Author's response to reviews

Title: Analysis of multiple single nucleotide polymorphisms closely positioned in the ovine PRNP gene using linear fluorescent probes and melting curve analysis

Authors:

David J French (david.french@lgc.co.uk)
Dominic Jones (dominic.jones@lgc.co.uk)
David G McDowell (david.mcdowell@lgc.co.uk)
Jim A Thomson (jim.thomson@lgc.co.uk)
Paul G Debenham (paul.debenham@lgc.co.uk)

Version: 3 Date: 1 June 2007

Author's response to reviews: see over
Dear Sir/Madam,

RE: MS 1950076591120625

Analysis of multiple single nucleotide polymorphisms closely positioned in the ovine PRNP gene using linear fluorescent probes and melting curve analysis.

David J French, Dominic Jones, David G McDowell, Jim A Thomson and Paul G Debenham

Again, many thanks for providing the constructive comments from the two reviewers. A revised manuscript has been uploaded and we address each of the reviewers concerns in turn below.

Version 3 of the above manuscript was submitted with one figure and four tables. Please also note the change of manuscript title.

Yours faithfully,

Dr David French Ph.D.

Reviewer 1: Gabriele Vaccari

Discretionary Revisions (which the author can choose to ignore)

We would rather not state detection of wild-type M137 and P168 since these do not affect the melting temperatures of probes designed to interrogate codons 136 and 171 respectively. If we were to include M137 and P168, we would equally have to mention G137, S135, S138, D150, R151, Y152, Y153, N156, M157, V169, D170, Y172, S173, and N174 which are covered by diagnostic probes but do not affect melting temperature. Furthermore, we cannot mention P168 since the 171 probe only covers two of the three bases of codon 168.

Reviewer 2: Bertram Brenig

Major points

1. Have removed “High-throughput genotyping” from the manuscript title, changing it to:

   “Analysis of multiple single nucleotide polymorphisms closely positioned in the ovine PRNP gene using linear fluorescent probes and melting curve analysis.”

2. The following sentence was removed from the abstract background section:

   “A key feature of melting curve analysis employing such linear probe systems is that a single fluorescent oligonucleotide is capable of simultaneously interrogating multiple sequence variations within homozygous and heterozygous genotypes.” [Page 2 of version 2]

   The following sentences were also removed from the introduction since the technology is described in referenced papers.
“The HyBeacon probe technology utilises fluorescent-labelled oligonucleotides to detect and identify specific nucleotide sequences post-amplification. Probes comprise specific linear oligonucleotides that possess fluorescent dye labels attached to internal nucleotides and a 3′ phosphate to prevent extension during polymerase chain reaction (PCR) based amplification.” [Page 6 of version 2]

“The advantage of employing HyBeacons for homogeneous sequence analysis derives from their ability to reliably identify homozygous and heterozygous samples using a single oligonucleotide probe in a single reaction vessel.” [Page 6 of version 2]

The advantage of employing dual-labelled probes will be published separately in the book “Molecular Beacons - Signalling Nucleic Acid Probes” (in press) which will appear in the Methods of Molecular Biology Series. The following sentence was added to describe the advantage of dual-labelling:

“Dual-labelled probes, possessing two fluorescent dye labels attached to internal nucleotides, exhibit considerably larger signal-to-noise ratios and melting peaks compared with single-labelled probes of identical sequence [23].” [Page 6]

The following sentence was modified:

“Furthermore, multiple polymorphisms may be analysed using a single probe, enabling the presence of additional polymorphisms flanking PRNP codons 136, 141, 154 and 171 to be detected through distinct shifts in probe Tm.” [Page 6]

3. a) We did not spend enough time developing a Tm-multiplex assay to find out which SNPs could cause problems. We did not employ a Tm-multiplex format for assay validation and have not included such information in the manuscript. The reviewer’s comment appears to refer to the cover letter of version 2 rather than the manuscript.

b) As mentioned in the revised manuscript, the bottle-neck for analysis was the rate at which we could prepare PCR plates. Each 384 well plate takes approximately 45 minutes to prepare using the Hamilton robot. Whilst preparation of 96 well plates will be slightly quicker, preparation of four 96 well plates would increase setup time. Furthermore, since internal genotype standards and negative controls are required for each PCR plate, the number of samples that could be analysed per day would be reduced considerably. No further changes made to the manuscript.

c) The 30 minute assay duration detailed on page 19 of version 2 referred to the capillary based LightCycler, not the LightCycler 480. This sentence has been clarified:

“HyBeacon assays may be performed in both high-throughput and ultra-rapid formats, where assays performed in glass capillaries using a LightCycler instrument are completed within 30 minutes.” [Page 21]

The capability of the LightCycler 480 has also been demonstrated for 136, 154 and 171 assays. We have trialed the instrument twice, once before the genotyping software was available and once after. Run times for PCR and melting curve analysis were approximately 45 minutes. Colour-multiplexing with HyBeacon probes has been demonstrated using the LightCycler 480, but not for PRNP genotyping as yet.

4. Two versions of the LightTyper software are employed for analysis. The newer version (1.5) of the software is faster and easier for analysts to use, but the peak calling ability was reduced
compared with the older version (1.1). Reanalysis of data to generate melting peak Tm and area