering the best health support services and solutions for a safer and healthier Queensland.

Queensland Health acknowledges the Traditional Owners of the land, and pays respect to Elders past, present and future.

Issues with the Quant Trio Validation

- This is by no means an exhaustive list, but reflects obvious errors from a relatively quick perusal of the sensitivity and reproducibility / repeatability sections of the report and the raw data.
    - The process consistently overestimates the quant value for single source DNA compared to the known DNA content (Table 6 & 7) with no explanation given as to why.
    - The process consistently overestimates the quant value for mixed source DNA compared to the known DNA content (Table 6 & 7) with no explanation given as to why.
    - The general overestimation of the quant is a likely reason why many reportedly high quant samples yield low rfu profiles. This should have been explored at the time of validation.
    - The incorrect type of t-test has been used throughout the experiment. Almost all the uses pertain to paired sample analysis whereas a t-test assuming different variances has been used.
    - Using the correct t-test, there are three groups that are significantly different between plates A & B from table 14 at the $p \leq 0.05$ level and 1 group that is borderline at $p=0.059$.
    - Using the correct t-test, there are two groups that are significantly different between plates A & C from table 16 at the $p \leq 0.05$ level and 1 group that is borderline at $p=0.057$.
    - Using the correct t-test, there are two groups that are significantly different between plates C & B from table 16 at the $p \leq 0.05$ level.
    - As stated numerous times in the past, it is inappropriate to compare more than two groups using t-tests as it leads to an increased potential for a Type 1 error to occur (falsely rejecting the null hypothesis when in fact it is true). This is exactly what occurs in the experiment 4 data when using the correct t-test. The correct analysis is to use a one-way ANOVA.
    - When testing a machine run or a process the sample unit is the machine run or the process itself. Thus, this study has only four samples which means it is a poor reflection of population of process runs it is trying to estimate. Having large numbers of repeated samples on a single run is meaningless (other than as an indicator of sample preparation variation) and is an example of pseudo-replication. As such, sections 4a & 4b do not meet the generally accepted five repeats and five reproductions that are recommended by groups such as ENFSI.
    - There is no exploration of the variation in the data. Quoting means without also quoting the SE or the SD is meaningless as it indicates little about the estimated population distribution. For example, the mean of 51 & 49 is the same as the mean of 0.5 and 99.5, but the first example is far more preferable when investigating machine accuracy.
    - Page 47 states that:

      "…the low t-test score at 0.01ng/uL is due to the low accuracy and the high variability at that DNA concentration level, therefore the t-test score of 0.00787 ($p \geq 0.05$) is not unexpected." A similar statement is found on p45-46.

      This sentence contains some errors.

        1. A t-test generates a probability 'p' (not a score) that is an estimate of the probability that any differences between groups are not due to chance. Thus, a $p=0.01$ implies that there is a 1% chance that any differences between the means are due to chance.
        2. Therefore, from above it follows that "low accuracy and high variability" would have exactly the opposite effect, as it would

mean there would be less chance the groups would be sufficiently separated to be significant.

3. P needs to be smaller than the decided threshold (so $p \geq 0.05$ should be $p \leq 0.05$).

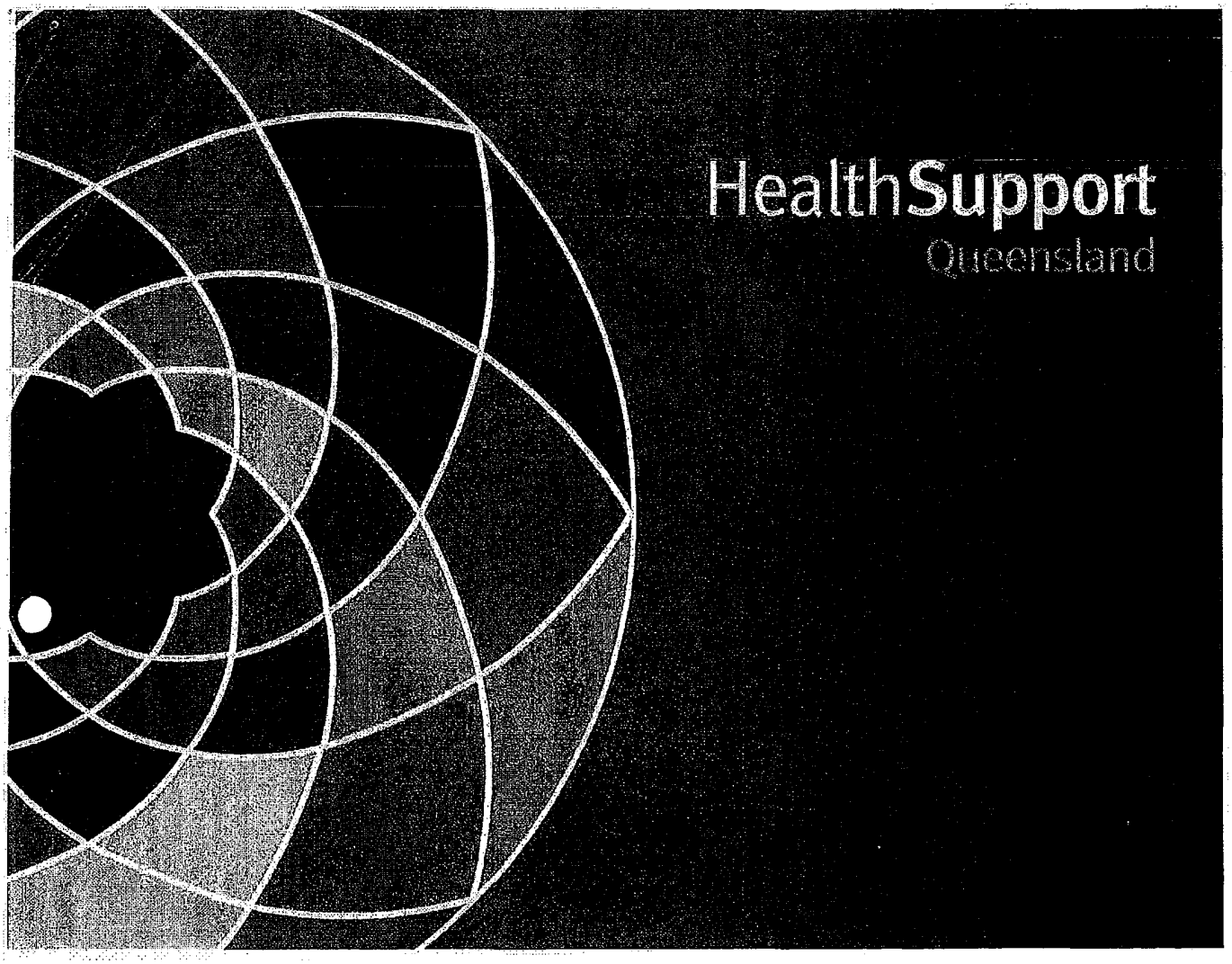
Risks (not specific to the Quant Trio validation only)

While the risk of the below occurring is very low, the potential effects of an occurrence is potentially extremely damaging.

- Defence asking for copies of validation studies (which has happened in the past) and seeking expert advice on the results.
- The rejection of DNA evidence due to inappropriately validated/ verified equipment
- Potentially having to rework hundreds or thousands of samples because of successful defence challenges.
- Losing scientific respect nationally by other DNA labs
- Losing the confidence and respect of the community because any successful defence challenge will be in the public arena
- Having to ride an ongoing wave of further defence challenges and corresponding s95 reports as the lab's underlying science will be viewed as weak.

Financial Costs (not specific to the Quant Trio validation only)

- With design improvements, the experiment could have been done to a much greater degree of scientific validity with minimal extra cost.
- The low quality of the validations / verifications means that the lab has a poor understanding of the variation expected from various pieces of equipment. This potentially leads to unnecessary re-amplification and ReGS in order to obtain consistent EPG results especially given that the Quant trio system is prone to over-estimating the quant, which could lead to under-amplification.
- There is a potential major cost of having to redo thousands of samples deemed inadequate due to insufficient validation quality in the event of a successful defence challenge.



Project Proposal #185

Validation of two QuantStudio™ 5 Real-Time PCR Systems

June 2017

Luke Ryan, Megan Mathieson and Cathie Allen

Project Proposal #185 Validation of two QuantStudio 5 Real-Time PCR Systems

Published by the State of Queensland (Queensland Health), June 2017

For more information contact:
Forensic DNA Analysis, Forensic and Scientific Services, Department of Health, GPO Box 48, Brisbane QLD 4001.

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: Luke Ryan
Title: Senior Scientist – Analytical
Phone: [redacted]
Email: [redacted]

Version history

Version	Date	Changed by	Description
1.0	10/05/2017	Luke Ryan	Document Created.
2.0	13/06/2017	Luke Ryan	Management Team Feedback

Document sign off

This document has been **approved** by:

Name	Position	Signature	Date
Cathie Allen	Managing Scientist	[redacted]	15/06/2017

The following officers have **endorsed** this document

Name	Position	Signature	Date
Justin Howes	Team Leader FRIT	[redacted]	13-06-2017

Name	Position	Signature	Date
Paula Brisotto	Team Leader ER & Q	[redacted]	15-06-2017

Name	Position	Signature	Date
Luke Ryan	Senior Scientist Analytical	[redacted]	13-06-2017

Name	Position	Signature	Date
Allan McNevin	Senior Scientist ER	[redacted]	15-06-2017

Name	Position	Signature	Date
Kirsten Scott	Senior Scientist Q & P	[redacted]	15/06/2017

Name	Position	Signature	Date
Sharon Johnstone	Senior Scientist Intel	[redacted]	15/06/2017

Name	Position	Signature	Date
Matthew Hunt	A/Senior Scientist Reporting 1	[redacted]	15/06/2017

Name	Position	Signature	Date
Kylie Rika	Senior Scientist Reporting 2	[redacted]	15/06/2017

Contents/

1. Purpose and Scope

1.1. Background

Forensic DNA Analysis has two 7500 Real-Time PCR instruments (7500s) which are used to analyse Quantifiler® Trio DNA quantification reactions. Both 7500s are at end of life and are being replaced under the Health Technology Equipment Replacement Program (HTER). The HTER process identified the QuantStudio™ 5 Real-Time PCR System (QS5) as the most suitable replacement for the 7500s. Two QS5s have been purchased.

Both QS5s will be validated for the analysis of Quantifiler® Trio DNA quantification reactions. The QS5s will be delivered with pre-installed protocols for the Quantifiler® Trio kit.

Validation and implementation of the two QS5s will be staggered. QS5-A will be validated first, whilst maintaining one 7500 in operation for routine processing. Once QS5-A has been validated and implemented the remaining 7500 will be removed from use and QS5-B will commence validation. The validation experiments for both QS5s will be the same.

1.2. Purpose

The purpose of this project is to validate both QS5-A and QS5-B for the analysis of Quantifiler® Trio DNA quantification reactions.

1.3. Scope

The QS5s will be validated only for Quantifiler® Trio DNA quantification reactions. The QS5s will be validated for casework and reference samples.

Both QS5-A and QS5-B will be validated separately as per this experimental design.

Quantifiler® Trio reaction setup will not be modified in this project, and will be as per QIS# 33407 Quantification of Extracted DNA using the Quantifiler® Trio DNA Quantification Kit.

The following experiments will be performed:

1. Sensitivity and Limit of Detection
2. Comparison of QS5 and 7500
3. Repeatability and Reproducibility
4. Y Intercept Thresholds

2. Governance

Project Personnel

- Project Manager: Luke Ryan – Senior Scientist, Analytical Team
- Senior Project Officer: Megan Mathieson, Senior Scientist, Analytical Team

Decision Making Group

- The Management Team and the Senior Project Officer are the decision making group for this project and may use the defined acceptance criteria in this project to cease part or all of the experimentation at any stage. The Decision Making Group may also make modifications to this Experimental Design as required, however this must be documented and retained with the original approved Experimental Design.
- The Senior Project Officer is included in the Decision Making Group in their capacity as an expert user.

Reporting

The Project Manager will provide a weekly project status update to the Team Leader, Evidence Recovery and Quality who will advise the Decision Making Group at the Management Team meetings and by exception as required.

3. Resources

The following resources are required for this validation/project:

3.1. Reagents

NIST Standard.

- 0.5% v/v Bleach White N Bright (Ecolab, NSW, AU)
- 5% v/v Trigene Advance (CEVA DEIVET Pty. Ltd. Seven Hills, NSW, AU)
- Ethanol (Recochem Incorporated, Wynnum, QLD, AU)
- TE-4 (Forensic DNA Analysis, Brisbane, QLD, AU)
- Quantifiler® Trio DNA Quantification kit (Life Technologies Applied Biosystems, Foster City, CA, US)

3.2. Materials

- MicroAmp® Optical 96-Well Reaction Plate with Barcode (Applied Biosystems by Life Technologies, Foster City, CA, USA)
- MicroAmp® Optical Adhesive Film (Life Technologies Applied Biosystems, Foster City, CA, US)
- Sterile 1.5 and 2 mL screw-cap tubes (SSI – Interpath, Heidelberg West VIC, AUS)

- Sterile 5 mL screw-cap tubes (Axygen Scientific Inc., Union City, CA, US)
- ART Filtered 1000 µL, 300 µL & 20p pipette tips (Molecular BioProducts Inc., San Diego, CA, US)
- F1-ClipTip pipette tips - 20µL, 50µL, 200µL & 1000 µL (Thermo Fisher Scientific Inc,)
- Combitips advanced® 0.5mL (Eppendorf Biopur, Hamburg, DE)
- Nunc™ Bank-It™ tubes and Caps (Nunc A/S DK-4000 Roskilde, Denmark)
- Rediwipes (Cello Paper Pty. Ltd., Fairfield, NSW, AU)
- Hamilton Conductive 50µL Filter Tips in Frames (Hamilton, Reno, NV, USA)
- Hamilton Conductive 300µL Filter Tips in Frames (Hamilton, Reno, NV, USA)

3.3. Equipment

- ID STARlet Automated Liquid Handler (Hamilton, Reno, NV, USA)
- QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific, Foster City, CA, US)
- Biological safety cabinets class II (ESCO, Lytton, QLD, AU)
- AB 7500 Real Time PCR System (Thermo Fisher Scientific, Foster City, CA, US)
- LaboGene Scanspeed 1248 Centrifuge (Labgear, Lynge, Denmark)
- Vortex Mixer VM1 (Ratek Instruments Pty Ltd, Melbourne, VIC, AU)
- MixMate (Eppendorf AG, Hamburg, DE)
- Micro centrifuge (Tomy, Tokyo, JP)
- Eppendorf 5424 centrifuge and Eppendorf 5804 centrifuge (Eppendorf, North Ryde, NSW, Australia)
- Milli-Q® Integral 3 (A10) System with Q-POD™ (Millipore™, Billerica, MA, USA)
- Pipettes (Eppendorf, Hamburg, DE and Thermo Fisher Scientific (Finnpipette), Waltham, MA, US)
- ClipTip Pipettes (Thermoscientific)
- Multi-step advanced® 0.5mL (Eppendorf Biopur, Hamburg, DE)

Forensic DNA Analysis Analytical Staff, Computer and instrument time, as well as bench space in Forensic DNA Analysis Analytical Laboratory will also be used for the duration of this project.

4. Methods

4.1. NIST Standard Creation

Basic concentrations

NIST standards will be used for this validation. NIST Standard sets A, B and C will be used to create serial dilutions using TE-4 buffer with final concentrations as per Table 1 below.

Once created, the serial dilutions of NIST A, B and C will be quantified in duplicate using the 7500.

Table 1: Serial Dilution NIST Standards

Sample Number	DNA Concentration (ng/µL)
1	5.0
2	1.0
3	0.5
4	0.1
5	0.09
6	0.07
7	0.05
8	0.03
9	0.01
10	0.009
11	0.008
12	0.007
13	0.006
14	0.005
15	0.004
16	0.003
17	0.002
18	0.001
19	0.0001

4.2. DNA Quantificaiton

Quantification using Quantifiler® Trio will be prepared using the ID STARlet according to QIS# 33407 Quantification of Extracted DNA using the Quantifiler® Trio DNA Quantification Kit.

Quantification using the 7500 will be performed as per QIS# 33407 Quantification of Extracted DNA using the Quantifiler® Trio DNA Quantification Kit.

Quantification using the QS5 will be as per the pre-installed scripts.

5. Experimental Design

5.1. Experiment 1: Sensitivity and Limit of Detection

Intent

Quantifiler® Trio has been shown to have a single source sensitivity down to concentrations of 5 pg/µL[1]. The validation of Quantifiler® Trio on the 7500s determined the Limit of Detection (LOD) to be 0.001 ng/µL[2]. Serial dilutions of NIST standards will be used to determine the LOD for Quantifiler® Trio on the QS5 instruments.

Experimental Design

NIST Standards A, B, and C, are derived from a single male donor, multiple female donors, and multiple male and female donors, respectively[3]. NIST standards A, B and C will be used to determine the LOD for Short Amplicon Target (SAT). NIST A only will be used to determine the LOD for the Y Target.

Serial dilutions of each NIST Standard (A, B and C) will be prepared using TE buffer for all samples as per Table 1 (Section 4.1).

Each serial dilution (1-19) of each NIST Standard (A, B and C) will be quantified in duplicate using Quantifiler® Trio and analysed on a QS5. Plates will be prepared according to Tables 2 and 3 below.

Table 2: NIST Standards Serial Dilutions – Platemap 1 of 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	STD 1 50 ng/µL	STD 5 0.005 ng/µL	NIST C 5.0 ng/µL	NIST B 0.5 ng/µL	NIST A 0.1 ng/µL	NIST C 0.09 ng/µL	NIST B 0.05 ng/µL	NIST A 0.03 ng/µL	NIST C 0.01 ng/µL	NIST B 0.008 ng/µL	NIST A 0.007 ng/µL	NIST C 0.006 ng/µL
B	STD 1 50 ng/µL	STD 5 0.005 ng/µL	NIST A 1.0 ng/µL	NIST C 0.5 ng/µL	NIST B 0.1 ng/µL	NIST A 0.07 ng/µL	NIST C 0.05 ng/µL	NIST B 0.03 ng/µL	NIST A 0.009 ng/µL	NIST C 0.008 ng/µL	NIST B 0.007 ng/µL	NIST A 0.005 ng/µL
C	STD 2 5.000 ng/µL	Reagent Blank	NIST B 1.0 ng/µL	NIST A 0.5 ng/µL	NIST C 0.1 ng/µL	NIST B 0.07 ng/µL	NIST A 0.05 ng/µL	NIST C 0.03 ng/µL	NIST B 0.009 ng/µL	NIST A 0.008 ng/µL	NIST C 0.007 ng/µL	NIST B 0.005 ng/µL
D	STD 2 5.000 ng/µL	NIST A 5.0 ng/µL	NIST C 1.0 ng/µL	NIST B 0.5 ng/µL	NIST A 0.09 ng/µL	NIST C 0.07 ng/µL	NIST B 0.05 ng/µL	NIST A 0.01 ng/µL	NIST C 0.009 ng/µL	NIST B 0.008 ng/µL	NIST A 0.006 ng/µL	NIST C 0.005 ng/µL
E	STD 3 0.500 ng/µL	NIST B 5.0 ng/µL	NIST A 1.0 ng/µL	NIST C 0.5 ng/µL	NIST B 0.09 ng/µL	NIST A 0.07 ng/µL	NIST C 0.05 ng/µL	NIST B 0.01 ng/µL	NIST A 0.009 ng/µL	NIST C 0.008 ng/µL	NIST B 0.006 ng/µL	NIST A 0.005 ng/µL
F	STD 3 0.500 ng/µL	NIST C 5.0 ng/µL	NIST B 1.0 ng/µL	NIST A 0.1 ng/µL	NIST C 0.09 ng/µL	NIST B 0.07 ng/µL	NIST A 0.03 ng/µL	NIST C 0.01 ng/µL	NIST B 0.009 ng/µL	NIST A 0.007 ng/µL	NIST C 0.006 ng/µL	NIST B 0.005 ng/µL
G	STD 4 0.050 ng/µL	NIST A 5.0 ng/µL	NIST C 1.0 ng/µL	NIST B 0.1 ng/µL	NIST A 0.09 ng/µL	NIST C 0.07 ng/µL	NIST B 0.03 ng/µL	NIST A 0.01 ng/µL	NIST C 0.009 ng/µL	NIST B 0.007 ng/µL	NIST A 0.006 ng/µL	NIST C 0.005 ng/µL
H	STD 4 0.050 ng/µL	NIST B 5.0 ng/µL	NIST A 0.5 ng/µL	NIST C 0.1 ng/µL	NIST B 0.09 ng/µL	NIST A 0.05 ng/µL	NIST C 0.03 ng/µL	NIST B 0.01 ng/µL	NIST A 0.008 ng/µL	NIST C 0.007 ng/µL	NIST B 0.006 ng/µL	Reagent Blank

Table 3: NIST Standards Serial Dilutions – Platemap 2 of 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	STD 1 50 ng/µL	STD 5 0.005 ng/µL	NIST C 0.004 ng/µL	NIST B 0.002 ng/µL	NIST A 0.001 ng/µL	NIST C 0.0001 ng/µL						
B	STD 1 50 ng/µL	STD 5 0.005 ng/µL	NIST A 0.003 ng/µL	NIST C 0.002 ng/µL	NIST B 0.001 ng/µL							
C	STD 2 5.000 ng/µL	Reagent Blank	NIST B 0.003 ng/µL	NIST A 0.002 ng/µL	NIST C 0.001 ng/µL							
D	STD 2 5.000 ng/µL	NIST A 0.004 ng/µL	NIST C 0.003 ng/µL	NIST B 0.002 ng/µL	NIST A 0.0001 ng/µL							
E	STD 3 0.500 ng/µL	NIST B 0.004 ng/µL	NIST A 0.003 ng/µL	NIST C 0.002 ng/µL	NIST B 0.0001 ng/µL							
F	STD 3 0.500 ng/µL	NIST C 0.004 ng/µL	NIST B 0.003 ng/µL	NIST A 0.001 ng/µL	NIST C 0.0001 ng/µL							
G	STD 4 0.050 ng/µL	NIST A 0.004 ng/µL	NIST C 0.003 ng/µL	NIST B 0.001 ng/µL	NIST A 0.0001 ng/µL							
H	STD 4 0.050 ng/µL	NIST B 0.004 ng/µL	NIST A 0.002 ng/µL	NIST C 0.001 ng/µL	NIST B 0.0001 ng/µL							

Data Analysis

Combined results from NIST A, B and C will be used to determine the LOD for the SAT and LAT. Results from NIST A only will be used to determine the LOD for the Y Target.

LOD will be determined based on the lowest expected concentration at which the observed DNA concentration is reliably detected across the majority of samples in the data set.

→ determined how?

Acceptance Criteria

The LOD for Quantifiler® Trio on the QS5 must be as good as or better than the sensitivity for Quantifiler® Trio on the 7500.

– measured how?

5.2. Experiment 2: Comparison of QS5 and 7500

Intent
To compare the QS5 and 7500 quantification results of NIST A, B and C serial dilutions.

Experimental Design
The 7500 and QS5 quantification results for the NIST A, B and C serial dilutions will be compared.

The 7500 data set will be drawn from the quantification results generated in Section 4.1. For each sample in the NIST A, B and C serial dilutions, the observed quantification result (from the 7500) will be compared to the expected concentration and percentage change calculated.

The QS5 data set will be drawn from the quantification results generated in Experiment 1. For each sample in the NIST A, B and C serial dilutions, the observed quantification result (from the QS5) will be compared to the expected concentration and percentage change calculated.

Data Analysis
The percentage change (expected vs observed) for the 7500 and QS5 will be compared. The instrument with the lowest percentage change will be the most accurate.

Assessment Criteria
The instrument with the lowest percentage change (observed vs expected quantification results) will be assessed as the most accurate. The QS5 will pass this experiment if it is more accurate than the 7500.

5.3. Experiment 3: Repeatability and Reproducibility

5.3.1. Experiment 3a: Repeatability

Intent
To assess repeatability for Quantifiler® Trio analysed on the QS5. Repeatability is an assessment of the whether the QS5 produces the same results when one sample set is processed in duplicate by one user, under the same conditions.

Experimental Design
Plates 1 and 2 from the Sensitivity and LOD experiment will be used for this experiment.

Data Analysis

Repeatability will be assessed by comparing the quantification results for each duplicate pair on the Plates 1 and 2 from the Sensitivity and LOD experiment. Results will be compared using percentage change.

Acceptance Criteria
QS5 will be assessed as acceptable if the results are as good or better than the results from the original Quantifiler® Trio validation[2].

5.3.2. Experiment 3b: Reproducibility

Intent
To assess reproducibility for Quantifiler® Trio analysed on the QS5. Reproducibility is an assessment of the whether QS5 produces the same results when one sample set is processed by different operators under different conditions.

Experimental Design
A second preparation of Plate 1 from the Sensitivity and LOD experiment will be prepared and analysed on the QS5. A second operator (different from the operator who prepared the plates in Experiment 1) will prepare these plates. The plate for this experiment will also be prepared on a different day.

Data Analysis
Reproducibility will be assessed by comparing the quantification results for each sample on for the Plate 1 (Operator 1 Day 1) and Plate 1 (Operator 2 Day 2). Results will be compared using percentage change.

Acceptance Criteria
QS5 will be assessed as acceptable if the results are as good or better than the results from the original Quantifiler® Trio validation[2].

5.4. Experiment 4: Y-Intercept Thresholds

Intent
Y-Intercept thresholds for the SAT, LAT and Y-Targets will be determined. The thresholds will be used for implementation of the QS5s with Quantifiler® Trio.

Experimental Design
Y-Intercept data from all plates run on the QS5 in this project will be used to calculate Y-Intercept thresholds for the SAT, LAT and Y-Target.

Thresholds calculated from project data will be used as implementation thresholds. Given that this data set is small, all runs post implementation will be added to the data set and the thresholds revised at least every 2 weeks for the first 3 months after implementation.

Data Analysis
Y-Intercept Thresholds will be calculated using: Average +/- 3 Standard Deviations.

? + only

Acceptance Criteria
The Y-Intercept thresholds for each target are instrument and kit specific, and are used to monitor performance over time. Therefore no acceptance criteria will be set.

But there must be some criteria. How else can we judge future performance?

6. Results and Data Compilation

The acceptance/assessment criteria for each experiment will be used to make an overall assessment as to whether the QS5s have been validated for analysis of Quantifiler® Trio DNA quantification assays.

If the Project Team forms the opinion that additional experiments are required before a final assessment can be made, application will be made to the Decision Making Group for a modification to this Experimental Design. The Decision Making Group is responsible for assessing this application and approving or rejecting it.

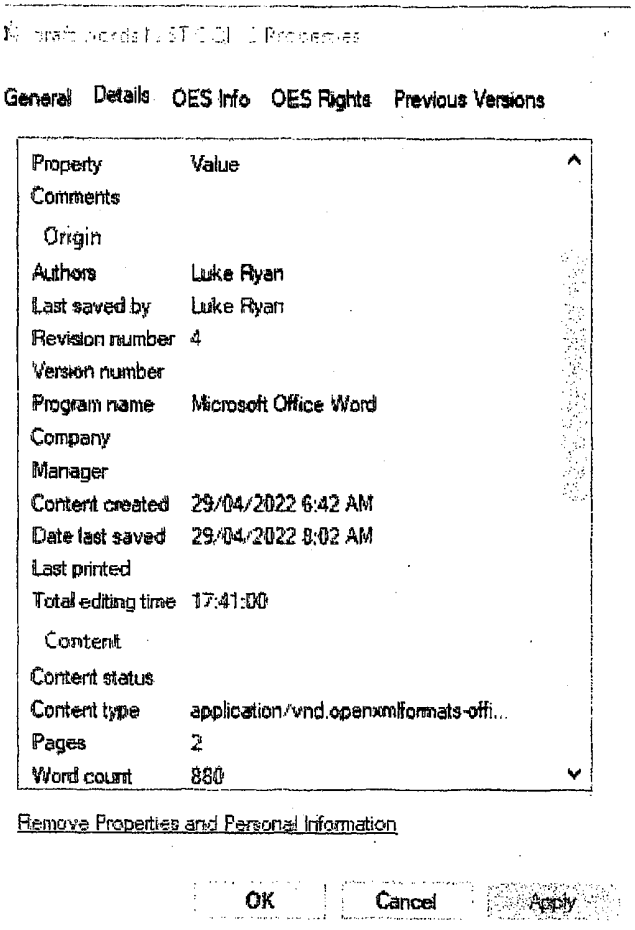
A final report will be produced which will compile all analyses, conclusion and recommendations. The final report will be prepared by the Project Group.

7. References

[1] Thermo Fisher Scientific, Quantifiler® HP and Trio DNA Quantification Kits User Guide, Publication Number 4485354, Revision A. Publication Number 4485354, Revision A ed2014.

[2] Validation of Quantifiler® Trio. P. Acedo, M. Mathieson, L. Ryan, C. Allen. September 2015. Forensic DNA Analysis.

[3] Certificate of Analysis – Standard Reference Material® 2372 Human DNA Quantitation Standard. National Institute of Standards & Technology.


Properties
General
Details
OES Info
OES Rights
Previous Versions
Property
Value
Comments
Origin
Authors
Luke Ryan
Last saved by
Luke Ryan
Revision number
4
Version number
Program name
Microsoft Office Word
Company
Manager
Content created
29/04/2022 6:42 AM
Date last saved
29/04/2022 8:02 AM
Last printed
Total editing time
17:41:00
Content
Content status
Content type
application/vnd.openxmlformats-offi...
Pages
2
Word count
880
Remove Properties and Personal Information
OK
Cancel
Apply

RP-07

In 2017, Forensic DNA Analysis purchased two QuantStudio™ 5 Real-Time PCR Systems (QS5) to replace the 7500 Real-Time PCR Systems instruments which were at end of life. The 7500 instruments were being used with the Quantifiler Trio DNA quantification kit to estimate the DNA concentration of samples. The QS5s were purchased to be used with the same Quantifiler Trio kit.

Project #185 "Validation of two QuantStudio™ 5 Real-Time PCR Systems" was conducted to validate DNA quantification using the Quantifiler Trio kit on the QS5 instruments. Project #185 was conducted in the first quarter of 2018 (the majority of labwork conducted in March 2018). Project documentation is stored in I:\Change Management\Proposal#151 to #200 (completed)\Proposal#185 - Validation of QS5. The purpose of this validation was to compare the 7500 and QS5 to assess whether the QS5 performs the same as or better than the 7500 and therefore is a suitable replacement. The validation was not a validation of the Quantifiler Trio kit.

The NIST SRM 2372 is a human DNA quantification standard, which includes three component genomic DNA materials labelled A, B and C. The NIST SRM 2372 was issued with a Certificate of Analysis on 08/01/2013, which provides apparent absorbance values (i.e. DNA concentration) for NIST SRM 2372 components A, B and C within specified uncertainty. This certificate was valid until expiry on 31 December 2017 after which the relative absorbance values of components A, B and C are not guaranteed.

Project #185 used the NIST SRM 2372 for Experiments 1: Sensitivity, Limit of Detection and Inaccuracy and Experiment 2: Comparison of QS5s and 7500. These experiments used the NIST SRM 2372 after the certificate of analysis had expired.

When reviewing this event the following considerations were noted:

- The NIST SRM 2372 was issued on 08/01/2013 and was therefore viable for an extended period up to the expiration date 31/12/2017. Given this extended certification period, it is not expected the NIST SRM 2372 would experience significant degradation or reduction in concentration in the 3-4 months after the certification expiry.
- Preparation of a serial dilution introduces variation at each serial dilution step due to pipetting error (up to 10% for less than 10 µL and up to 5% for greater than 10 µL), which is compounded with each successive step. Therefore it is expected there will be inaccuracy in the individual serial dilutions.
- The DNA quantification step uses real time PCR which has run to run variation.
- Quantification is an estimation of the DNA concentration only and has been shown to have variation (+/- 30%) in successive internal validations.
- This validation was primarily a comparative study to determine whether the QS5 was a suitable replacement for the 7500.

In Experiment 1 NIST SRM 2372 A, B and C were used in serial dilution to compare the 7500 and QS5 in terms of sensitivity, limit of detection and inaccuracy. Duplicate serial dilutions of NIST SRM 2372 A, B and C were prepared and run on 7500-A, QS5-A and QS5-B.

Percentage inaccuracy compared the QS5 and 7500 when estimating the DNA concentration for each sample in the serial dilution sample set (A, B and C standards). As such, the accuracy of the concentration of each sample in the serial dilution (and therefore the starting concentration of the NIST SRM 2372) is not the critical element of this experiment. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments across the range of concentrations in the serial dilution. Therefore the

use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of this experiment demonstrated comparable performance between the 7500-A, QS5-A and QS5-B.

The Limit of Detection (LOD) of 0.001 ng/µL threshold was determined and set in the PowerPlex®21 PCR amplification kit validation based on the DNA concentration required to reliably obtain reportable DNA profiles. The QS5 validation LOD experiment was intended to compare the performance of the QS5 and 7500 when analysing samples with concentrations above and below the LOD. The NIST SRM 2372 serial dilutions used in Experiment 1 were used again for this experiment.

As with the percentage inaccuracy experiment, the accuracy of the concentration of each serial dilution is not the critical element of this experiment as the QS5/7500 were assessed at several concentrations above and one concentration below the LOD. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments at these reducing concentrations. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of this experiment showed comparable performance between the 7500-A, QS5-A and QS5-B and recommended the LOD remain at 0.001 ng/µL.

Experiment 2 was a statistical comparison of the performance of the 7500-A, QS5-A and QS5-B using the results of the NIST SRM 2372 serial dilutions. Because this was a comparison the accuracy of the concentration of each serial dilution is not the critical element of this experiment. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of the statistical analysis demonstrated there was no significant difference between the 7500-A, QS5-A and QS5-B.

This assessment of the use of the NIST SRM 2372 in Project #185 after the certificate of analysis had expired has shown that it the original assessments and conclusions made in Project #185 are valid and the QS5 is appropriate for use.

Report for QIS OQI as of 25/09/2022 11:58:40 AM

Report for QIS OQI -

56218 Use of NIST standard in Project#185

OQI Details

Status	Follow-Up
Subject	During writing of Project#206 project plan update, which will incorporate NIST 2372a, it was observed that the quantification values of the NIST standard used in Project#185 correspond to an old NIST standard.
Source of OQI	Internal Problem
Date Identified	20/04/2022

OQI Creator Contact Details

Creator	Thomas NURTHEN
Organisational Unit/s	Reporting 2
Service/s	Forensic and Scientific Service
Site Location/s	Coopers Plains

Investigator/Actioner Contact Details

Actioner	Luke RYAN
Organisational Unit/s	Analytical
Service/s	Forensic and Scientific Service
Site Location/s	Coopers Plains

Investigation Details

Investigation Completed 03/05/2022 **Root Cause Type** | Documentation

Investigation Details

In 2017, Forensic DNA Analysis purchased two QuantStudio™ 5 Real-Time PCR Systems (QS5) to replace the 7500 Real-Time PCR Systems instruments which were at end of life. The 7500 instruments were being used with the Quantifiler Trio DNA quantification kit to estimate the DNA concentration of samples. The QS5s were purchased to be used with the same Quantifiler Trio kit.

Project #185 "Validation of two QuantStudio™ 5 Real-Time PCR Systems" was conducted to validate DNA quantification using the Quantifiler Trio kit on the QS5 instruments. Project #185 was conducted in the first quarter of 2018 (the majority of labwork conducted in March 2018). Project documentation is stored in I:\Change Management\Proposal#151 to #200 (completed) \Proposal#185 - Validation of QS5. The purpose of this validation was to compare the 7500 and QS5 to assess whether the QS5

performs the same as or better than the 7500 and therefore is a suitable replacement. The validation was not a validation of the Quantifiler Trio kit.

The NIST SRM 2372 is a human DNA quantification standard, which includes three component genomic DNA materials labelled A, B and C. The NIST SRM 2372 was issued with a Certificate of Analysis on 08/01/2013, which provides apparent absorbance values (i.e. DNA concentration) for NIST SRM 2372 components A, B and C within specified uncertainty. This certificate was valid until expiry on 31 December 2017 after which the relative absorbance values of components A, B and C are not guaranteed.

Project #185 used the NIST SRM 2372 for Experiments 1: Sensitivity, Limit of Detection and Inaccuracy and Experiment 2: Comparison of QS5s and 7500. These experiments used the NIST SRM 2372 after the certificate of analysis had expired.

When reviewing this event the following considerations were noted:

- The NIST SRM 2372 was issued on 08/01/2013 and was therefore viable for an extended period up to the expiration date 31/12/2017. Given this extended certification period, it is not expected the NIST SRM 2372 would experience significant degradation or reduction in concentration in the 3-4 months after the certification expiry.

- Preparation of a serial dilution introduces variation at each serial dilution step due to pipetting error (up to 10% for less than 10 µL and up to 5% for greater than 10 µL), which is compounded with each successive step. Therefore it is expected there will be inaccuracy in the individual serial dilutions.

- The DNA quantification step uses real time PCR which has run to run variation.

- Quantification is an estimation of the DNA concentration only and has been shown to have variation (+/- 30%) in successive internal validations.

- This validation was primarily a comparative study to determine whether the QS5 was a suitable replacement for the 7500.

In Experiment 1 NIST SRM 2372 A, B and C were used in serial dilution to compare the 7500 and QS5 in terms of sensitivity, limit of detection and inaccuracy. Duplicate serial dilutions of NIST SRM 2372 A, B and C were prepared and run on 7500-A, QS5-A and QS5-B.

Percentage inaccuracy compared the QS5 and 7500 when estimating the DNA concentration for each sample in the serial

dilution sample set (A, B and C standards). As such, the accuracy of the concentration of each sample in the serial dilution (and therefore the starting concentration of the NIST SRM 2372) is not the critical element of this experiment. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments across the range of concentrations in the serial dilution. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of this experiment demonstrated comparable performance between the 7500-A, QS5 -A and QS5-B.

The Limit of Detection (LOD) of 0.001 ng/µL threshold was determined and set in the PowerPlex®21 PCR amplification kit validation based on the DNA concentration required to reliably obtain reportable DNA profiles. The QS5 validation LOD experiment was intended to compare the performance of the QS5 and 7500 when analysing samples with concentrations above and below the LOD. The NIST SRM 2372 serial dilutions used in Experiment 1 were used again for this experiment.

As with the percentage inaccuracy experiment, the accuracy of the concentration of each serial dilution is not the critical element of this experiment as the QS5/7500 were assessed at several concentrations above and one concentration below the LOD. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments at these reducing concentrations. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of this experiment showed comparable performance between the 7500-A, QS5-A and QS5-B and recommended the LOD remain at 0.001 ng/µL.

Experiment 2 was a statistical comparison of the performance of the 7500-A, QS5-A and QS5-B using the results of the NIST SRM 2372 serial dilutions. Because this was a comparison the accuracy of the concentration of each serial dilution is not the critical element of this experiment. The critical element is the use of the same serial dilution to test each of the 7500-A, QS5-A and QS5-B to enable comparative performance assessments. Therefore the use of the NIST SRM 2372 post-expiry does not affect the validity of this experiment. The results of the statistical analysis demonstrated there was no significant difference between the 7500-A, QS5-A and QS5-B.

This assessment of the use of the NIST SRM 2372 in Project #185 after the certificate of analysis had expired has shown that it the original assessments and conclusions made in Project #185 are valid and the QS5 is appropriate for use.

Preformed By Luke RYAN

Action Details

Action Complete	03/05/2022	**Action Fix Type**	DocumentationOQI Recorded added to
Title	Project #185 folder		
Action Description	A record of this OQI is to be added to the Project #185 folder for future reference.		

Task Details

No Tasks found

Follow-up And Approval

No Follow Up and Approval Information Available for this OQI

Associations

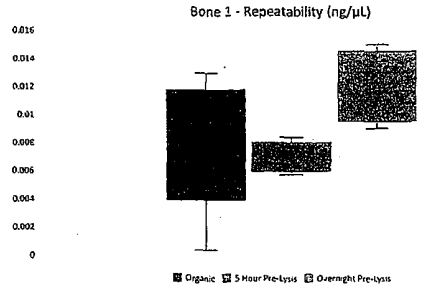
No Associations found

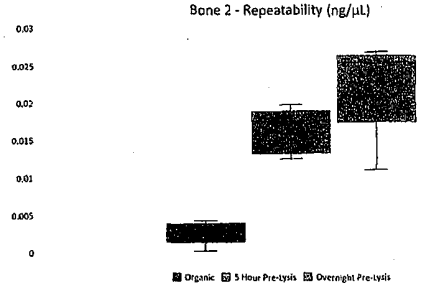
Records

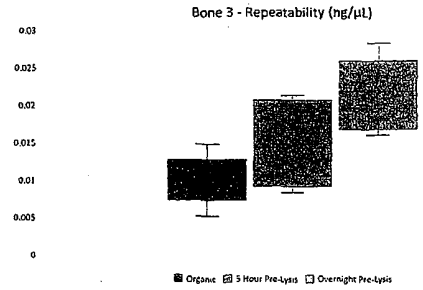
No Records found

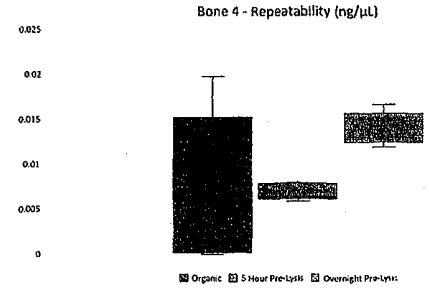
56218 Use of NIST standard in Project#185

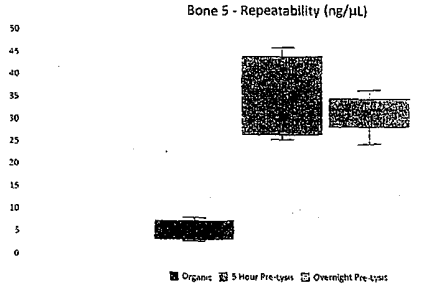
Bone 1	Organic	0.01296	0.00032	0.00747	0.01054	0.00956
	5 Hour Pre-Lysis	0.00568	0.00835	0.00694	0.00619	0.00754
	Overnight Pre-Lysis	0.014	0.015	0.009	0.01	0.013
Bone 2	Organic	0.00361	0.00274	0.00311	0.00038	0.00444
	5 Hour Pre-Lysis	0.01811	0.0177	0.01412	0.01253	0.01984
	Overnight Pre-Lysis	0.01112	0.02394	0.02606	0.02558	0.02697
Bone 3	Organic	0.00965	0.0052	0.01075	0.01485	0.01017
	5 Hour Pre-Lysis	0.00896	0.0084	0.02136	0.01434	0.01998
	Overnight Pre-Lysis	0.02185	0.01757	0.02838	0.02348	0.01603
Bone 4	Organic	0.01979	0.01077	0.00044	0	0.00999
	5 Hour Pre-Lysis	0.00792	0.00643	0.00787	0.00591	0.00643
	Overnight Pre-Lysis	0.0166	0.013	0.01471	0.01193	0.01296
Bone 5	Organic	3.90087	7.82573	2.79068	3.52519	6.41383
	5 Hour Pre-Lysis	25.25375	27.4191	45.80926	41.76097	34.41719
	Overnight Pre-Lysis	24.29558	31.8261	31.8615	32.42724	36.27103











RP-09

reproducibility — Bone 1

	Day 1	Day 2	Day 3	Day 4	Day 5
Organic	0.01296	0.00999	0.0099	0.00974	0.00647
5 Hour Pre	0.00568	0.00922	0.01038	0.01321	0.01056
Overnight	0.01409	0	0.01114	0.01741	0.01084

reproducibility — bone 2

	Day 1	Day 2	Day 3	Day 4	Day 5
Organic	0.00361	0.00196	0.00458	0.00713	0.00692
5 Hour Pre	0.01811	0.00614	0.02598	0.02227	0.01705
Overnight	0.01112	0.00013	0.00521	0.02773	0.01552

reproducibility — bone 3

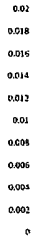
	Day 1	Day 2	Day 3	Day 4	Day 5
Organic	0.00965	0.00953	0.00785	0.00723	0.01027
5 Hour Pre	0.00996	0.1384	0.02495	0.03412	0.03583
Overnight	0.02185	0.00665	0.02976	0.03259	0.00743

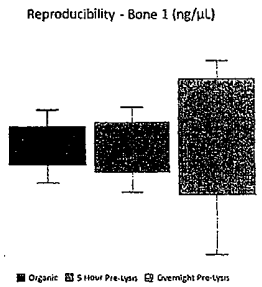
reproducibility — bone 4

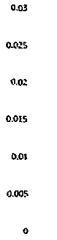
	Day 1	Day 2	Day 3	Day 4	Day 5
Organic	0.01979	0.01093	0.01657	0.01294	0.01324
5 Hour Pre	0.00792	0.00261	0.00632	0.01793	0.01821
Overnight	0.0166	0	0	0.0101	0.00604

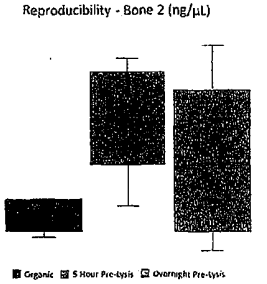
reproducibility — bone 5

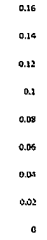
	Day 1	Day 2	Day 3	Day 4	Day 5
Organic	3.90037	69.92214	42.1617	46.41128	41.4768
5 Hour Pre	25.25375	11.38112	41.33673	51.12676	65.29295
Overnight	24.29558	0.11813	44.00444	44.07124	19.29339

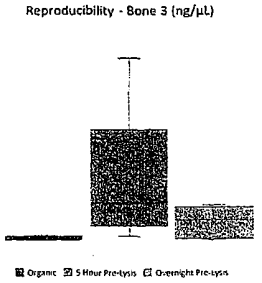


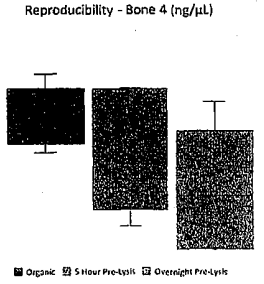


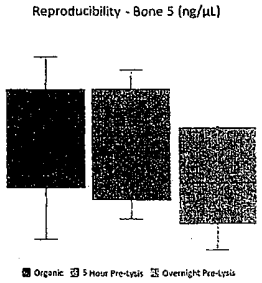










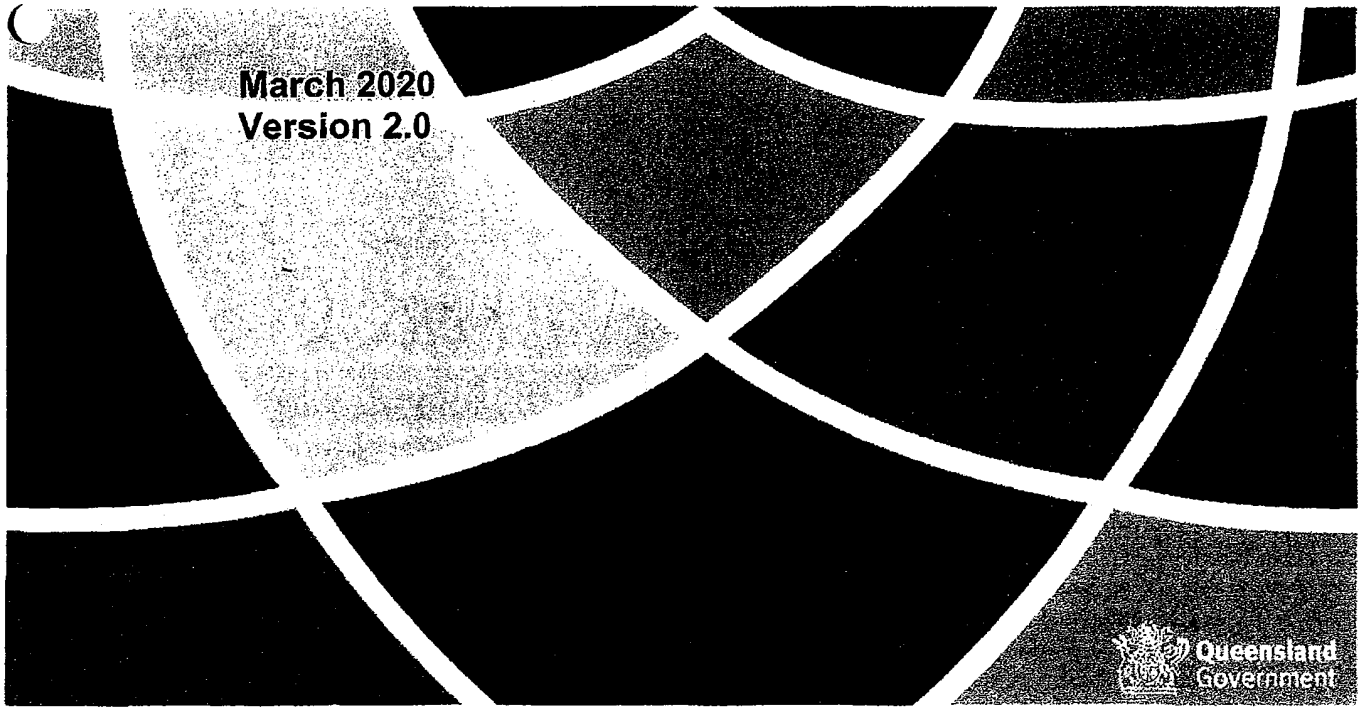


Project Report #192

Validation of QIAsymphony® SP for Bone Extraction

Supplementary Repeatability and Reproducibility

Melissa Cipollone, Luke Ryan, Megan Mathieson and Cathie Allen



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: Luke Ryan
Title: Senior Scientist – Analytical
Phone: [redacted]
Email: [redacted]

Version history

Version	Date	Changed by	Description
1.0	December 2019	Melissa Cipollone	Creation of Document
2.0	March 2020	Melissa Cipollone, Luke Ryan, Megan Mathieson	Document Feedback

Document sign off

This document has been **approved** by:

Name	Position	Signature	Date
Cathie Allen	Managing Scientist	[redacted]	24/03/2020

The following officers have **endorsed** this document

Name	Position	Signature	Date
Justin Howes	Team Leader FRIT	[redacted]	23.03.2020
Paula Brisotto	Team Leader ER & Q		23-03-2020
Luke Ryan	Senior Scientist Analytical		23-03-2020
Allan McNevin	Senior Scientist ER		24.03.2020
Kirsten Scott	Senior Scientist Q & P		23/03/2020
Allison Lloyd	A/Senior Scientist Intel		23/03/2020
Sharon Johnstone	Senior Scientist Reporting 1		23/03/2020
Kylie Rika	Senior Scientist Reporting 2		24/03/2020

Contents

Abstract

Forensic DNA Analysis currently uses an organic extraction for the extraction of DNA from bone and teeth. Phenol chloroform isoamyl alcohol is used in the organic extraction process and is a chemical hazard to the operator. The organic extraction process is time consuming and labour intensive. One organic extraction batch contains a maximum of 12 bone/teeth samples and takes an operator a full day to complete which is relatively inefficient and is the rate limiting step in the processing of bone/teeth samples in the Analytical Team.

Forensic DNA Analysis currently uses the QIAsymphony® SP instrument for automated DNA extraction of a range of substrate and sample types (QIS# 33758), but not for bones/teeth. QIAsymphony® SP DNA extractions can process up to 96 samples per batch, and it is possible for one operator to run up to two full runs of 96 samples in a day. QIAGEN have developed protocols for pre-lysis and on-deck protocols for bones/teeth and other casework samples which have been used as the basis for the protocols to be tested in this validation.

The purpose of this project was to conduct further repeatability and reproducibility experiments for the QIAsymphony® SP bone extraction using both the 5 hour and overnight pre-lysis protocols, and to compare these results to the current organic extraction protocol.

The results obtained from this experiment show the 5 hour and overnight pre-lysis QIAsymphony® extractions are comparable to the current organic extraction with the overnight pre-lysis QIAsymphony® protocol the preferred method for routine processing.

Introduction

Forensic DNA Analysis currently performs automated DNA extractions on a range of sample types and substrates using a QIAGEN® QIAsymphony® SP/AS instrument. The QIAsymphony® SP/AS instrument is a modular automated system which enables the processing of up to 96 samples on a single run. The QIAsymphony® SP module is used for the extraction and purification of DNA from forensic casework and reference samples. It uses pre-programmed optimized protocols and the QIAGEN® cartridge-based magnetic-particle chemistry kit, the QIAsymphony® DNA Investigator Kit.

The original validation of the QIAsymphony® SP/AS did not include bone or teeth extraction. Forensic DNA Analysis currently have two QIAsymphony® SP/AS instruments and the use of these instruments for bone/teeth extraction would be particularly beneficial in the event of a large scale disaster victim identification (DVI), as it will dramatically increase the efficiency and processing capacity of bone/teeth DNA extractions. Furthermore, organic extraction involves the use of phenol chloroform isoamyl alcohol which is a chemical hazard, therefore implementing an alternative protocol would remove this hazard.

Processing bone extractions on the QIAsymphony® SP would also provide benefits and efficiencies to training and maintenance of competency. The low numbers of routinely submitted bones/teeth make initial training, and subsequent maintenance of competency, lengthy and difficult to coordinate. Extraction of bones/teeth on the QIAsymphony® would be included in the standard QIAsymphony® casework training module, and not a separate organic extraction competency as it currently is.

Following the completion of the first validation experiments it was decided additional repeatability and reproducibility experiments were required. The following experiments were performed to test and compare repeatability and reproducibility of three extraction protocols:

- Repeatability Experiment:
  - Current organic extraction
  - QIAGEN pre-lysis with overnight incubation and QIAsymphony® SP extraction
  - QIAGEN pre-lysis with 5 hour incubation and QIAsymphony® SP extraction
- Reproducibility Experiment over 5 days:
  - Current organic Extraction
  - QIAGEN pre-lysis with overnight incubation and QIAsymphony® SP extraction
  - QIAGEN pre-lysis with 5 hour incubation and QIAsymphony® SP extraction

Resources and Methods

All reagents, materials and equipment used in this project were as specified in the approved in-house document Project #192 Validation of QIAsymphony® Bone Extraction - Supplementary R&R. This document will be referred to as the experimental design.

All samples used in this verification were selected, analysed and interpreted as outlined in the experimental design.

Sample Selection

Five powdered bone samples were retained from the Freezer Mill Project #209.

Bone Sample	Laboratory Number
Bone 1	[redacted]
Bone 2	[redacted]
Bone 3	[redacted]
Bone 4	[redacted]
Bone 5	[redacted]

(*Exhibit registered in Auslab)

Table 1: Bone samples used in this Validation

Experiments and Results

Experiment 1 – Repeatability

Purpose

The purpose of the repeatability experiment was to extract human genomic DNA from powdered bone using three different extraction methods and compare the results. The compared methods were:

- The current validated method of extracting DNA from bone and teeth using organic extraction.
- The QIAGEN pre-lysis method with the samples being incubated for 5 hours only and then extracted on the QIAsymphony® SP instrument.
- The QIAGEN pre-lysis method with the samples being incubated overnight and then extracted on the QIAsymphony® SP instrument.

Results

Tabulated results are provided in Appendices 1-5. The repeatability quantification results for bones 1-5 are shown in Figures 1-5. The number of alleles obtained for bones 1 to 5 are shown in Figure 6. It should be noted the allele count for some samples were obtained after a microcon concentration procedure (refer to tabulated results in Appendices 1-5).

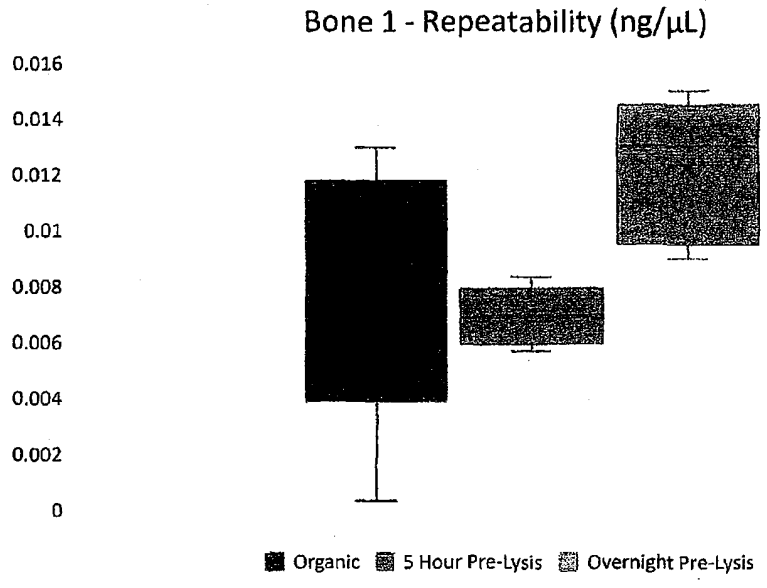


Figure 1: Representation of repeatability data for Bone 1 using Quant Values

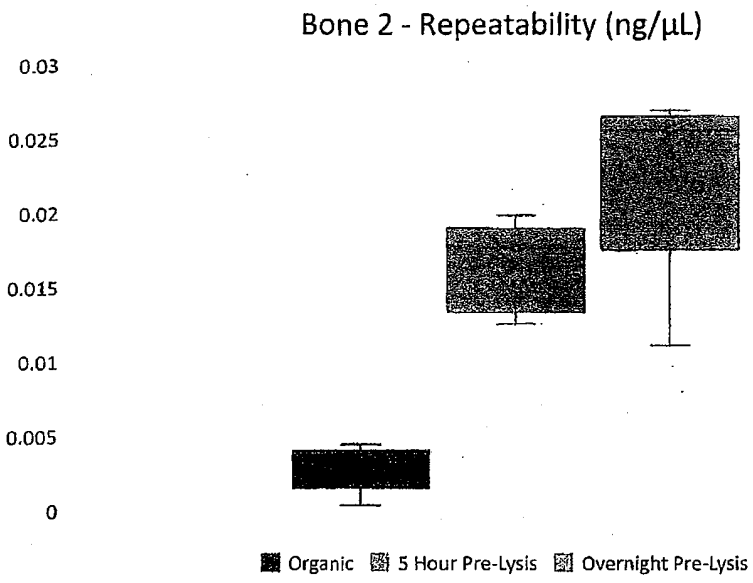


Figure 2: Representation of repeatability data for Bone 2 using Quant Values

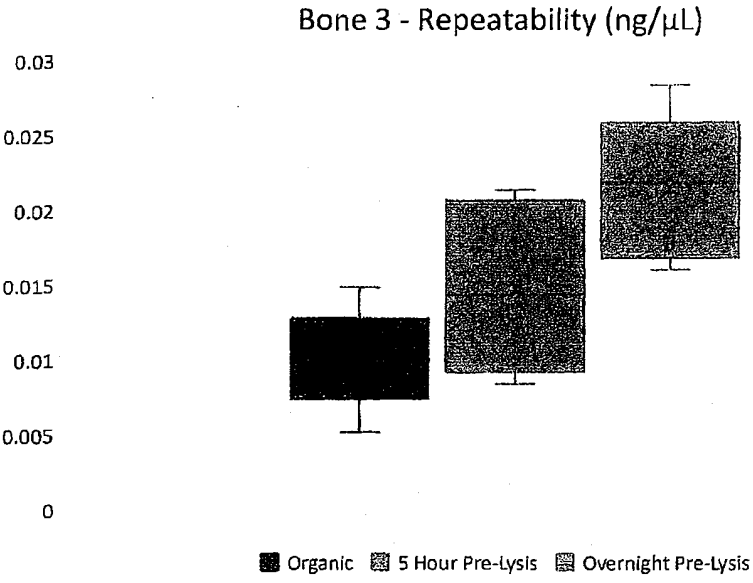


Figure 3: Representation of repeatability data for Bone 3 using Quant Values

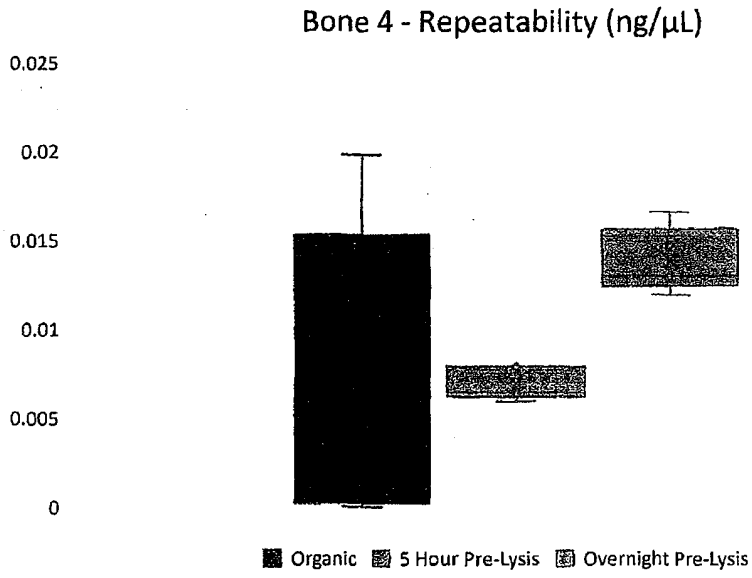


Figure 4: Representation of repeatability data for Bone 4 using Quant Values

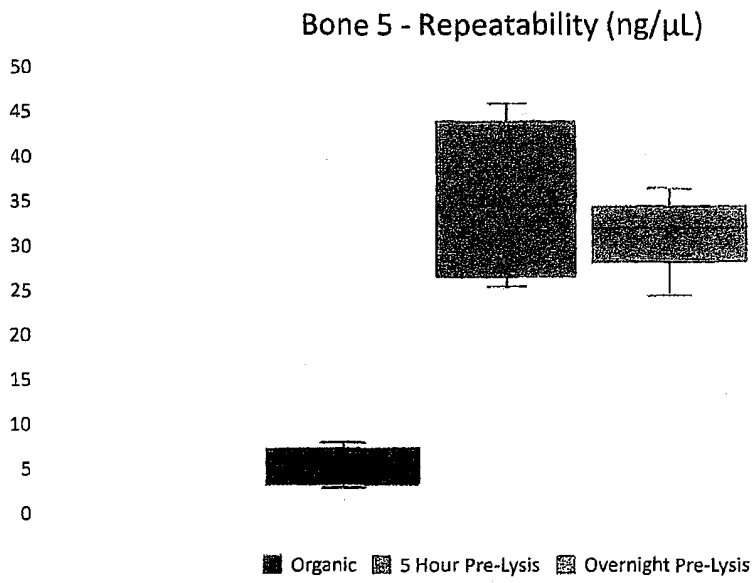


Figure 5: Representation of repeatability data for Bone 5 using Quant Values

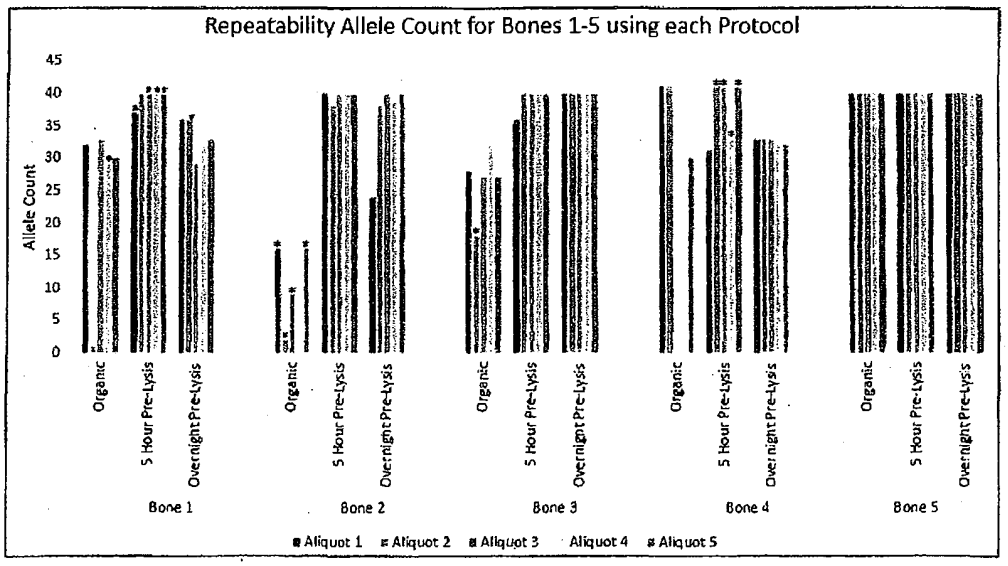


Figure 6: Representation of repeatability data for Bones 1-5 using Allele counts. * indicates samples which have undergone microcon concentration.

Discussion

The results from the current validated method (organic extraction) were used as a benchmark to compare the results from the QIAsymphony® protocols.

Repeatability for each extraction protocol varied between samples with no apparent consistency or trend. The organic and both QIAsymphony® protocols appeared to have a comparable level of repeatability with no one protocol being more or less repeatable consistently across the 5 samples (refer to Figures 1 - 5).

The quantification results for bones 1-4 were lower than bone 5 for all extraction protocols. This is likely due to the quality of the bone samples given the consistency across each of the three extraction protocols.

The overnight pre-lysis QIAsymphony® extraction gave higher quantification results for each of the 5 replicates of bones 2, 3 and 5 (as per Figures 2, 3 and 5) than the organic extraction. For bones 1 and 4, the mean quantification results (across the 5 replicates) were higher for the overnight pre-lysis QIAsymphony® than the organic extraction (as per Figures 1 and 4).

The QIAsymphony® extraction with 5 hour pre-lysis gave higher mean quantification results (across the 5 replicates) than the organic extraction for bones 2, 3 and 5. For bones 1 and 4, although the mean quantification result was lower for the QIAsymphony® extraction, quantification results overall overlapped and were comparable. It should be noted for bones 1 and 4, the range of results for the organic extraction were much wider than the QIAsymphony® extraction with 5 hour pre-lysis, which meant that although some organic replicates gave higher quantification results, some also gave lower quantification results.

Sample extracts quantified in the range 0.001-0.0088 ng/µL underwent microcon concentration prior to amplification to mimic real processing conditions. As stated

previously, bones 1-4 gave low quantification results which resulted in a number of samples undergoing microcon concentration. Across all samples tested, 6 organic extraction samples and 8 QIAsymphony® 5 hour pre-lysis extraction samples underwent microcon concentration. No QIAsymphony® overnight pre-lysis samples underwent microcon concentration (see Appendices 1 – 5). Given the final DNA profile results include samples which have and have not undergone microcon concentration, the final profile and allele count results (refer to Figure 6) have only been used to assess any negative impact the extraction protocols may have had on profile quality. No negative impact on profile quality was noted for any of the extraction protocols.

Overall this repeatability experiment has shown that the organic and both QIAsymphony® protocols are comparable, with the overnight lysis generally giving higher quantification results than the 5 hour lysis. This fits with intuitive expectations as increased reaction time could be expected to give higher yields.

Experiment 2 - Reproducibility

Purpose

The purpose of the reproducibility experiment is to test the reproducibility of results from each extraction protocol when performed by five independent scientists. One aliquot from each sample was tested per protocol for the reproducibility experiments (75 aliquots in total not including controls).

The compared methods were done over a 5 day period by 5 different operators:

- Current organic extraction
- The QIAGEN pre-lysis method with the samples being incubated for 5 hours only and then extracted on the QIAsymphony® SP instrument.
- The QIAGEN pre-lysis method with the samples being incubated overnight and then extracted on the QIAsymphony® SP instrument.

The five independent analytical scientists who conducted each of the reproducibility experiments are:

1	Scientist 1
2	Scientist 2
3	Scientist 3
4	Scientist 4
5	Scientist 5

Table 2: The five independent scientists used for the reproducibility validation

Results

Tabulated results are provided in Appendices 6-8. The reproducibility quantification results for bones 1-5 are shown in figures 7-11. The number of alleles obtained for bones 1-5 are shown in Figure 12. It should be noted the allele count for some samples were obtained after a microcon concentration procedure (refer to tabulated results in Appendices 6-8).

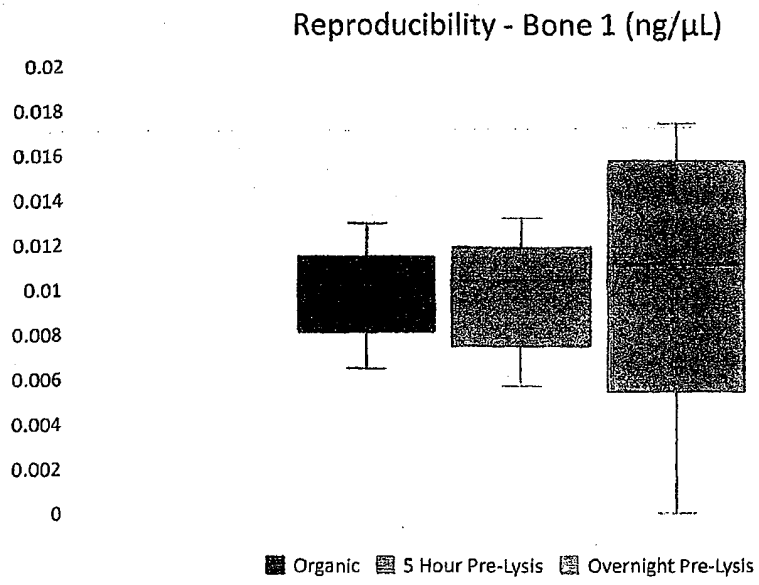


Figure 7: Reproducibility results for Bone 1 using each protocol and quant values

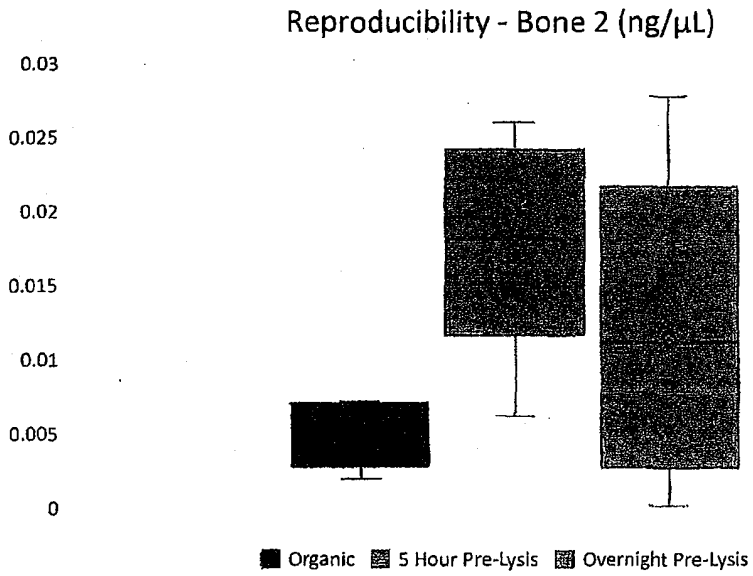


Figure 8: Reproducibility results for Bone 2 using each protocol and quant values

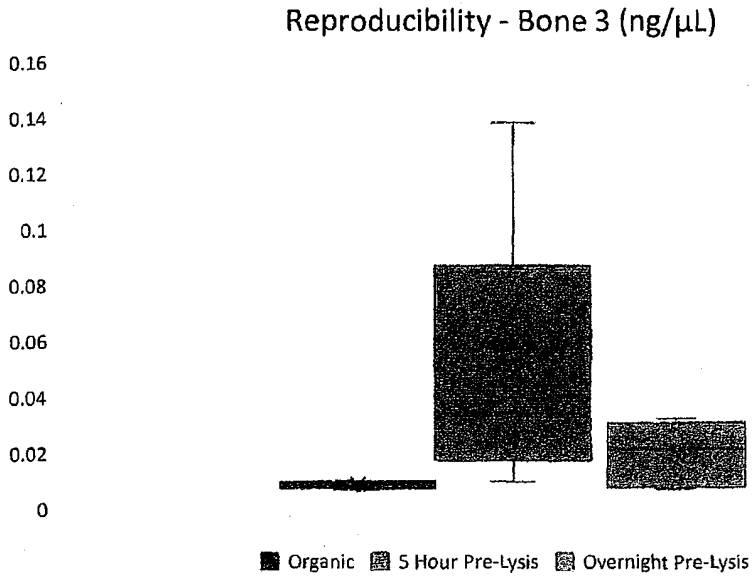


Figure 9: Reproducibility results for Bone 3 using each protocol and quant values

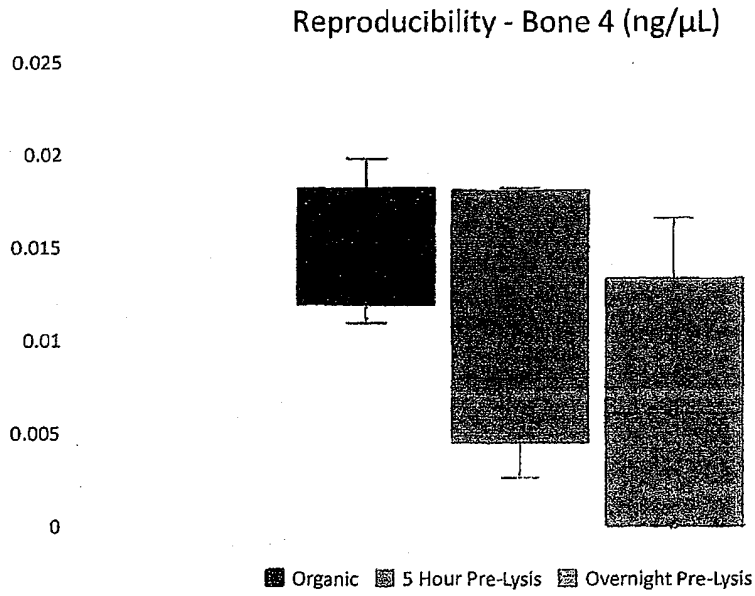


Figure 10: Reproducibility results for Bone 4 using each protocol and quant values

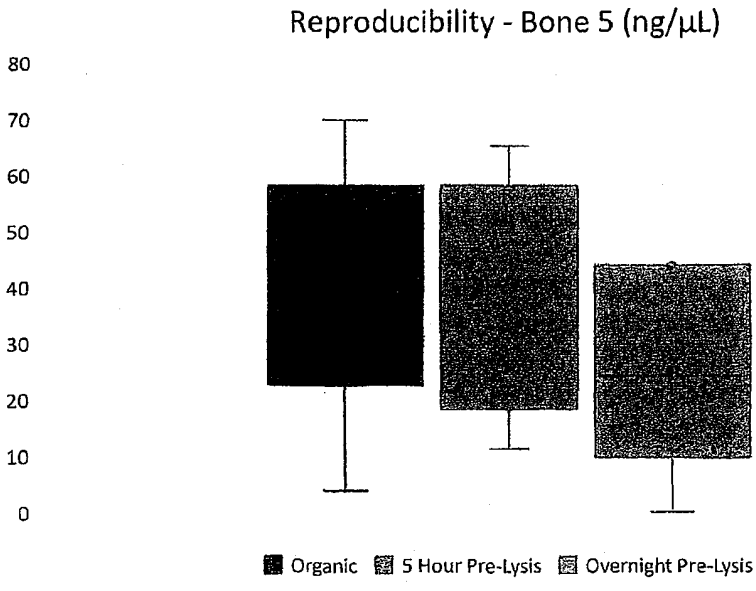


Figure 11: Reproducibility results for Bone 5 using each protocol and quant values

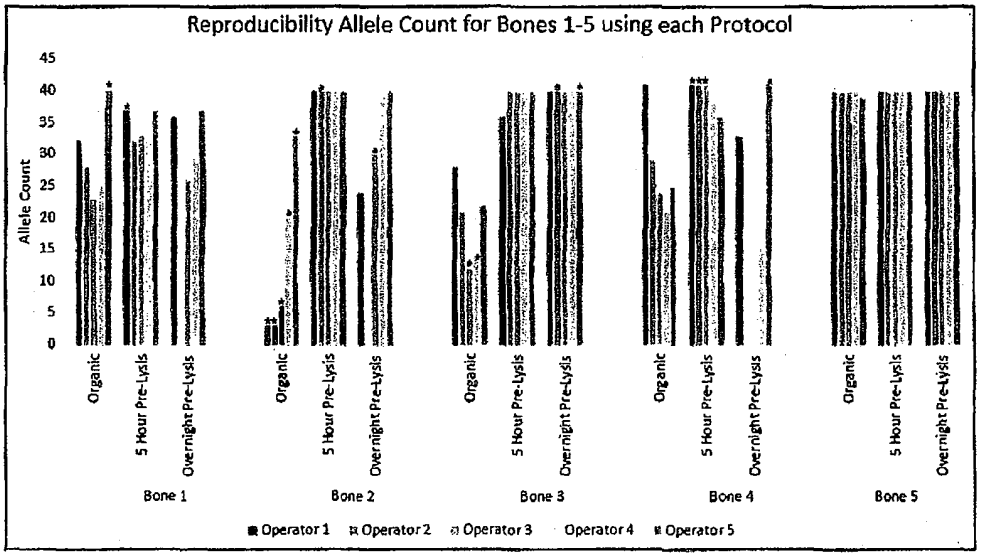


Figure 12: Reproducibility results for Bones 1-5 using allele counts for each protocol. * indicates samples which have undergone microcon concentration.

Discussion

Similar to repeatability, the reproducibility for each extraction protocol varied between samples with no apparent consistency or trend. No one protocol appeared to be more or less reproducible with consistency across the 5 samples. The organic and both QIAsymphony® protocols appeared to have a comparable level of reproducibility.

As with repeatability, the quantification results for bones 1-4 were lower than bone 5 for all extraction protocols. This is likely due to the quality of the bone samples given the consistency across each of the three extraction protocols.

For bones 1, 2 and 3, both QIAsymphony® protocols gave higher mean quantification results (across the 5 operators) than the organic extraction. For Bone 4, the organic protocol gave high average quantification results than both QIAsymphony protocols. The maximum quantification results were comparable across the three protocols protocols (0.01979, 0.01821 and 0.01660 ng/µL for organic, QIAsymphony 5 hour pre-lysis and respectively) however both QIAsymphony protocols gave more samples with lower quantification results (when compared to the Organic protocol). This was particularly evident for the overnight protocol, where two replicates gave a zero quantification result. This may be a sample specific issue as this trend was not replicated in the other bones.

For bone 5, the organic and QIAsymphony® with 5 hour pre-lysis gave comparable results, while the QIAsymphony® overnight pre-lysis extraction gave a lower mean quantification result.

Sample extracts quantified in the range 0.001-0.0088 ng/µL underwent microcon concentration prior to amplification to mimic real processing conditions. Bones 1-4 gave low quantification results which resulted in a number of samples undergoing microcon concentration. Across all samples tested, 7 organic, 5 QIAsymphony® 5 hour pe-lysis and 4 QIAsymphony® overnight lysis samples underwent microcon

concentration (see Appendices 6 - 8 for details of specific samples). Given the final DNA profile results include samples which have and have not undergone microcon concentration, the final profile and allele count results have been used only to assess any negative impact the extraction protocols may have had on profile quality. No negative impact on profile quality was noted for any of the extraction protocols.

Overall this experiment showed the QIAGEN protocols using either the 5 hour or overnight incubations gave DNA quantification results which were comparable to the organic extraction.

Additional Analysis – IPCCT

Purpose

To provide comparative analysis of IPCCT results for the tested bone extraction protocols. The compared methods were:

- The current validated method of extracting DNA from bone and teeth using organic extraction.
- The QIAGEN pre-lysis method with the samples being incubated for 5 hours only and then extracted on the QIAsymphony® SP instrument.
- The QIAGEN pre-lysis method with the samples being incubated overnight and then extracted on the QIAsymphony® SP instrument.

Results

Raw data IPCCT results can be located in the Change Management folder (I:\Change Management\Proposal#192 - QIAsymphony Bone Extraction\Supplementary R&R\Results – Supp R&R.xls). Figures 13-17 below contain the IPCCT results for bones 1-5.

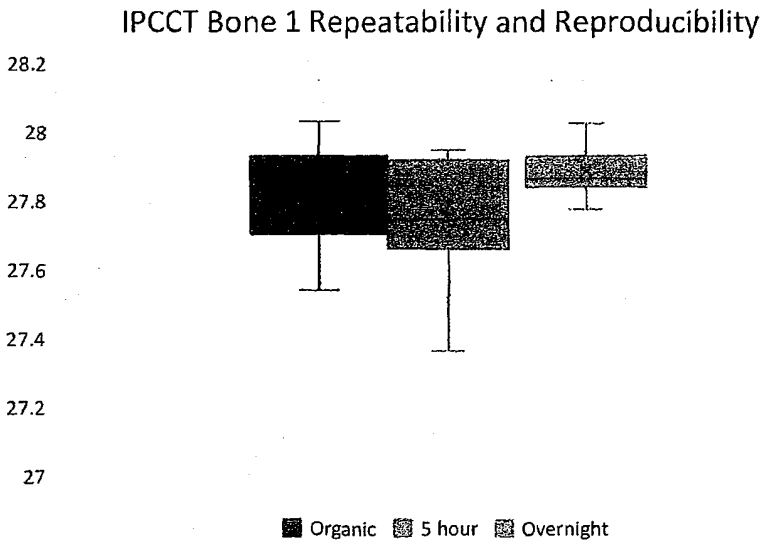


Figure 13: IPCCT results for Bone 1 – Combined Repeatability and Reproducibility data.

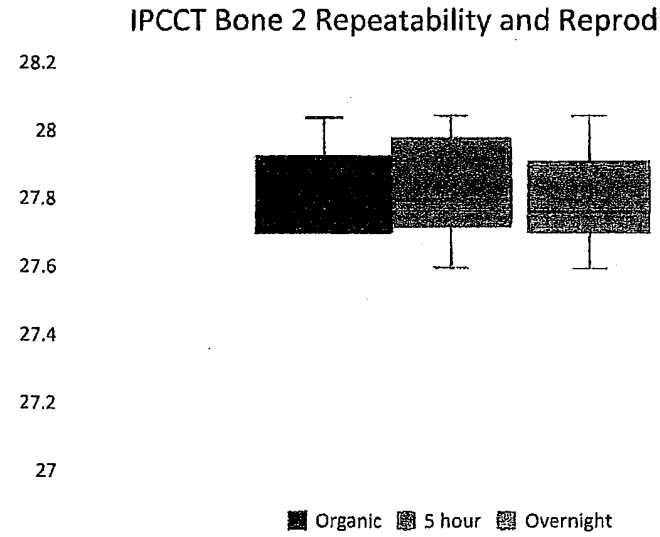


Figure 14: IPCCT results for Bone 2 – Combined Repeatability and Reproducibility data.

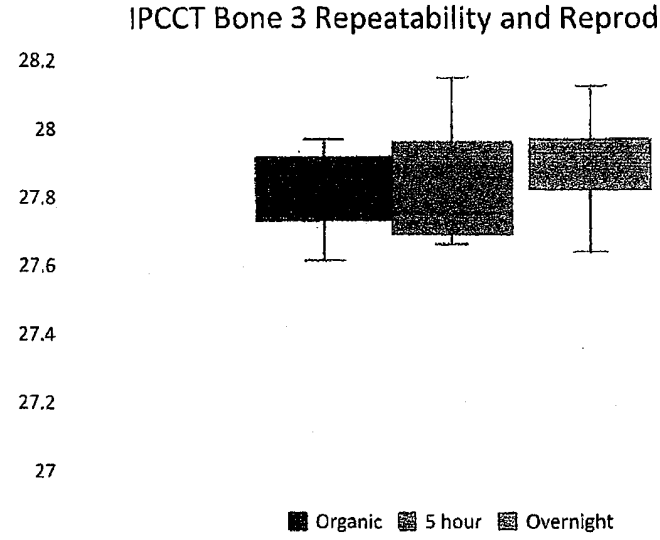


Figure 15: IPCCT results for Bone 3 – Combined Repeatability and Reproducibility data.

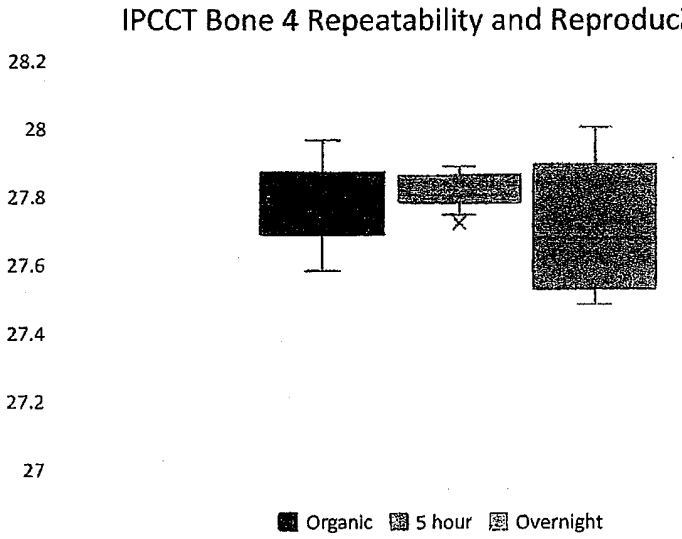


Figure 16: IPCCT results for Bone 4 – Combined Repeatability and Reproducibility data.

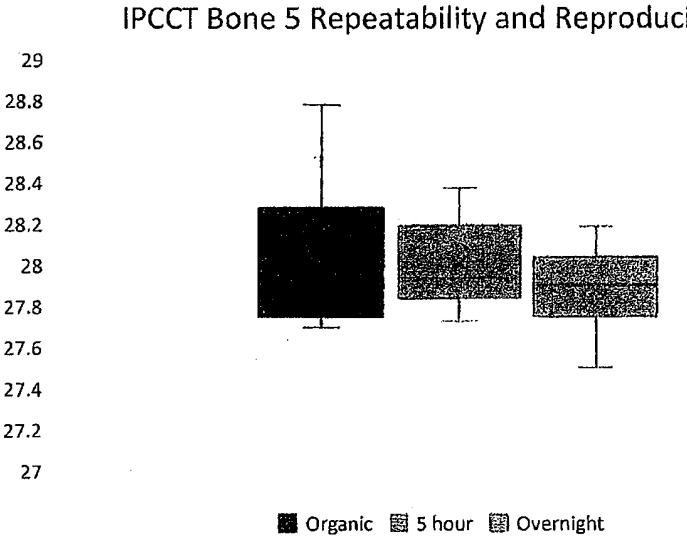


Figure 17: IPCCT results for Bone 5 – Combined Repeatability and Reproducibility data.

Discussion

The combined repeatability and reproducibility IPCCT results for bones 1-5 were compared across the three extraction protocols. Bones 1-5 showed comparable IPCCT results across each of the three tested protocols, with no indication of inhibition.

Conclusion and Recommendations

Overall the results of these additional experiments have shown that the QIAsymphony SP extraction with both 5 hour and overnight pre-lysis produce comparable DNA yields and repeatability and reproducibility to the current organic extraction. There was some evidence that the overnight pre-lysis produced higher DNA yields than the 5 hour pre-lysis and this fits with intuitive expectations given the longer reaction time. It should be noted that it is routine practice for multiple samples from a single bone to be submitted for DNA analysis, which may mitigate and/or compensate for some of the sample to sample variability observed in this validation.

As noted in the discussion, sample extracts quantified in the range 0.001-0.0088 ng/µL underwent microcon concentration. Samples underwent microcon concentration 13 times for the organic extraction, 8 times for the QIAsymphony 5 hour pre-lysis protocol and 4 times for the QIAsymphony overnight pre-lysis protocol. This indicates the QIAsymphony overnight pre-lysis protocol gave extracts with less samples in the 0.001-0.0088ng/µL microcon concentration range.

In addition to workflow efficiency improvements, implementation of the QIAsymphony for bone extraction also improves occupational health and safety for staff by removing the use of phenol chloroform in the organic extraction.

It is therefore recommended that:

- The DNA extraction of bones on the QIAsymphony SP is implemented as a replacement for organic extraction.
- The organic extraction SOP is archived.
- The overnight pre-lysis is used for routine, non-time critical bone processing given the evidence of higher DNA yields.
- The 5 hour pre-lysis protocol is considered for use where
    - there is a large number of samples and/or where time critical processing is required (i.e. for DVIs), or
    - samples are expected to provide good DNA yields and there is sufficient material for retesting if required.

References

Agullera, M., Micic, B., Acedo, P., Ryan, L. and Allen, C. (2016) Validation of the QIAsymphony6® SP/AS Modules [Final Report].

- QIS 34039 Extracting DNA from Bone and Teeth
- QIS 34045 Quantification of Extracted DNA using the Quantifiler® Trio DNA Quantification Kit
- QIS 34052 Amplification of Extracted DNA using the PowerPlex®21 System
- QIS 34112 STR fragment analysis of PowerPlex® 21 profiles using GeneMapper® ID-X software
- QIS 34131 Capillary Electrophoresis Quality (CEQ) Check
- QIS 34132 DNA Extraction and Quantification of samples using the QIAsymphony® SP and AS – FR

Appendix 1 - Table of Results: Bone 1

Bone	Method	Barcode	Quant Value ng/µL	Rework	Allele Count
1	Organic	[redacted]	0.01296		32
1	Organic		0.00032		1
1	Organic		0.00747	Microcon	33
1	Organic		0.01054		29
1	Organic		0.00956		30
1	Pre-Lysis 5 hour		0.00568	Microcon	37
1	Pre-Lysis 5 hour		0.00835		40
1	Pre-Lysis 5 hour		0.00694	Microcon	40
1	Pre-Lysis 5 hour		0.00619	Microcon	40
1	Pre-Lysis 5 hour		0.00754	Microcon	40
1	Pre-Lysis Overnight		0.014		36
1	Pre-Lysis Overnight		0.015		36
1	Pre-Lysis Overnight		0.009		29
1	Pre-Lysis Overnight		0.010		32
1	Pre-Lysis Overnight		0.013		33

Table 3: Repeatability for Bone 1 using the three different methods tested including the Organic Extraction, the QIAsymphony® Pre-Lysis (5 hour) and Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction.

Appendix 2 - Table of Results: Bone 2

Bone	Method	Barcode	Quant Value ng/µL	Rework	Allele Count
2	Organic	[redacted]	0.00361	Microcon	16
2	Organic		0.00274	Microcon	2
2	Organic		0.00311	Microcon	9
2	Organic		0.00038		0
2	Organic		0.00444	Microcon	16
2	Pre-Lysis 5 hour		0.01811		40
2	Pre-Lysis 5 hour		0.0177		38
2	Pre-Lysis 5 hour		0.01412		40
2	Pre-Lysis 5 hour		0.01253		40
2	Pre-Lysis 5 hour		0.01984		40
2	Pre-Lysis Overnight		0.01112		24
2	Pre-Lysis Overnight		0.02394		38
2	Pre-Lysis Overnight		0.02606		40
2	Pre-Lysis Overnight		0.02558		39
2	Pre-Lysis Overnight		0.02697		40

Table 4: Repeatability for Bone 2 using the three different methods tested including the Organic Extraction, the QIAsymphony® Pre-Lysis (5 hour) and Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction

Appendix 3 - Table of Results: Bone 3

Bone	Method	Barcode	Quant Value ng/µL	Rework	Allele Count
3	Organic	[redacted]	0.00965		28
3	Organic		0.00520	Microcon	18
3	Organic		0.01075		27
3	Organic		0.01485		32
3	Organic		0.01017		27
3	Pre-Lysis 5 hour		0.00996		36
3	Pre-Lysis 5 hour		0.00840		40
3	Pre-Lysis 5 hour		0.02136		40
3	Pre-Lysis 5 hour		0.01434		40
3	Pre-Lysis 5 hour		0.01998		40
3	Pre-Lysis Overnight		0.02185		40
3	Pre-Lysis Overnight		0.01757		40
3	Pre-Lysis Overnight		0.02838		40
3	Pre-Lysis Overnight		0.02348		40
3	Pre-Lysis Overnight		0.01603		40

Table 5: Repeatability for Bone 3 using the three different methods tested including the Organic Extraction, the QIAsymphony® Pre-Lysis (5 hour) and Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction

Appendix 4 - Table of Results: Bone 4

Bone	Method	Barcode	Quant Value ng/µL	Rework	Allele Count
4	Organic	[redacted]	0.01979		41
4	Organic		0.01077		41
4	Organic		0.00044		0
4	Organic		0.0000		0
4	Organic		0.00999		30
4	Pre-Lysis 5 hour		0.00792		31
4	Pre-Lysis 5 hour		0.00643	Microcon	41
4	Pre-Lysis 5 hour		0.00787	Microcon	41
4	Pre-Lysis 5 hour		0.00591	Microcon	33
4	Pre-Lysis 5 hour		0.00643	Microcon	41
4	Pre-Lysis Overnight		0.01660		33
4	Pre-Lysis Overnight		0.01300		33
4	Pre-Lysis Overnight		0.01471		33
4	Pre-Lysis Overnight		0.01193		32
4	Pre-Lysis Overnight		0.01296		32

Table 6: Repeatability for Bone 4 using the three different methods tested including the Organic Extraction, the QIAsymphony® Pre-Lysis (5 hour) and Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction

Appendix 5 - Table of Results: Bone 5

Bone	Method	Barcode	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
5	Organic	[redacted]	3.90037			40
5	Organic		7.82573	Dilution	714153595	40
5	Organic		2.79068			40
5	Organic		3.52519			40
5	Organic		6.41383	Dilution	714153603	40
5	Pre-Lysis 5 hour		25.25375	Dilution	714153523	40
5	Pre-Lysis 5 hour		27.41910	Dilution	714153534	40
5	Pre-Lysis 5 hour		45.80926	Dilution	714153540	40
5	Pre-Lysis 5 hour		41.76097	Dilution	714153556	40
5	Pre-Lysis 5 hour		34.41719	Dilution	714153567	40
5	Pre-Lysis Overnight		24.29558	Dilution	724204372	40
5	Pre-Lysis Overnight		31.82610	Dilution	724204381	40
5	Pre-Lysis Overnight		31.86150	Dilution	724204390	40
5	Pre-Lysis Overnight		32.42724	Dilution	724204407	40
5	Pre-Lysis Overnight		36.27103	Dilution	724204416	40

Table 7: Repeatability for Bone 5 using the three different methods tested including the Organic Extraction, the QIAsymphony® Pre-Lysis (5 hour) and Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction

Appendix 6 - Reproducibility Table of Results for Organic Extraction

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele Count
Day 1	Scientist 1	1	[redacted]	0.01296		32
Day 2	Scientist 2	1		0.00999		28
Day 3	Scientist 3	1		0.00990		23
Day 4	Scientist 4	1		0.00974		25
Day 5	Scientist 5	1		0.00647	Microcon	40

Table 8: Reproducibility results for the Current Organic Extraction for Bone 1

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele Count
Day 1	Scientist 1	2	[redacted]	0.00361	Microcon	3
Day 2	Scientist 2	2		0.00196	Microcon	3
Day 3	Scientist 3	2		0.00458	Microcon	6
Day 4	Scientist 4	2		0.00713	Microcon	20
Day 5	Scientist 5	2		0.00692	Microcon	33

Table 9: Reproducibility results for the Current Organic Extraction for Bone 2

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele Count
Day 1	Scientist 1	3	[redacted]	0.00965		28
Day 2	Scientist 2	3		0.00953		21
Day 3	Scientist 3	3		0.00785	Microcon	12
Day 4	Scientist 4	3		0.00723	Microcon	13
Day 5	Scientist 5	3		0.01027		22

Table 10: Reproducibility results for the Current Organic Extraction for Bone 3

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele Count
Day 1	Scientist 1	4	[redacted]	0.01979		41
Day 2	Scientist 2	4		0.01093		29
Day 3	Scientist 3	4		0.01657		24
Day 4	Scientist 4	4		0.01294		21
Day 5	Scientist 5	4		0.01324		25

Table 11: Reproducibility results for the Current Organic Extraction for Bone 4

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	5	[redacted]	3.90037			40
Day 2	Scientist 2	5		69.92214	Dilution	714140860	40
Day 3	Scientist 3	5		42.16170	Dilution	720478242	40
Day 4	Scientist 4	5		46.41128	Dilution	714153578	40
Day 5	Scientist 5	5		41.47680	Dilution	714153589	39

Table 12: Reproducibility results for the Current Organic Extraction for Bone 5

Appendix 7 - Reproducibility Table of Results for the QIAsymphony Pre-Lysis (5 hour incubation)

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	1	[redacted]	0.00568	Microcon		37
Day 2	Scientist 2	1		0.00922			32
Day 3	Scientist 3	1		0.01038			33
Day 4	Scientist 4	1		0.01321			33
Day 5	Scientist 5	1		0.01056			37

Table 13: Reproducibility results for the QIAsymphony® Pre-Lysis (5 hour incubation) and QIAsymphony® SP Extraction for Bone 1

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	2	[redacted]	0.01811			40
Day 2	Scientist 2	2		0.00614	Microcon		40
Day 3	Scientist 3	2		0.02598			40
Day 4	Scientist 4	2		0.02227			40
Day 5	Scientist 5	2		0.01705			40

Table 14: Reproducibility results for the QIAsymphony® Pre-Lysis (5 hour incubation) and QIAsymphony® SP Extraction for Bone 2

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	3	[redacted]	0.00996			36
Day 2	Scientist 2	3		0.1384			40
Day 3	Scientist 3	3		0.02495			40
Day 4	Scientist 4	3		0.03412			40
Day 5	Scientist 5	3		0.03583			40

Table 15: Reproducibility results for the QIAsymphony® Pre-Lysis (5 hour incubation) and QIAsymphony® SP Extraction for Bone 3

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	4	[redacted]	0.00792	Microcon		41
Day 2	Scientist 2	4		0.00594	Microcon		41
Day 3	Scientist 3	4		0.00632	Microcon		41
Day 4	Scientist 4	4		0.01793			38
Day 5	Scientist 5	4		0.01821			36

Table 16: Reproducibility results for the QIAsymphony® Pre-Lysis (5 hour incubation) and QIAsymphony® SP Extraction for Bone 4

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele count
Day 1	Scientist 1	5	[redacted]	25.25375	Dilution	714153523	40
Day 2	Scientist 2	5		11.38112	Dilution	718880513	40
Day 3	Scientist 3	5		41.33673	Dilution	718880491	40
Day 4	Scientist 4	5		51.12676	Dilution	718880541	39
Day 5	Scientist 5	5		65.29295	Dilution	718880530	40

Table 17: Reproducibility results for the QIAsymphony® Pre-Lysis (5 hour incubation) and QIAsymphony® SP Extraction for Bone 5

Appendix 8 - Reproducibility Table of Results for the QIAsymphony Pre-Lysis (Overnight incubation)

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele count
Day 1	Scientist 1	1	[redacted]	0.01409		36
Day 2	Scientist 2	1		0.00000		0
Day 3	Scientist 3	1		0.01114		26
Day 4	Scientist 4	1		0.01741		30
Day 5	Scientist 5	1		0.01084		37

Table 18: Reproducibility results for the QIAsymphony® Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction for Bone 1

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	
Day 1	Scientist 1	2	[redacted]	0.01112		24
Day 2	Scientist 2	2		0.00013		0
Day 3	Scientist 3	2		0.00521	Microcon	30
Day 4	Scientist 4	2		0.02773		40
Day 5	Scientist 5	2		0.01552		40

Table 19: Reproducibility results for the QIAsymphony® Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction for Bone 2

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele count
Day 1	Scientist 1	3	[redacted]	0.02185		40
Day 2	Scientist 2	3		0.00865	Microcon	40
Day 3	Scientist 3	3		0.02976		40
Day 4	Scientist 4	3		0.03259		40
Day 5	Scientist 5	3		0.00743	Microcon	40

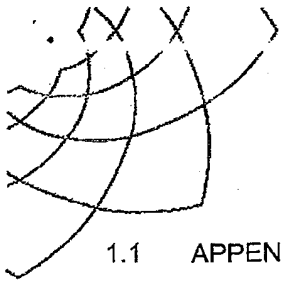
Table 20: Reproducibility results for the QIAsymphony® Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction for Bone 3

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Allele count
Day 1	Scientist 1	4	[redacted]	0.01660		33
Day 2	Scientist 2	4		0.0000		0
Day 3	Scientist 3	4		0.0000		0
Day 4	Scientist 4	4		0.01010		15
Day 5	Scientist 5	4		0.00604	Microcon	41

Table 21: Reproducibility results for the QIAsymphony® Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction for Bone 4

Day	Operator	Bone Sample	Barcode Number	Quant Value ng/µL	Rework	Rework Barcode	Allele Count
Day 1	Scientist 1	5	[redacted]	24.29558	Dilution	724204372	40
Day 2	Scientist 2	5		0.11813			40
Day 3	Scientist 3	5		44.00444	Dilution	723695325	40
Day 4	Scientist 4	5		44.07124	Dilution	718880557	40
Day 5	Scientist 5	5		19.29339	Dilution	718880524	40

Table 22: Reproducibility results for the QIAsymphony® Pre-Lysis (Overnight incubation) and QIAsymphony® SP Extraction for Bone 5



HealthSupport
Queensland
Forensic and Scientific Services

1.1 APPENDIX 3: Implementation Plan for project leaders

Successful project implementation may require numerous tasks to be completed either prior to implementation, or shortly after the implementation date. Some of the considerations/tasks that may be required are listed below; however this is not intended to be a comprehensive list of tasks as each project will have different implementation requirements. Project leaders should devise and submit a comprehensive implementation plan for management review. Once complete, the checklist should be submitted to the quality team for filing with the signed project documents.

Task	Details	Date Completed
Staff Training	All current QIAsymphony trainers to be assessed at CTT using RCC given similarity of bone and other substrate protocols.	24/03/2020
Staff Training	All current QIAsymphony operators (assessed as competent) will be assessed as competent using RCC given similarity of bone and other substrate protocols.	24/03/2020
Add to minor change register	Ensure that implementation has been added to the minor changes register	24/03/2020
Communication	Communicate to staff and other stakeholders – by meetings and emails.	24/03/2020
SOP	Archive Organic extraction SOP (QIS# 34039)	24/03/2020
SOP	Add bone extraction protocols to QIS # 34132 DNA Extraction and Quantification of Samples Using the QIAsymphony® SP and AS Modules	23/03/2020

Forensic DNA Analysis

Analysis of data for stutter threshold selection for VFP

Analysis conducted by Emma Caunt, Cassandra James, Angela Adamson & Rhys Parry to be presented to the Verifiler team

August 2021

Introduction

Stutter thresholds are required to enable the scientist to interpret DNA profiles and to assign a number of contributors to a DNA profile; these stutter thresholds will be called 'intuitive stutter thresholds'.

Prior to the introduction of continuous profile interpretation, binary thresholds were applied. These thresholds were calculated using the mean + 3SD. In terms of stutter, this meant that the scientist could be confident that any peak above this threshold was likely to be allelic.

With the introduction of continuous profile interpretation, it is no longer necessary for the scientist to place this degree of certainty on a peak being allelic. STRmix will take into account a number of characteristics of the profile to calculate a weighting of the peak being stutter or allelic which is ultimately factored into the likelihood ratio (LR).

Since the introduction of STRmix into the Forensic DNA Analysis laboratory, casework experience has shown that in some instances using an intuitive stutter threshold of mean + 3SD is too high and does not align with STRmix modelling.

In order to close the gap between the intuitive stutter thresholds and the STRmix stutter ratios a decision has been made to reduce the intuitive stutter thresholds to the mean + 2SD. In some instances, the data shows that the mean + 2SD value may still be too high and therefore the 95% confidence interval in the regression line has been used.

This document shows the stutter data analysis for Project #213 – Verifiler Plus and the suggested intuitive stutter thresholds for the interpretation of Verifiler Plus profiles.

-1 rpt stutter

The observed stutter ratios per allele were plotted for each locus and the regression line was determined. The 95% confidence interval of the regression line was also plotted along with the mean + 2SD values for each allele. Where appropriate the same information was plotted for LUS (longest uninterrupted stretch). The R^2 value and equation for the observed regression line is also displayed. The LUS value is provided by STRmix based on alleles being sequenced. If an allele has not been sequenced, then the LUS value is not known and therefore a stutter threshold based on the allele will be used as this is the best information available. If the LUS value is not observed in the data but a value is known, the 95% regression line for LUS vs SR will be used. Some alleles have more than

one LUS value based on differing structures so in these instances STRmix has provided an average of these values which provides LUS values with decimals.

For the loci where the LUS Look Up Table [1] states that "the line formula for the plot LUS vs SR is used to calculate the SR for each allele", the LUS data was used to calculate the intuitive stutter thresholds. These loci are D3S1358, vWA, D6S1043, D8S1179, D21S11, D19S433, FGA, D1S1656, TH01, D12S391 and D2S1338.

Alleles or LUS values that have less than three data points will be treated as non-observed alleles.

D3S1358

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold. One exception to this is the LUS value of 15.5 where the mean + 2SD value is skewed by one outlier. For this LUS value the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR (

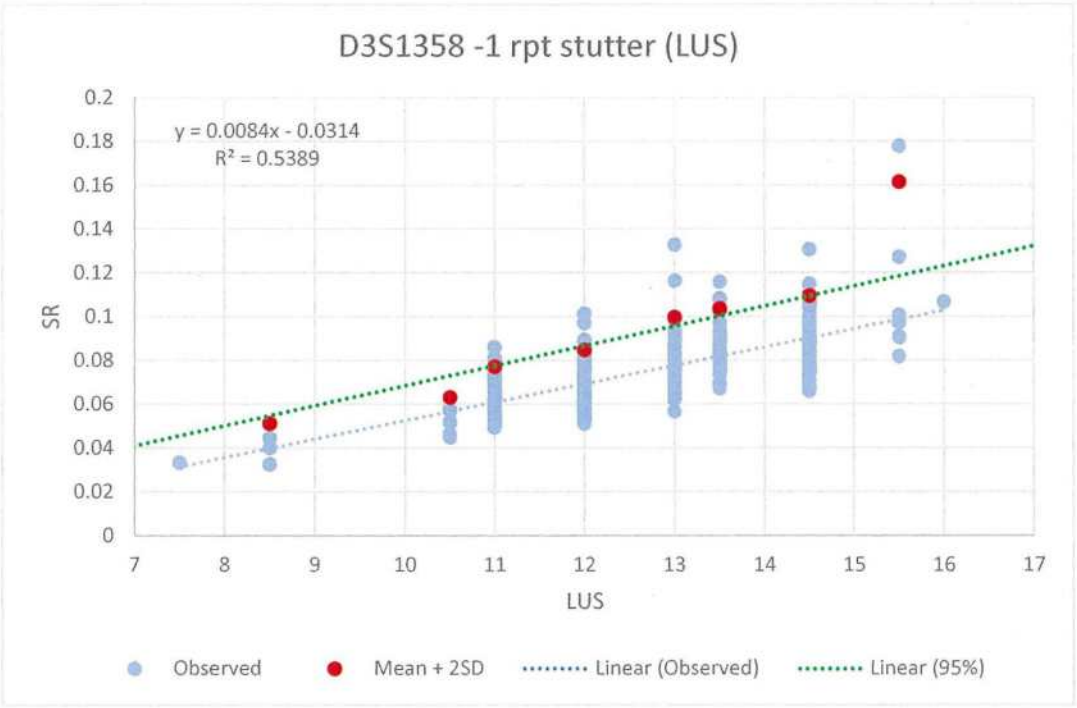


Figure 1).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (

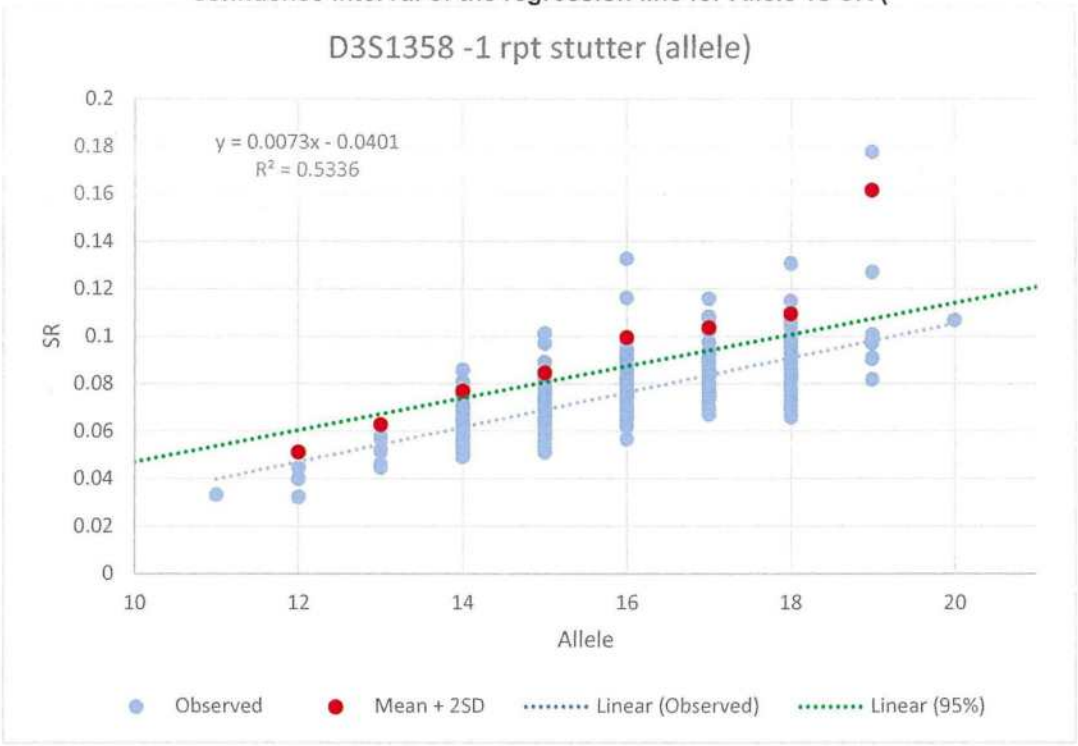


Figure 2).

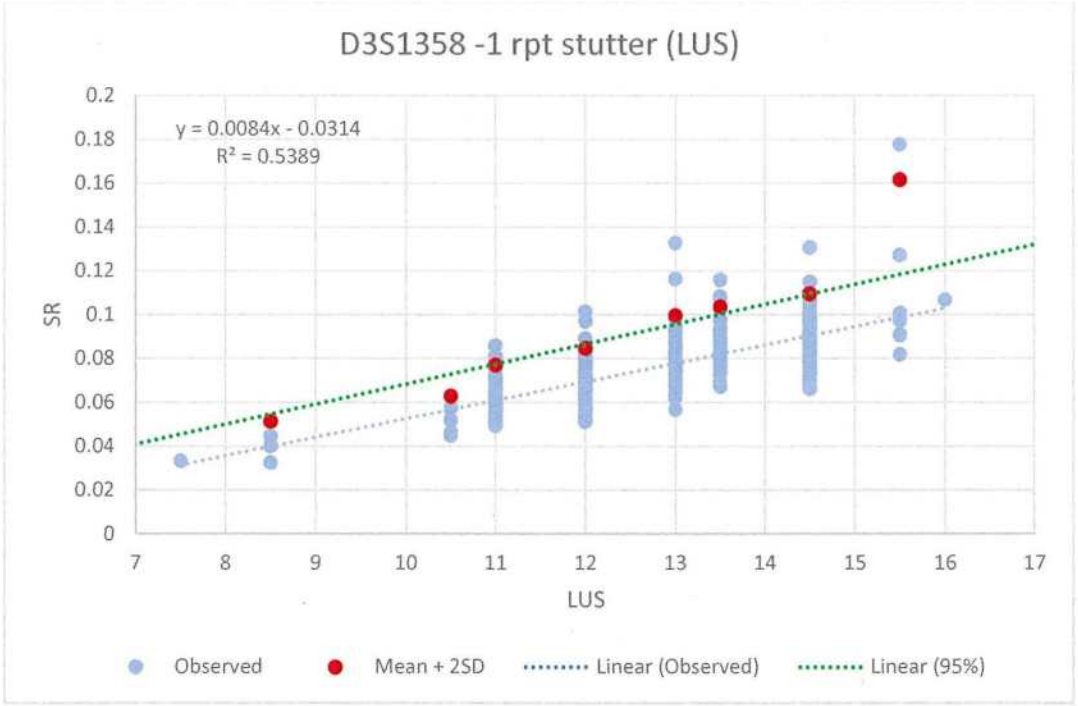


Figure 1

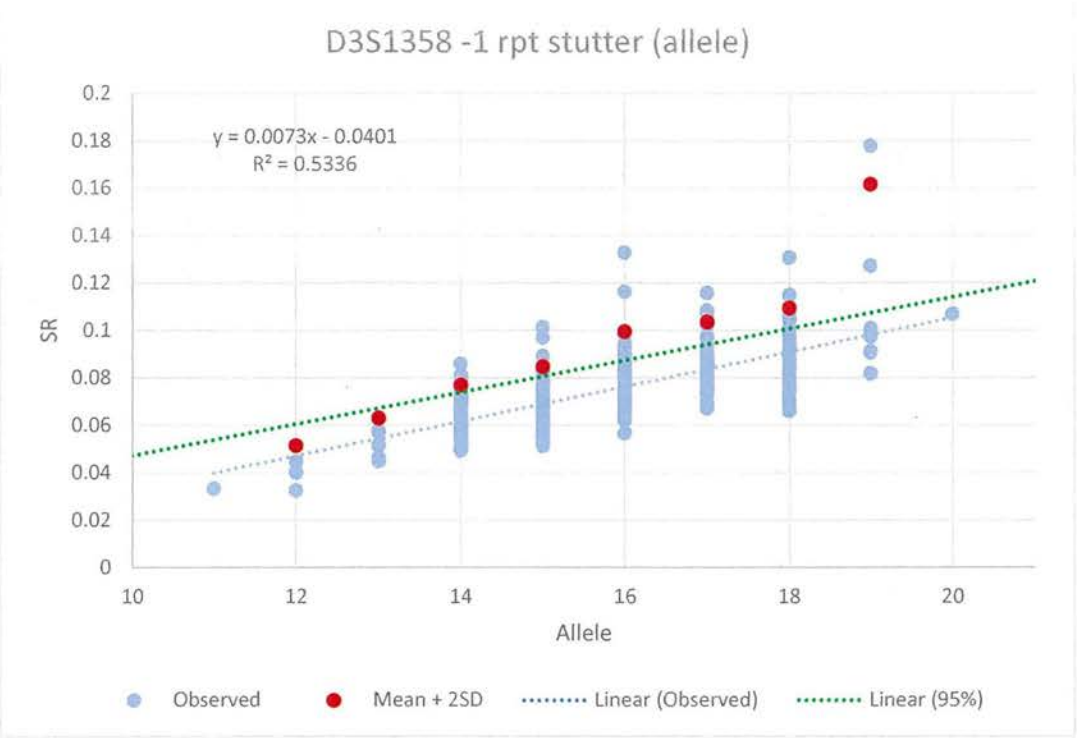


Figure 2

vWA

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 3).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 4).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS value of 13 are above the regression line value and therefore the average observed stutter ratio for the LUS value of 13 will be used in the STRmix stutter exceptions file.

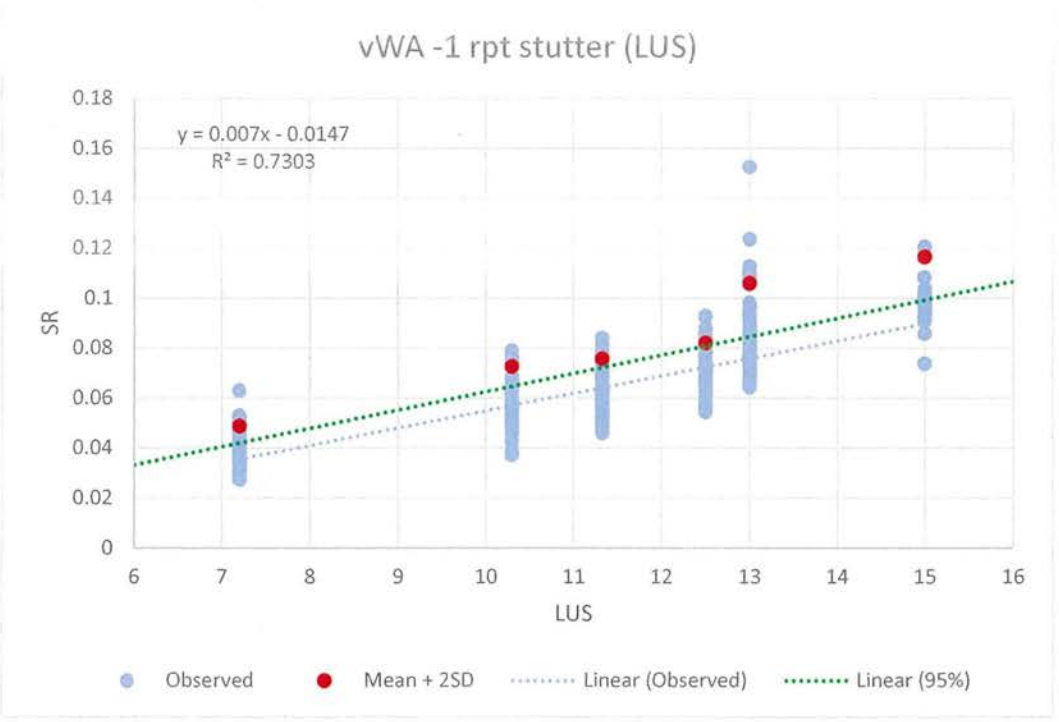


Figure 3

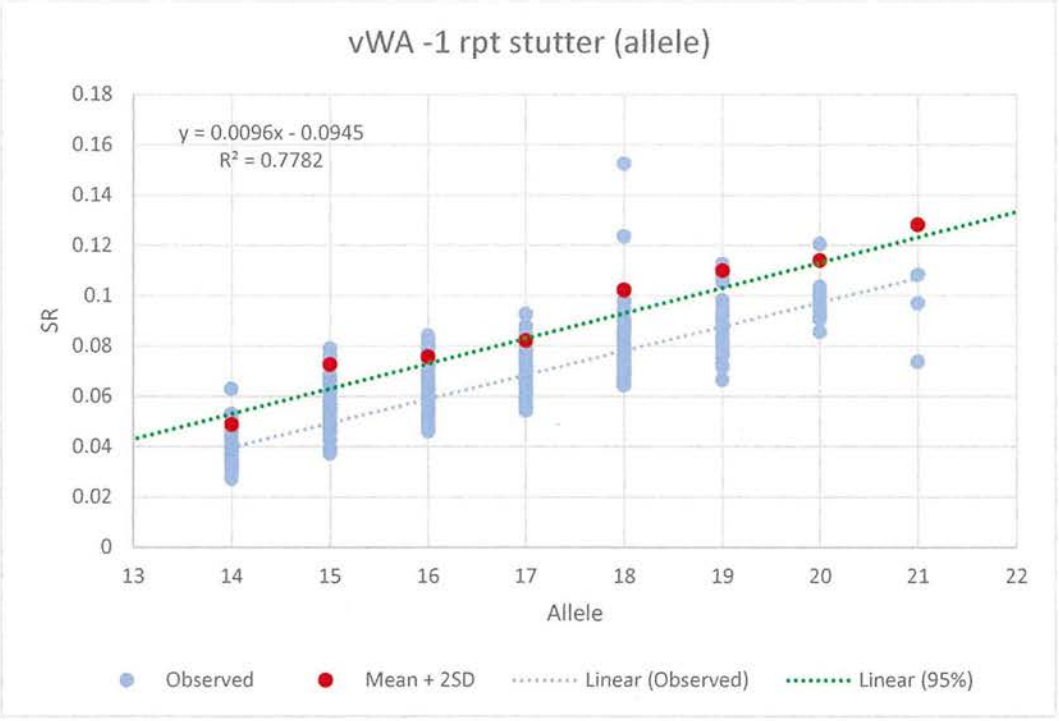


Figure 4

D16S539

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles. One exception to this is the 8 allele where the mean + 2SD value is skewed by one outlier. For this allele the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (

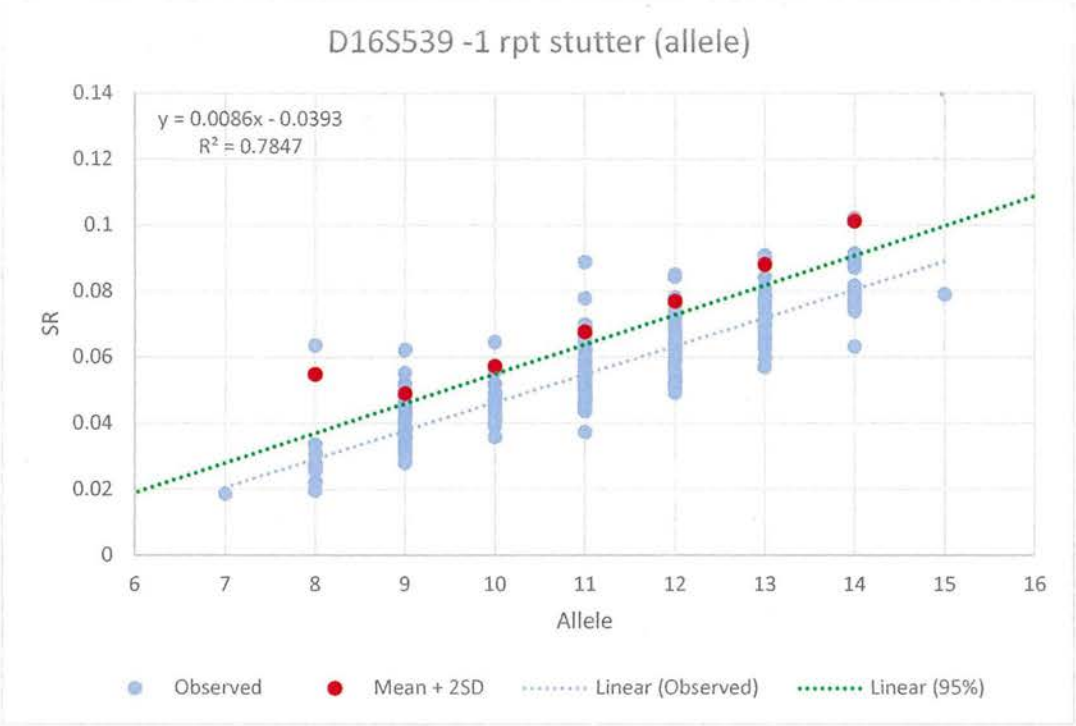


Figure 5).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

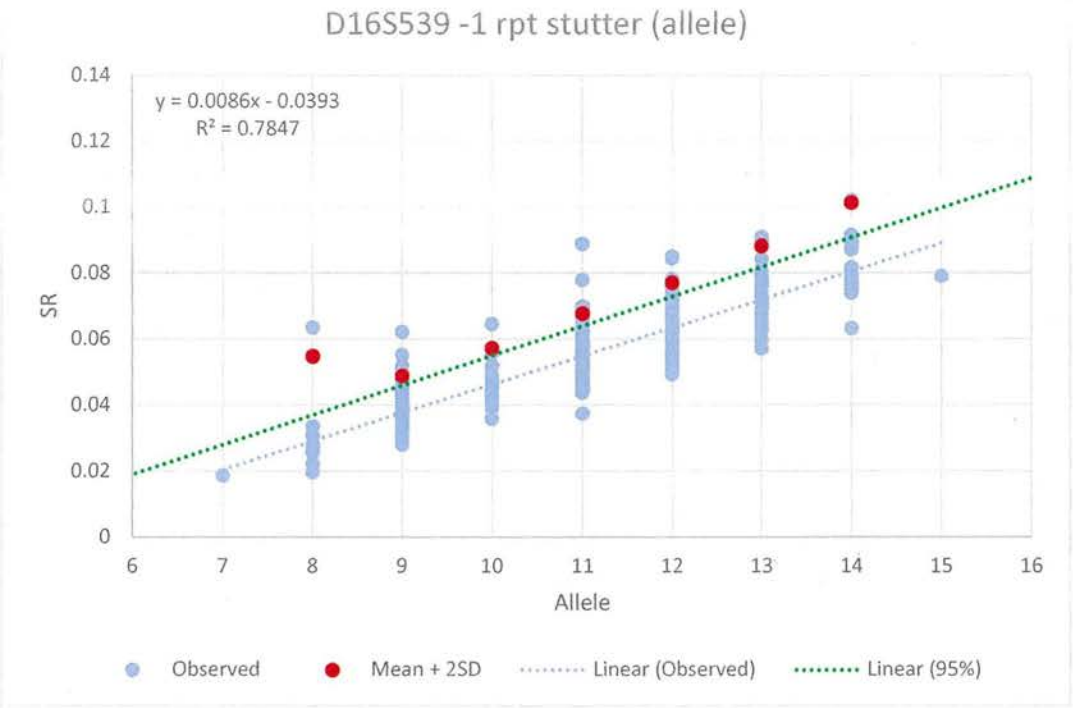


Figure 5

CSF1PO

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 6).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

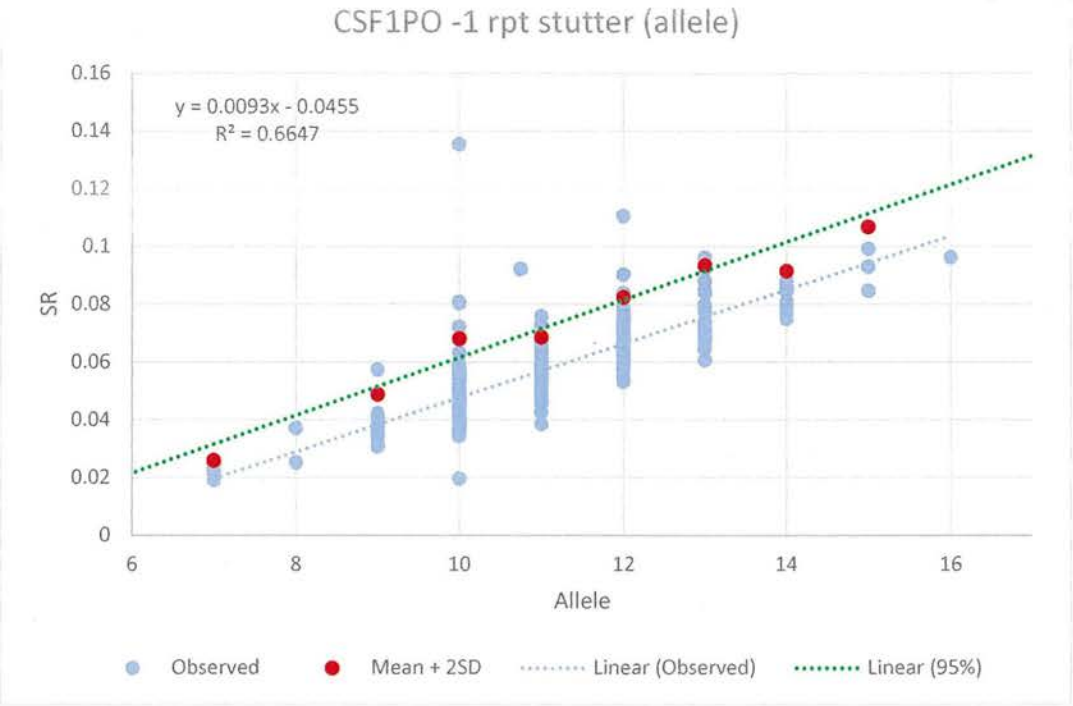


Figure 6

D6S1043

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (

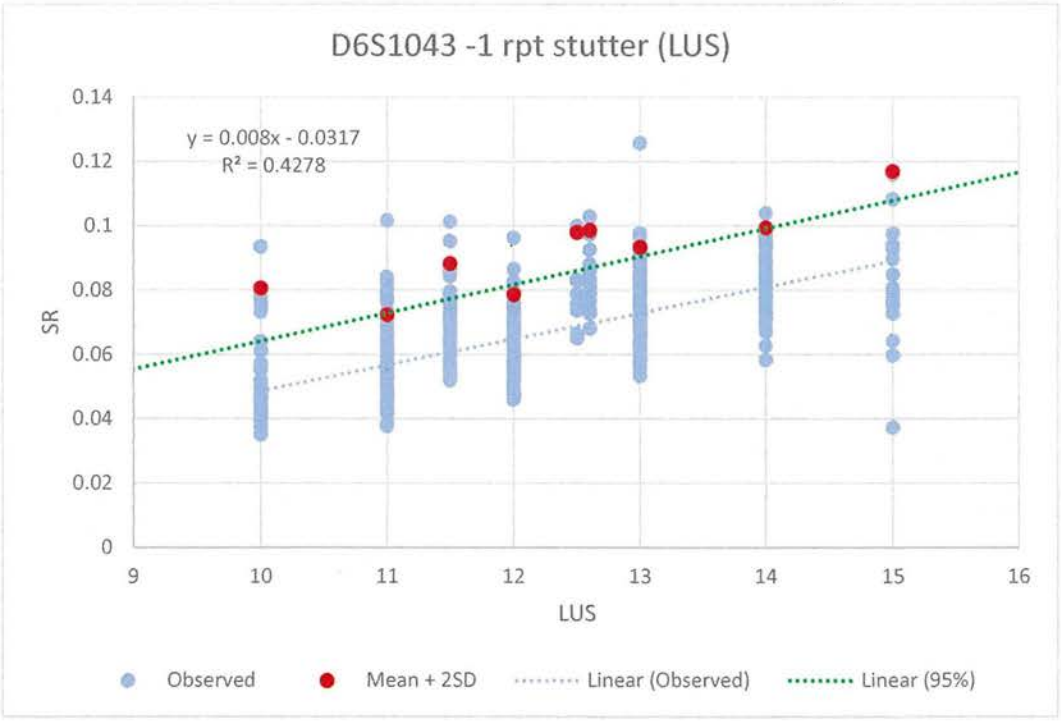


Figure 7).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 8).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS values of 11.5, 12.5 and 12.6 are above the regression line value and therefore the average observed stutter ratio for LUS values of 11.5, 12.5 and 12.6 will be used in the STRmix stutter exceptions file.

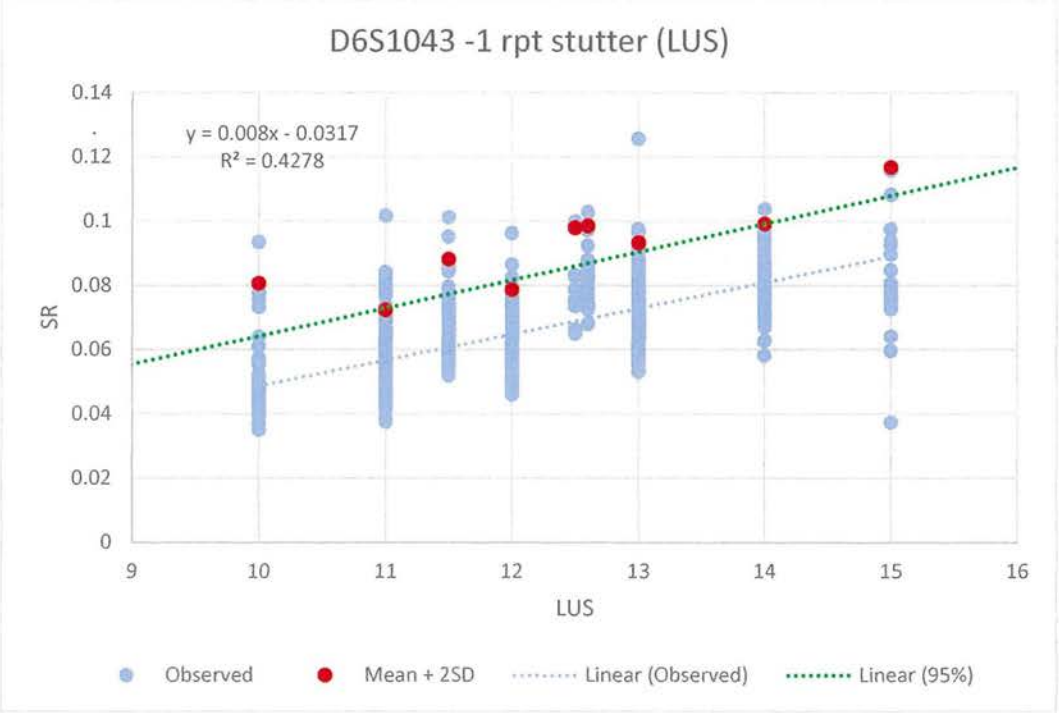


Figure 7

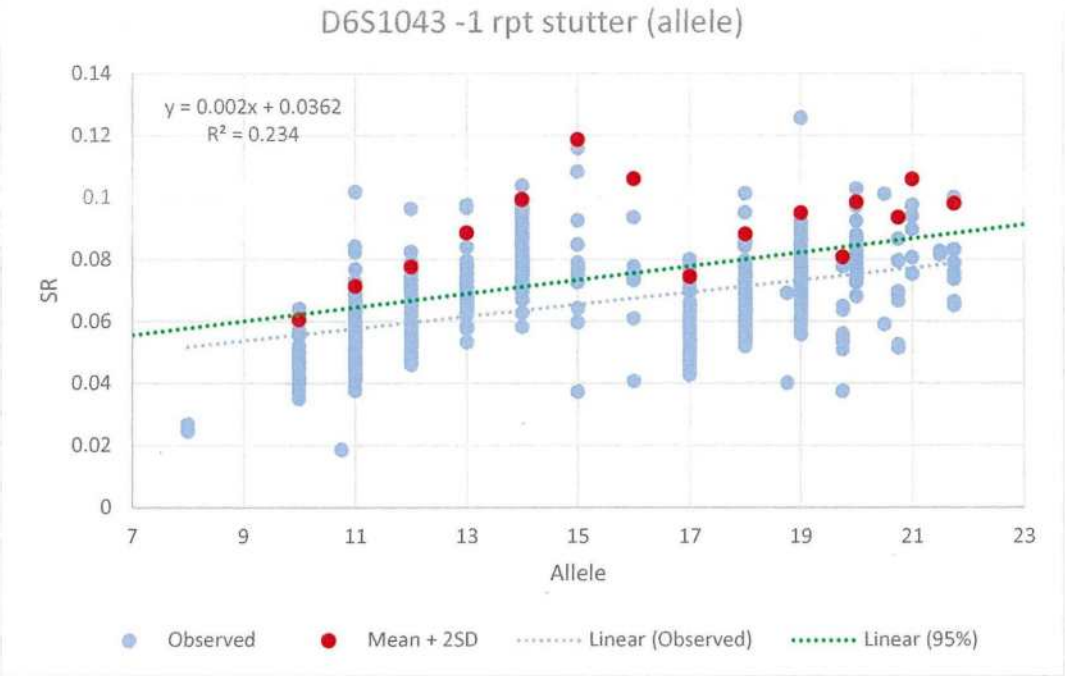


Figure 8

D8S1179

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 9).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 10).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS values of 9.5 and 10.3 are above the regression line value and therefore the average observed stutter ratio for the LUS values of 9.5 and 10.3 will be used in the STRmix stutter exceptions file.

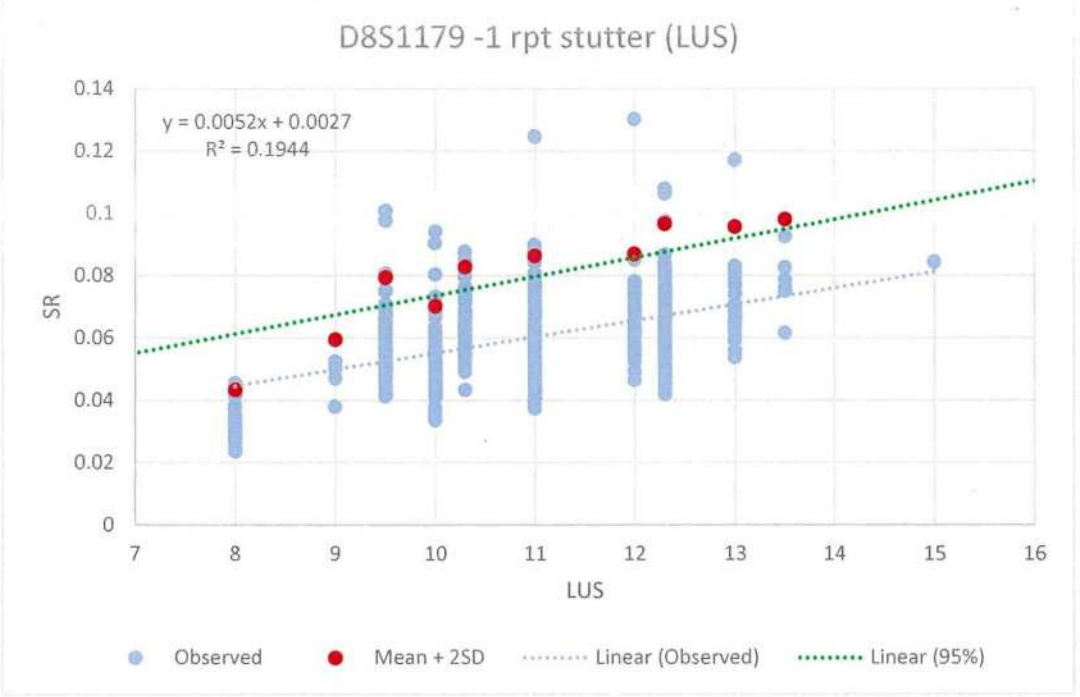


Figure 9

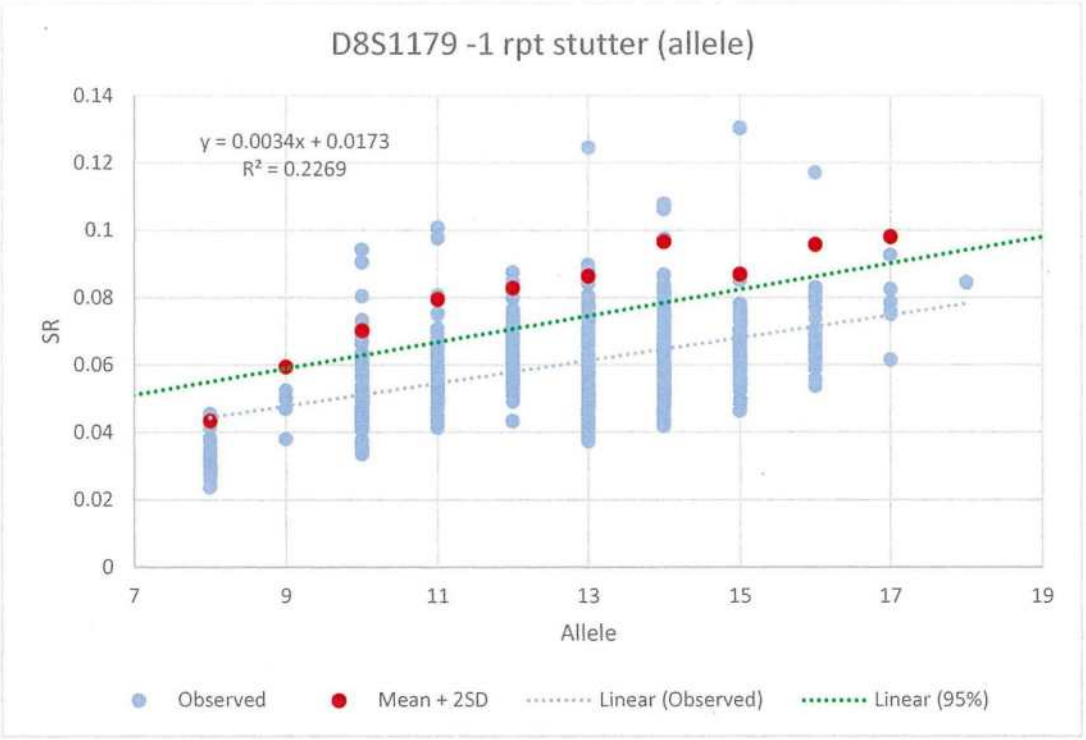


Figure 10

D21S11

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold. Exceptions to this are the LUS values of 14 and 15 where the mean + 2SD value is skewed by one outlier. For these LUS values the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR (Figure 11).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 12).

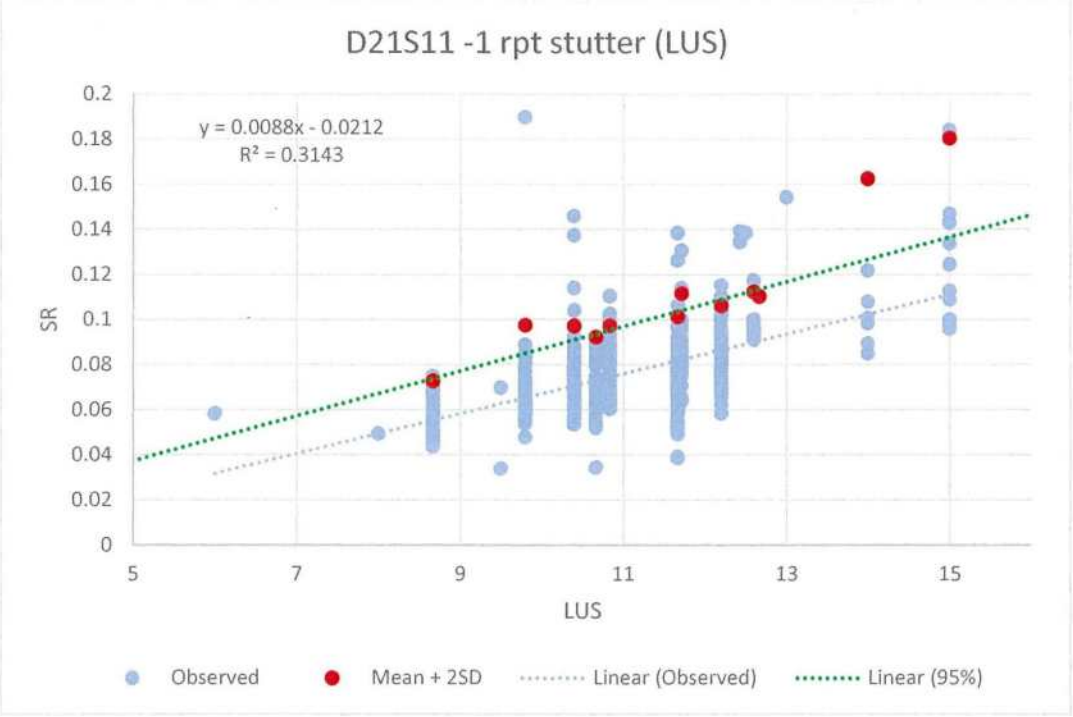


Figure 11

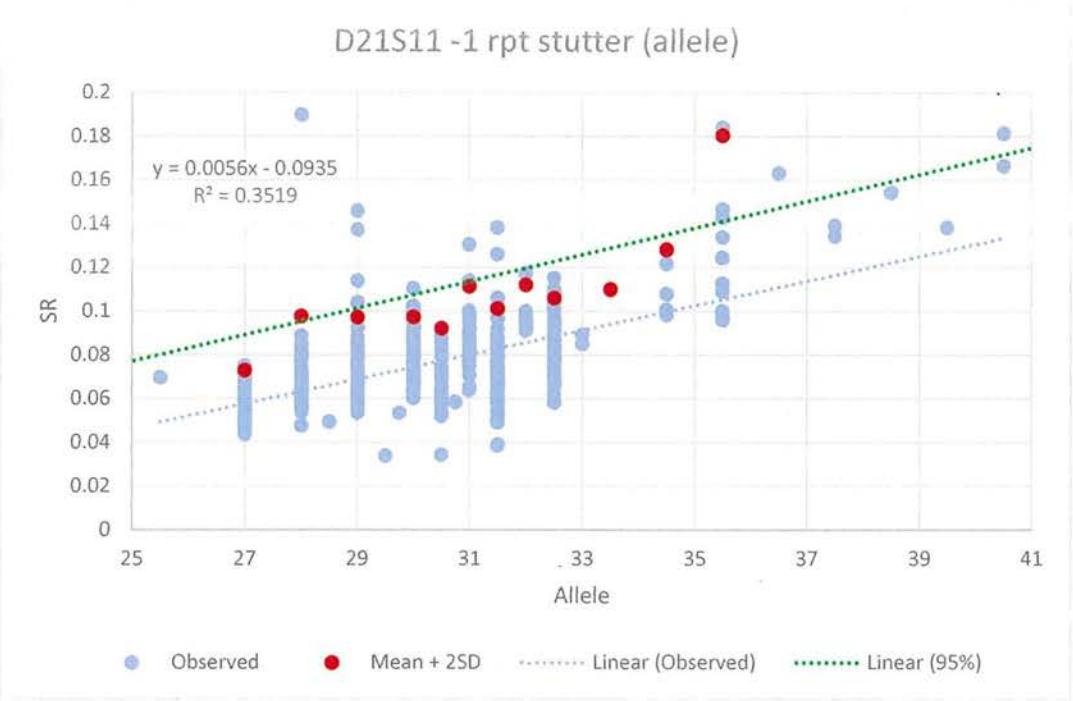


Figure 12

D18S51

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 13).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

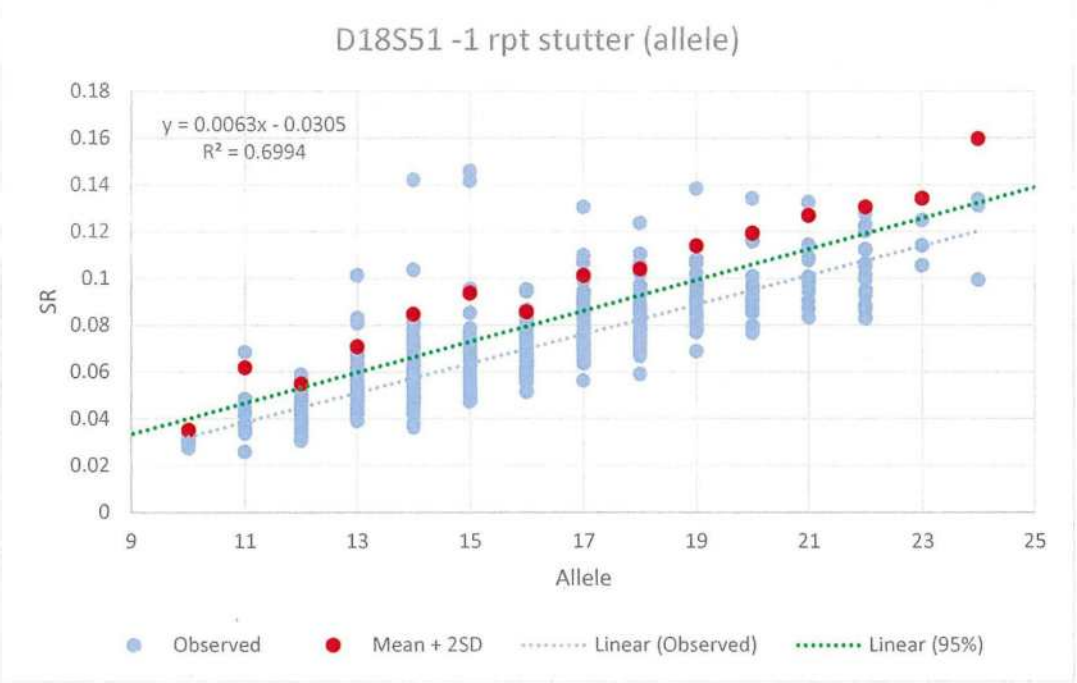

D18S51 -1 rpt stutter (allele)
y = 0.0063x - 0.0305
R² = 0.6994
SR
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
9
11
13
15
17
19
21
23
25
Allele
Observed
Mean + 2SD
Linear (Observed)
Linear (95%)

Figure 13

D5S818

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (

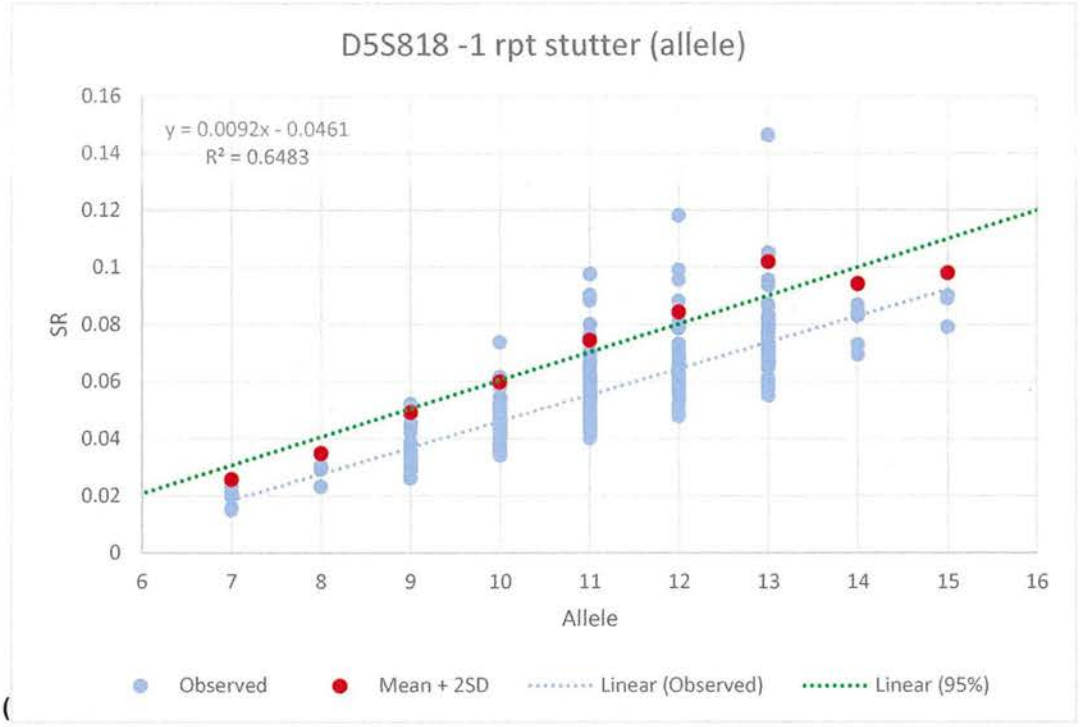


Figure 14).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

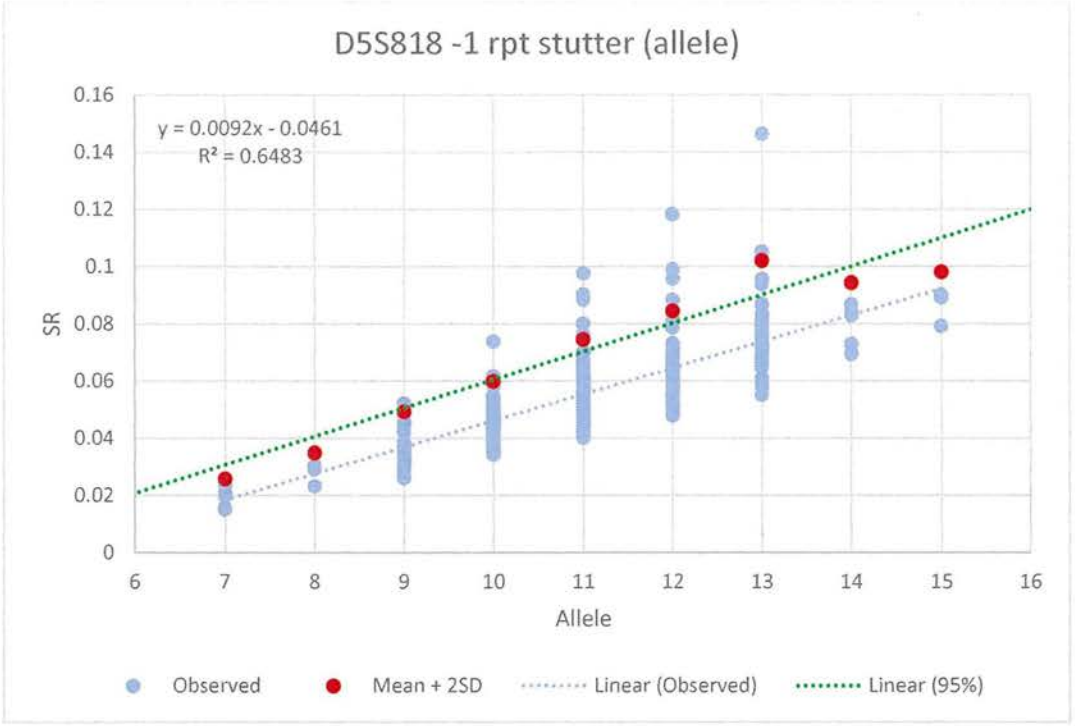


Figure 14

D2S441

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 15).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

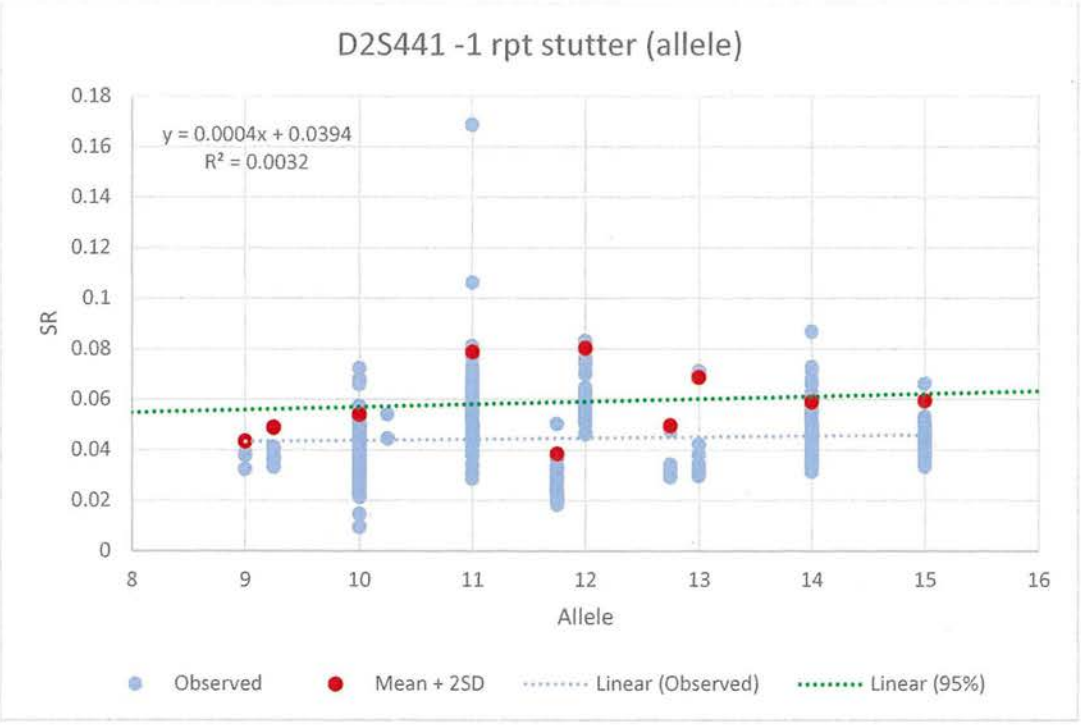


Figure 15

D19S433

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 16).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 17).

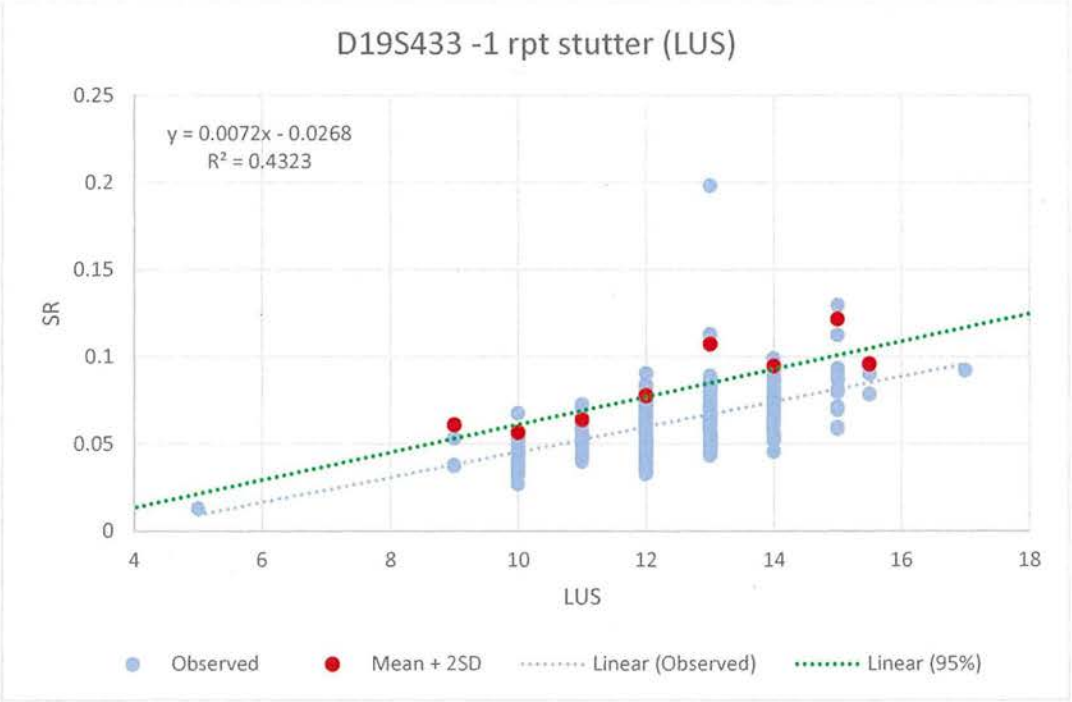


Figure 16

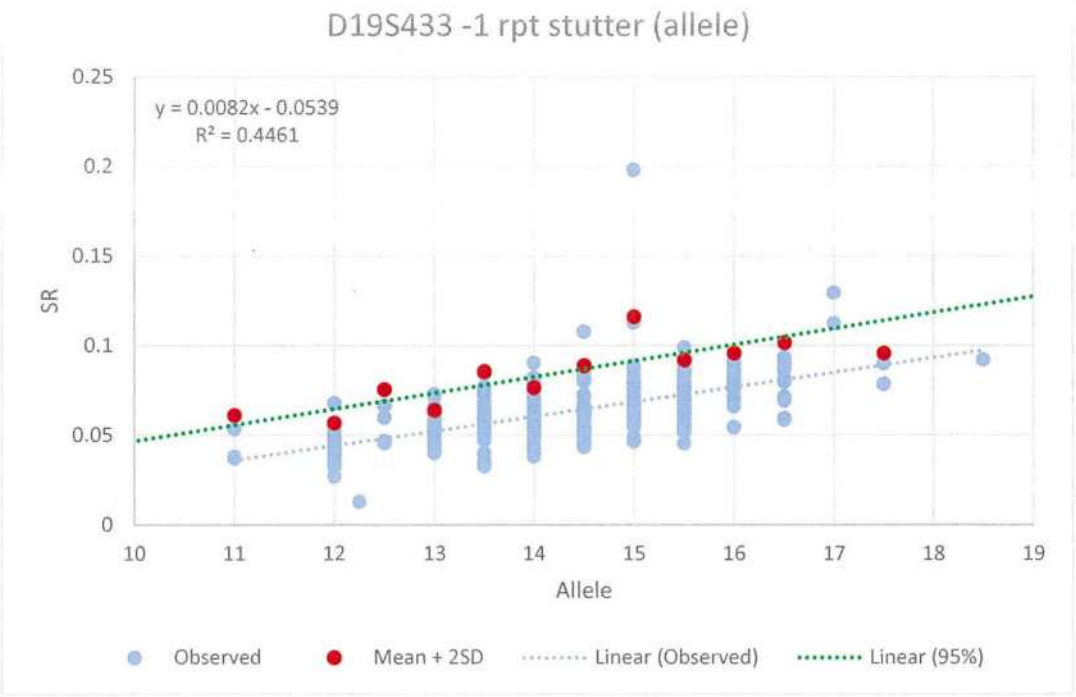


Figure 17

FGA

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 18). It is noted that the intuitive stutter threshold for the LUS value of 16.3 appears very high in comparison to the rest of the data, however using the 95% confidence interval of the regression line for LUS vs SR may result in an intuitive stutter threshold that is too low. If the intuitive stutter threshold is too low the risk is that the number of contributors may be overestimated; this is considered to be a lesser risk than underestimating the number of contributors. Therefore, for the LUS value of 16.3, the 95% confidence interval of the regression line for LUS vs SR will be used.

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 19).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS values of 15.3 and 16.3 are above the regression line value and therefore the average observed stutter ratio for LUS values of 15.3 and 16.3 will be used in the STRmix stutter exceptions file.

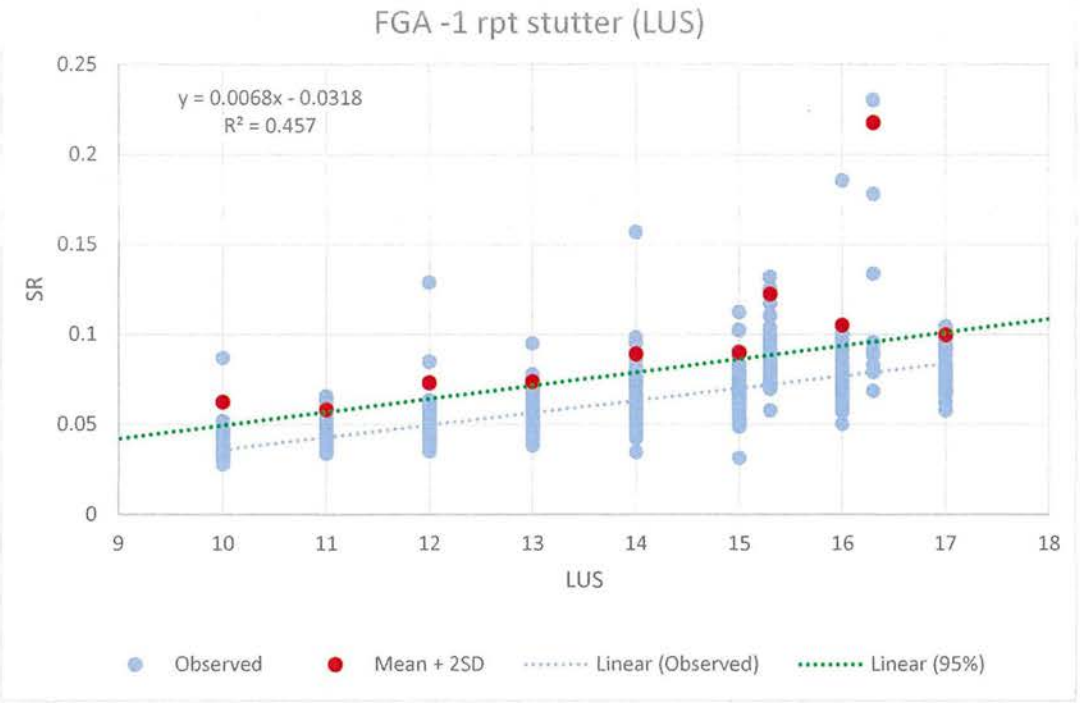


Figure 18

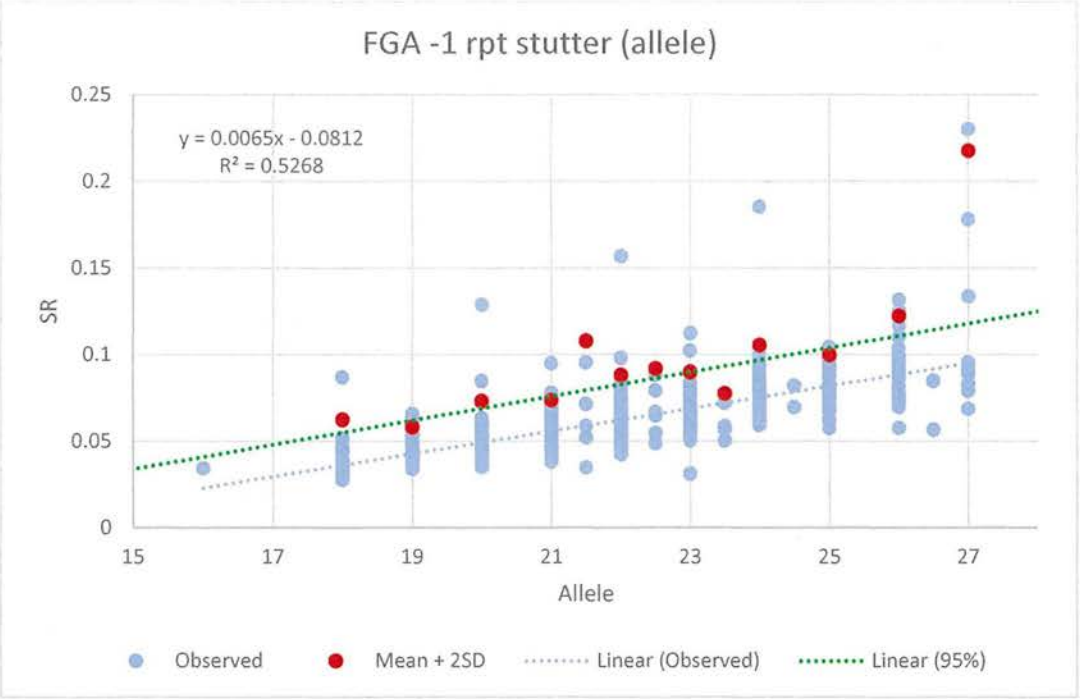


Figure 19

D10S1248

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 20).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

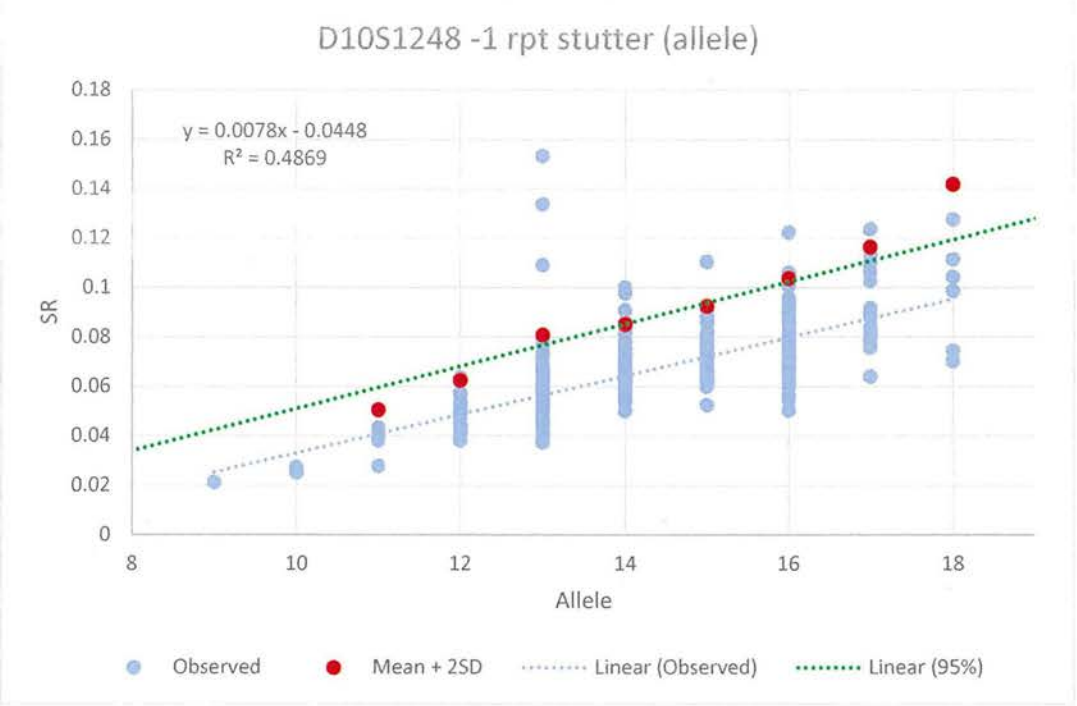


Figure 20

D22S1045

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 21).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

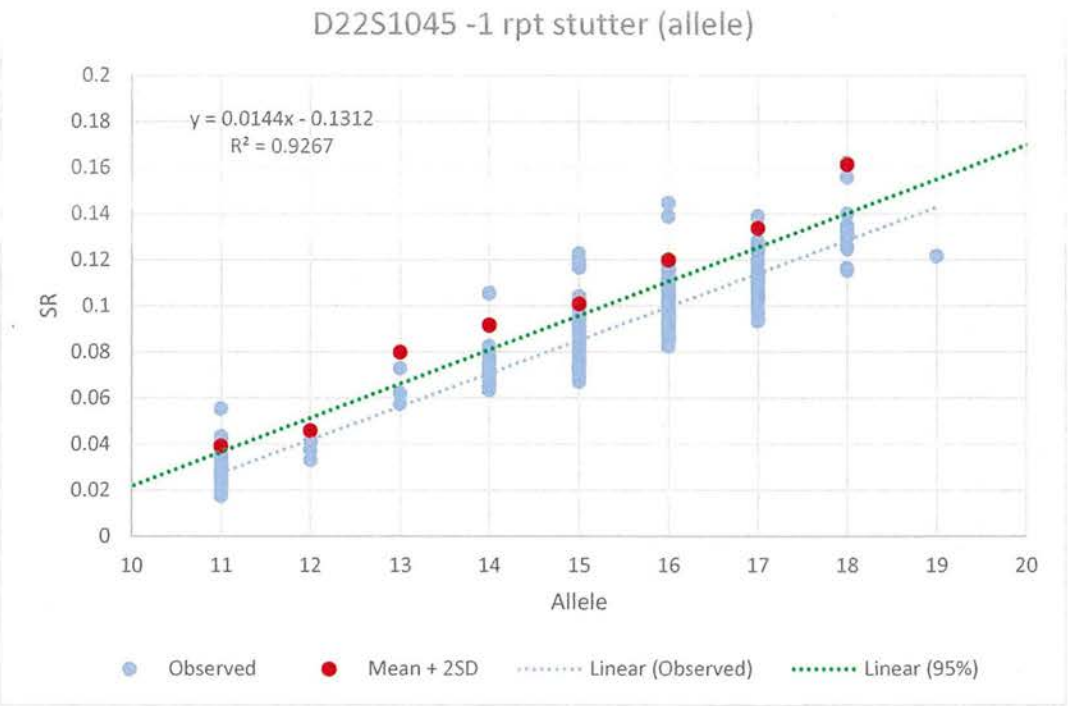


Figure 21

D1S1656

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 22).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 23).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS value of 13 are above the regression line value and therefore the average observed stutter ratio for LUS value of 13 will be used in the STRmix stutter exceptions file.

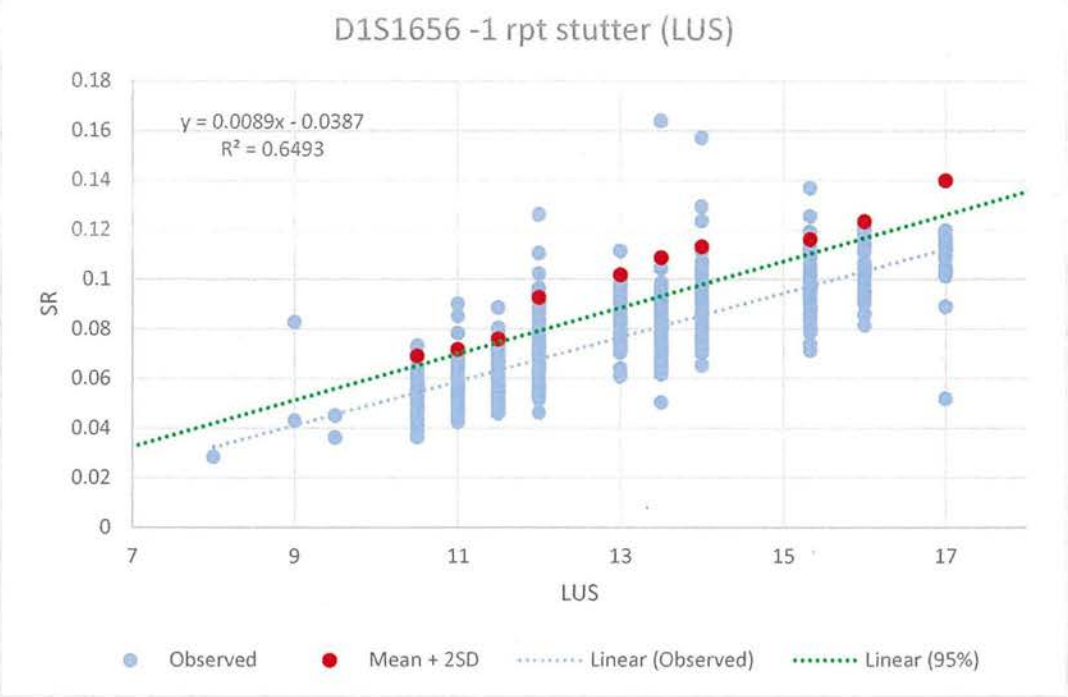


Figure 22

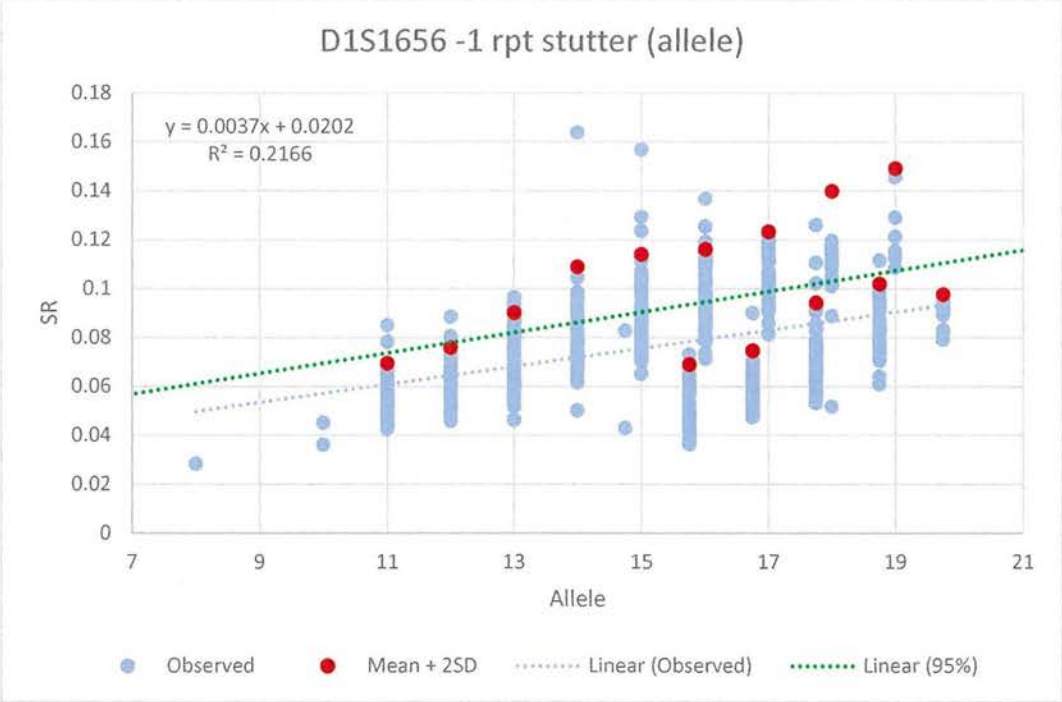


Figure 23

D13S317

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 24).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

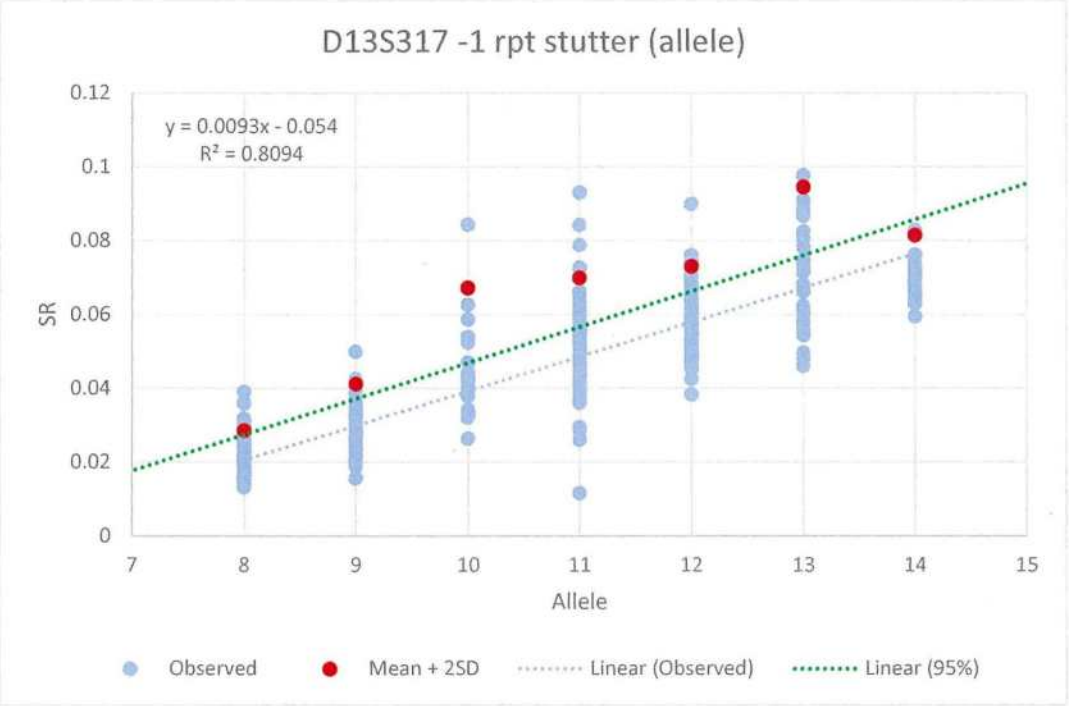


Figure 24

D7S820

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 25). One exception to this is the 7 allele where the mean + 2SD value is skewed by one outlier. For the 7 allele the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

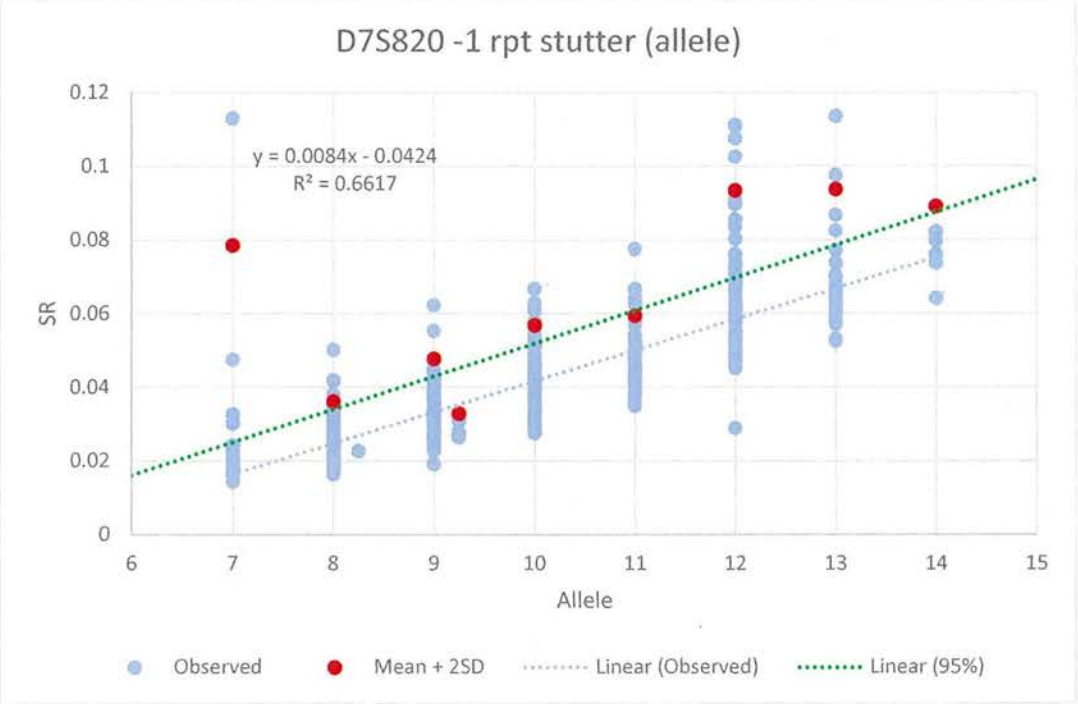


Figure 25

Penta E

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 26). One exception to this is the 26 allele where the mean + 2SD value is skewed by one outlier. For the 26 allele the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

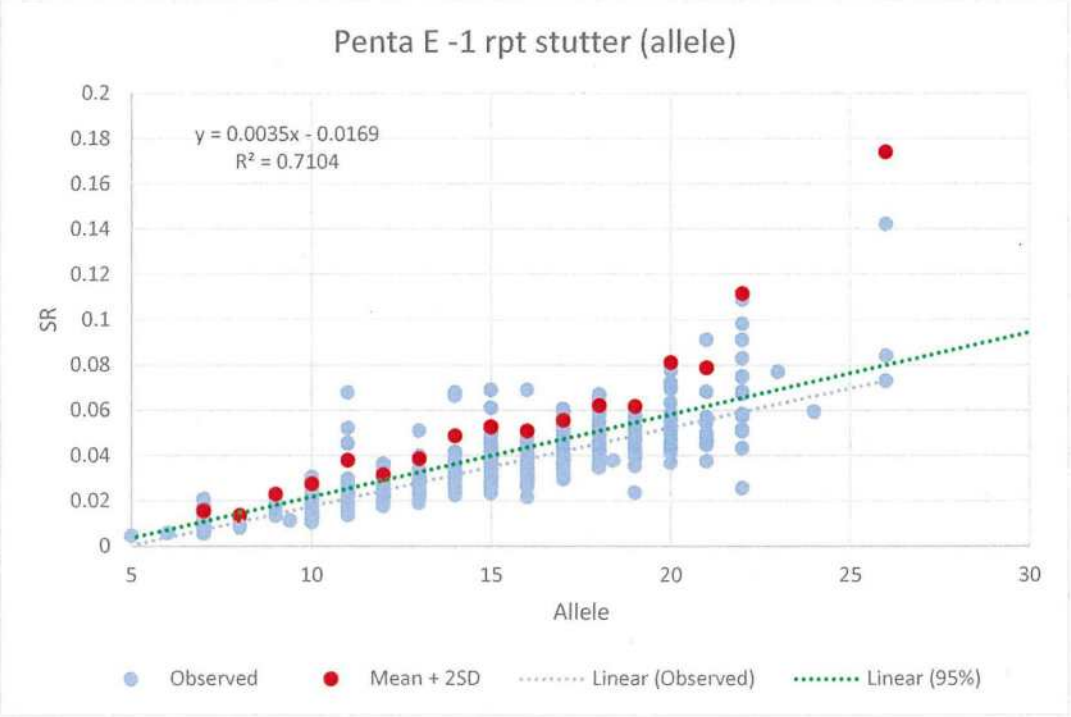


Figure 26

Penta D

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 27). Exceptions to this are the 10 and 15 alleles where the mean + 2SD value is skewed by one outlier. For the 10 and 15 alleles the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

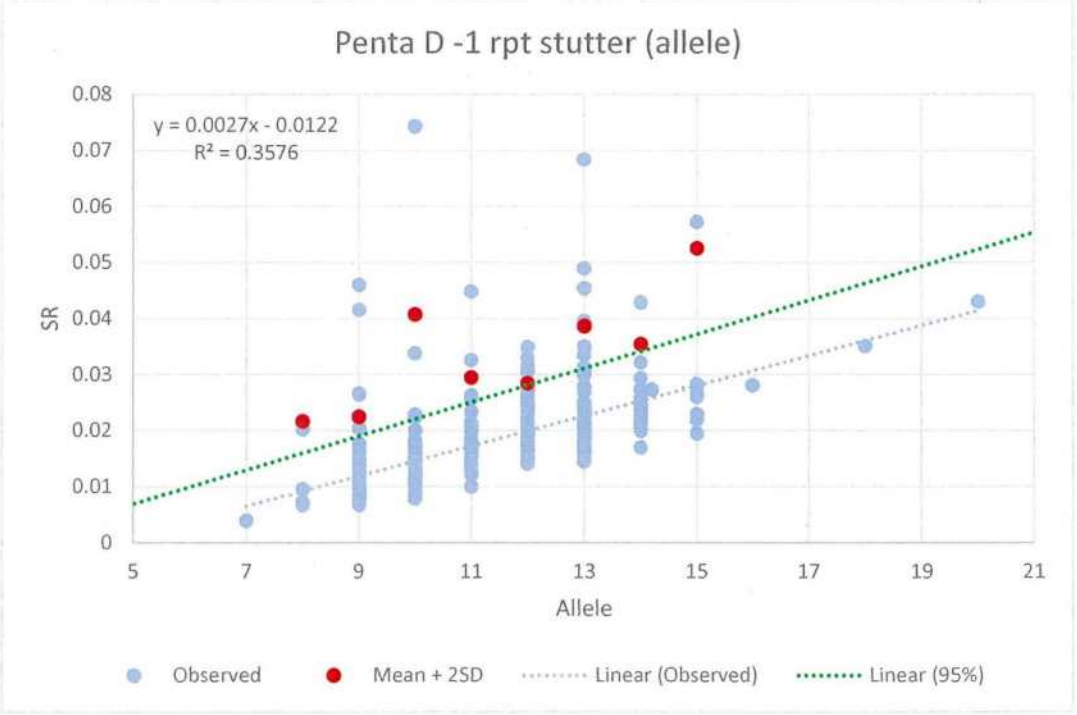


Figure 27

TH01

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 28).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 29).

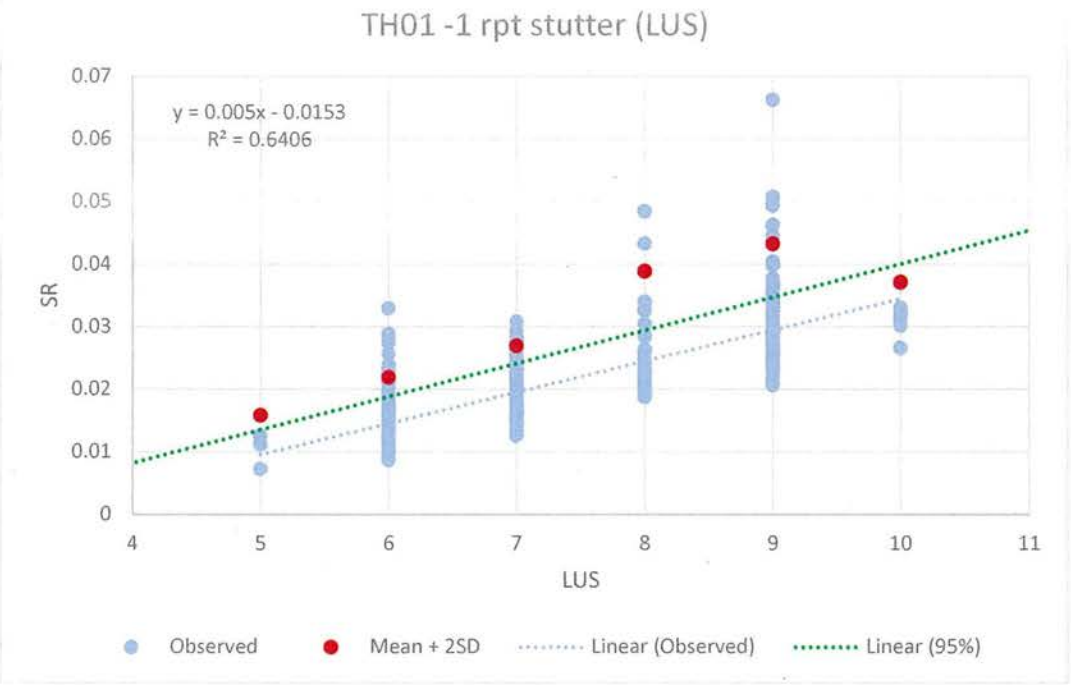


Figure 28

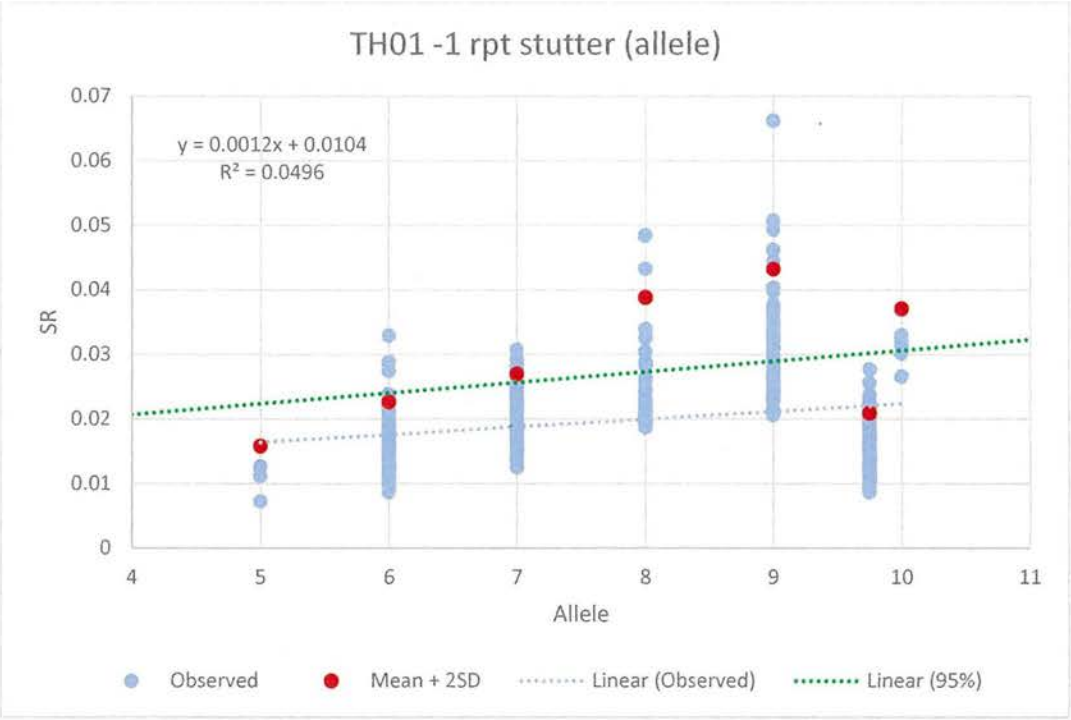


Figure 29

D12S391

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 30).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 31).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS values of 8, 11.286 and 15.2 are above the regression line value and therefore the average observed stutter ratio for LUS values of 8, 11.286 and 15.2 will be used in the STRmix stutter exceptions file.

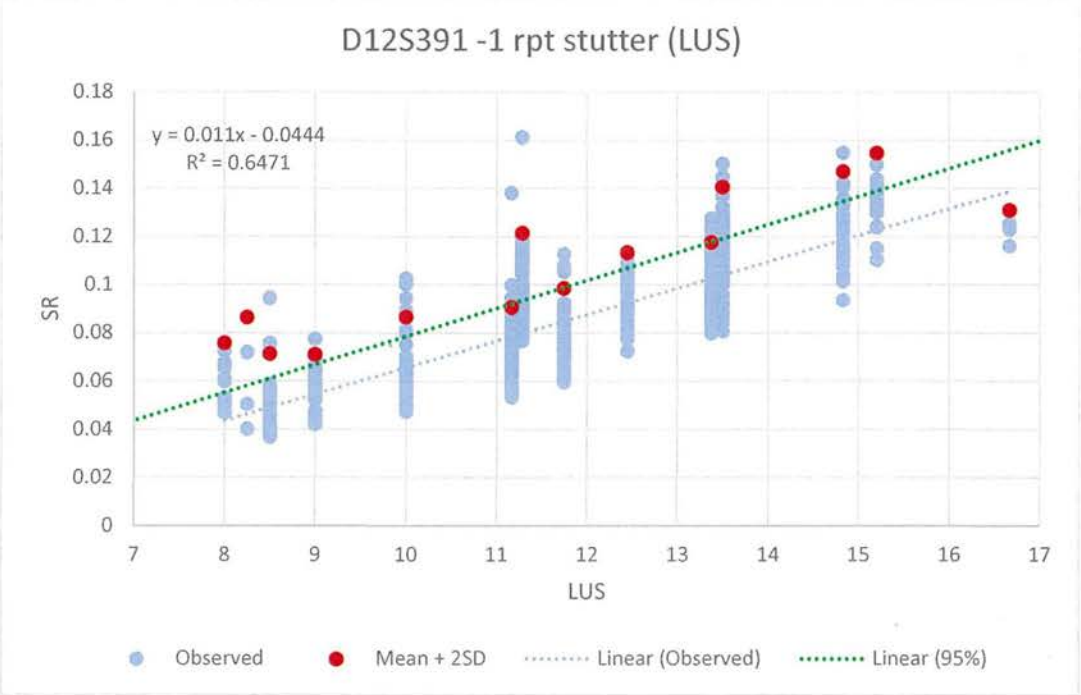


Figure 30

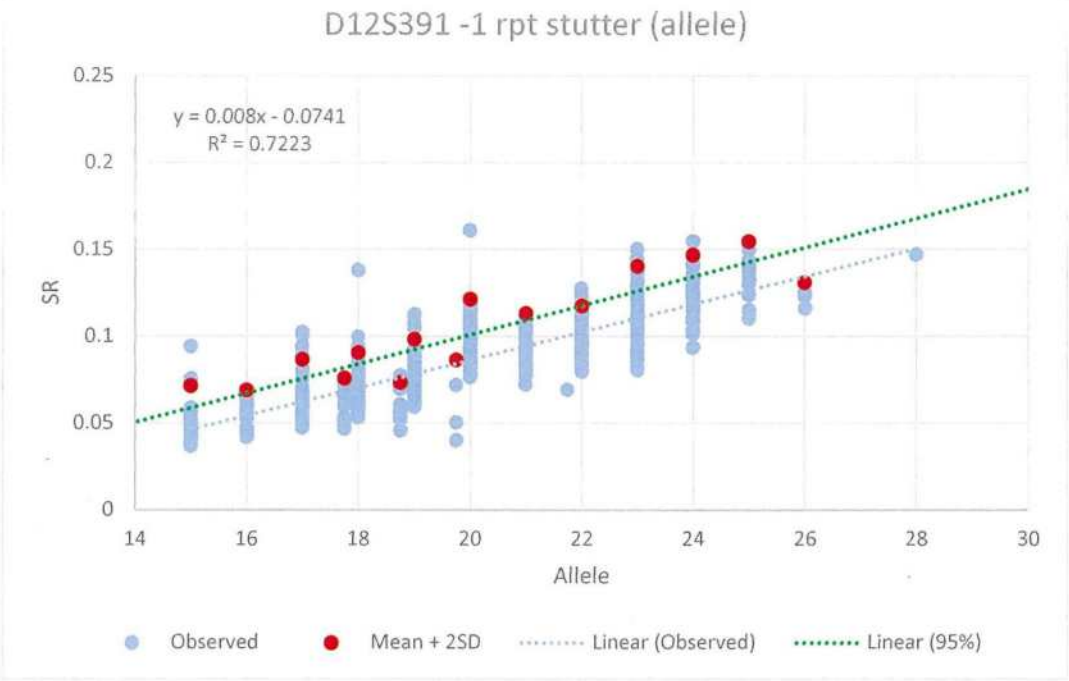


Figure 31

D2S1338

Where the LUS value is known and has been observed within the data, the LUS mean + 2SD value will be used for the intuitive stutter threshold (Figure 32).

Where the LUS value is known but has not been observed within the data, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for LUS vs SR.

Where the LUS value is not known, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR (Figure 33).

Additional STRmix considerations:

The STRmix stutter ratios will be calculated from the observed LUS regression line. The majority of the data points for the LUS values of 11.5, 12.25, 12.5 and 16.5 are above the regression line value and therefore the average observed stutter ratio for the LUS values of 11.5, 12.25, 12.5 and 16.5 will be used in the STRmix stutter exceptions file.

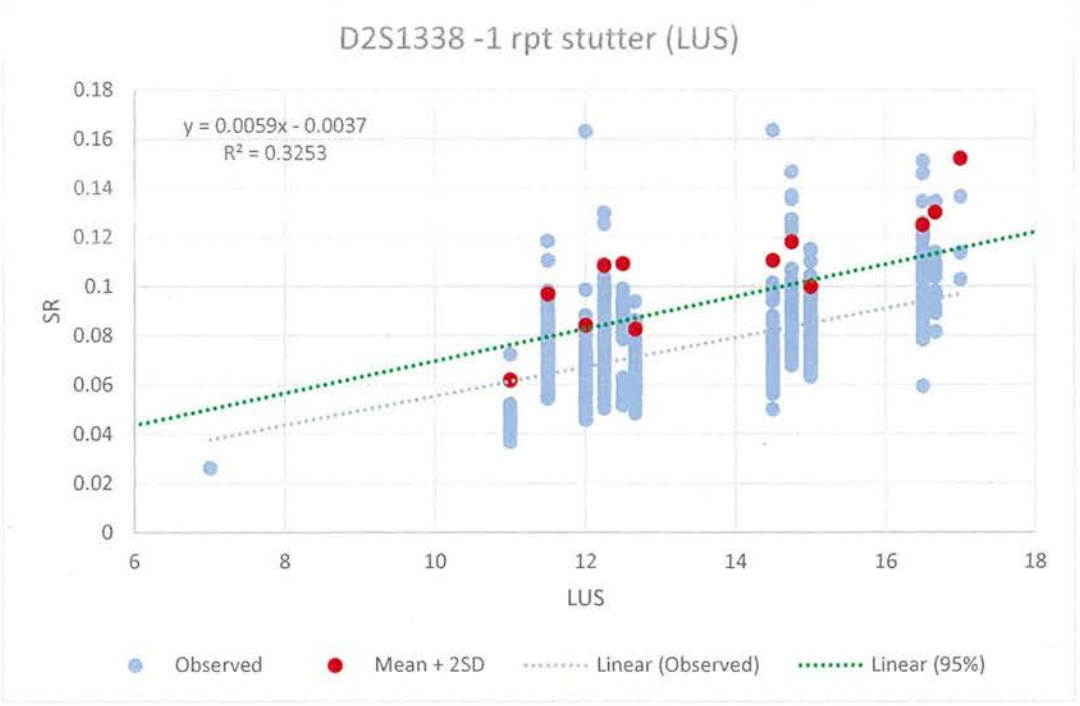


Figure 32

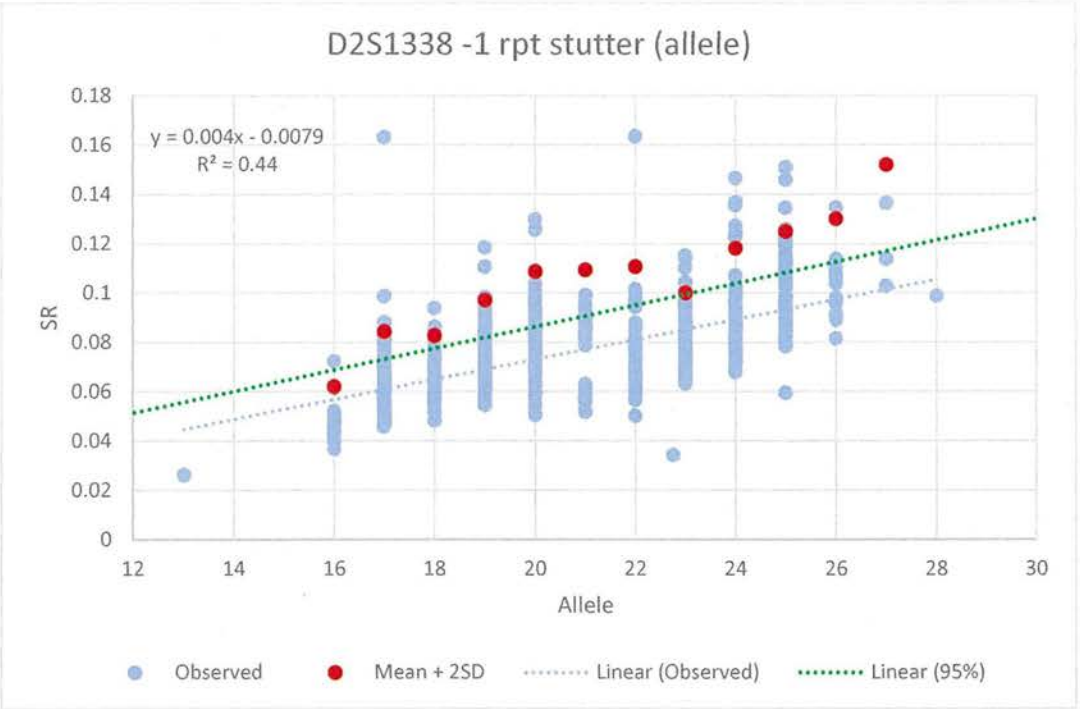


Figure 33

TPOX

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 34).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

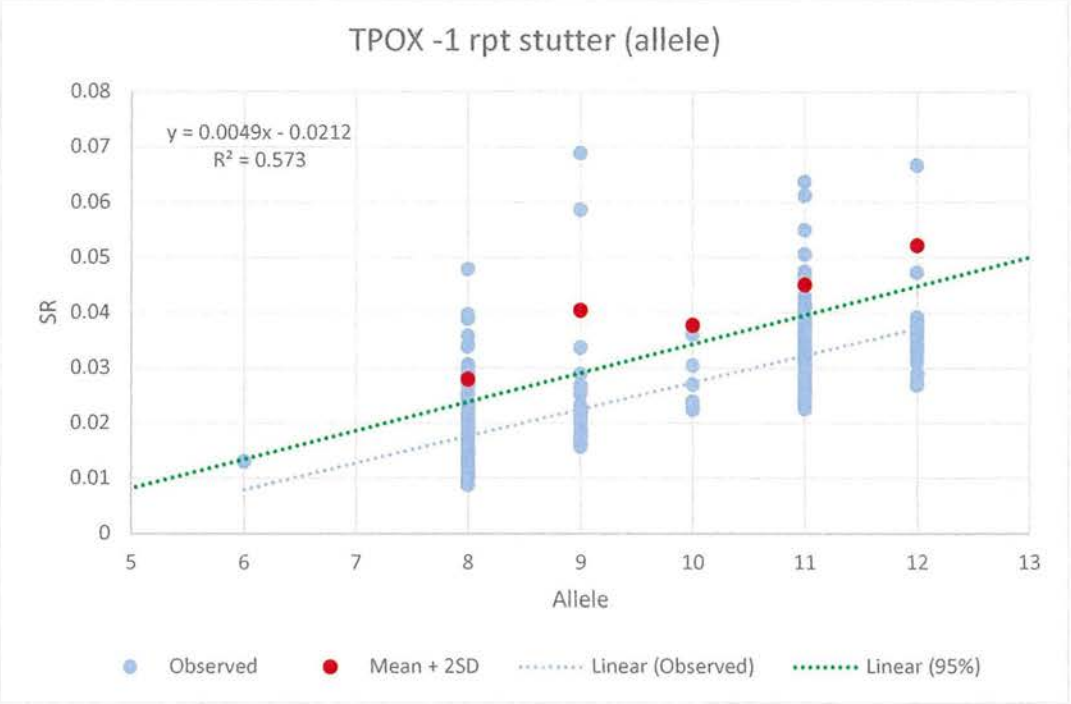


Figure 34

+1 rpt stutter

The observed stutter ratios per allele were plotted for each locus and the regression line was determined. With the exception of locus D22S1045, there was no correlation between the allelic designation and stutter ratio; for this reason, stutter ratios were considered per locus rather than per allele.

The per locus +1 rpt intuitive stutter thresholds were calculated using mean + 2SD (Table 1).

D22S1045 is considered on a per allele basis.

Locus	# data points	Mean +2SD
D3S1358	384	1.61%
vWA	228	2.65%
D16S539	375	2.62%
CSF1PO	414	2.55%
D6S1043	378	2.47%
D8S1179	346	2.76%
D21S11	466	2.75%
D18S51	256	5.89%
D5S818	316	3.04%
D2S441	182	2.15%
D19S433	50	2.52%
FGA	244	2.01%
D10S1248	13	5.52%
D22S1045	673	Per allele
D1S1656	562	2.71%
D13S317	214	2.96%
D7S820	230	2.92%
Penta E	54	5.88%
Penta D	33	2.99%
TH01	13	1.26%
D12S391	163	4.13%
D2S1338	39	3.48%
TPOX	28	1.43%

Table 1

D22S1045

The per allele mean + 2SD value will be used for the intuitive stutter threshold for all observed alleles (Figure 35).

For all non-observed alleles, the intuitive stutter threshold will be calculated from the 95% confidence interval of the regression line for Allele vs SR.

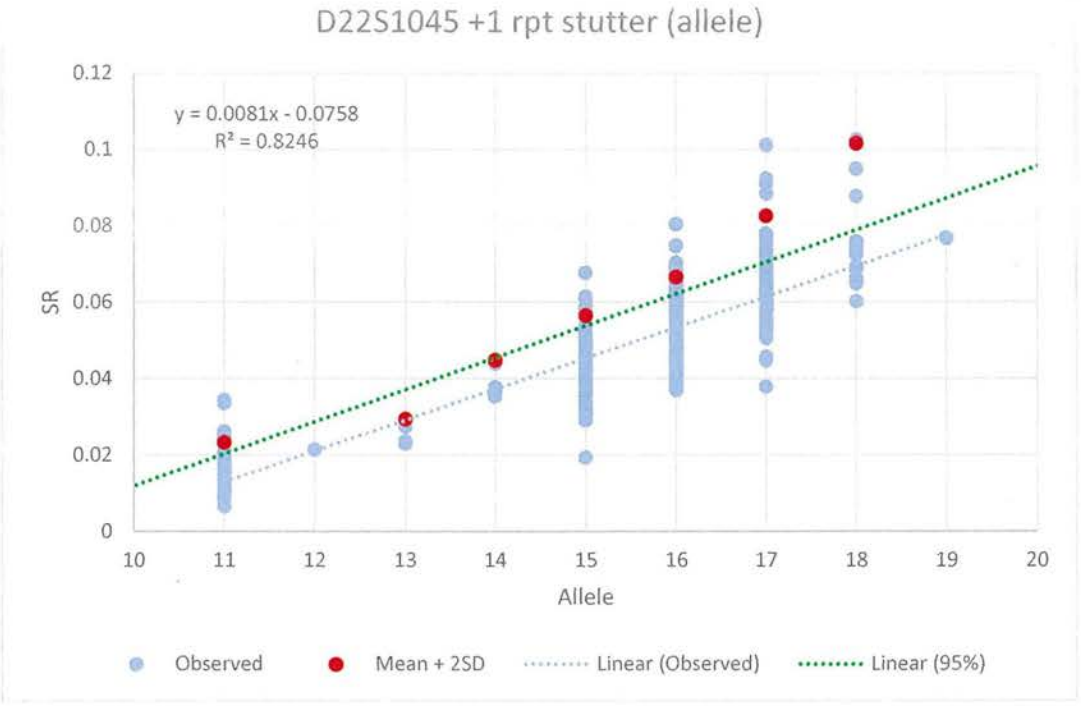


Figure 35

-2 rpt stutter

The observed stutter ratios per allele were plotted for each locus and the regression line was determined. There was no correlation between the allelic designation and stutter ratio; for this reason, stutter ratios were considered per locus rather than per allele.

Loci Penta D and TH01 had very few data points meaning that the incidence of -2 rpt stutter for these loci is low. -2 rpt stutter will not be considered for these loci.

The per locus -2 rpt intuitive stutter thresholds were calculated using mean + 2SD (Table 2).

Locus	# data points	Mean +2SD
D3S1358	203	1.23%
vWA	238	1.34%
D16S539	98	1.13%
CSF1PO	96	1.38%
D6S1043	123	1.27%
D8S1179	116	1.19%
D21S11	195	1.35%
D18S51	158	1.54%
D5S818	96	1.46%
D2S441	41	1.22%
D19S433	129	1.19%
FGA	132	2.04%
D10S1248	75	1.51%
D22S1045	328	1.62%
D1S1656	324	1.54%
D13S317	44	1.57%
D7S820	55	1.68%
Penta E	7	1.94%
Penta D	1	N/A
TH01	3	N/A
D12S391	120	2.29%
D2S1338	214	1.72%
TPOX	10	0.74%

Table 2

Other stutter types

D22S1045

D22S1045 displays a +2 rpt stutter product. Due to the number of observations (n=115) this stutter should be considered in the intuitive interpretation of profiles.

The per locus +2 rpt intuitive stutter threshold was calculated using mean + 2SD.

The +2 rpt stutter threshold for D22S1045 is 1.1%.

D1S1656

D1S1656 displays a -2 bp stutter product (also known as half back stutter and n-2). Due to the large number of observations (n=775) this stutter should be considered in intuitive interpretations. This stutter product is also formed off of the -1 rpt stutter to the parent allele rather than a -6 bp product being formed off of the parent allele. This is demonstrated by the ratio of the -2 bp stutter associated with the -1 rpt stutter being similar to the ratio of the -2 bp stutter associated with the parent allele. STRmix is not able to model a stutter product of a -1 rpt peak and therefore it is recommended that this stutter type not be modelled in STRmix but be removed at plate reading stage.

Since this is a binary threshold in that the stutter type will not be modelled in STRmix, a whole locus threshold using mean + 3SD was calculated.

The -2bp stutter threshold for D1S1656 is 3.4%.

References

1. ESR, STRmix Technical and Scientific Support. 2015. *LUS Look up referenced v3.xlsx*. [online] Available at: <https://support.strmix.com/support/solutions/articles/1000220995-lus-look-up-table> [Accessed 19 January 2021].

RP-11

Angela Adamson

From: Kirsten Scott
Sent: Wednesday, 23 June 2021 2:08 PM
To: Cassandra James; Emma Caunt; Angela Adamson; Rhys Parry
Cc: Paula Brisotto; Justin Howes; Sharon Johnstone
Subject: Query authorship RE: Verifiler Stutter

Cassie, Emma, Angela and Rhys,

Thanks for the extensive analysis, and hard work that has produced this document.
It is incredibly valuable and I appreciate the contribution and hours put in.

I do however feel a little uncomfortable about how we are proceeding with authorship on this one.
Given it is a verifiler document I needed some insight into how this decision on authorship was made.

We have a Verifiler team (those staff are clearly identified), and we have incorporated workshops with other staff due to their expertise and skills in this area.
I think this was a great idea, and it has been worthwhile.

However this document does not contain all verifiler reporting and interpretation sub-project staff, or all staff that were invited to the workshop.
So I am surprised, and am unsure how the decision was made on authorship.

Given that we still have a long was to go on written reports under the Verifiler project banner - I would like some clarity on contributions and authorship, as it will only become more complex as we proceed.
My personal preference would be that all Verifiler reporting and interpretation reports were co-authored by Sharon, Emma and Cassie (as a minimum) so that it is clear that you support the document as written.

If we incorporate other staff I would appreciate an explicit discussion on this.

Kirsten

From: Cassandra James <[redacted]>
Sent: Wednesday, 23 June 2021 9:40 AM
To: Sharon Johnstone <[redacted]>; Kylie Rika <[redacted]>; Allison Lloyd <[redacted]>; Thomas Nurthen <[redacted]>; Justin Howes <[redacted]>; Allan McNevin <[redacted]>; Kirsten Scott <[redacted]>; Paula Brisotto <[redacted]>; Luke Ryan <[redacted]>
Cc: Angela Adamson <[redacted]>; Emma Caunt <[redacted]>; Rhys Parry <[redacted]>
Subject: Verifiler Stutter

Hi All,

Please find attached the following document that summarises our suggested Verifiler stutter thresholds. Please read and provide any feedback to these decisions as soon as possible as we need to decide the stutter thresholds before we can move on with any other Verifiler interpretation.

Please reply feedback to Emma, Angela, Cassie and Rhys by Monday 28th June 2021

Many thanks
Cassie, Angela and Rhys

Cassandra James
Scientist – Forensic Reporting and Intelligence Team

Forensic DNA Analysis, Forensic & Scientific Services
Prevention Division, Queensland Health

Please note that I may be working from a different location during the COVID-19 Pandemic. The best contact method is via email.

p [redacted]
a 39 Kessels Road, Coopers Plains, QLD 4108
e [redacted] **w** **www.health.qld.gov.au/fss**

Queensland Health acknowledges the Traditional Owners of the land, and pays respect to Elders past, present and emerging.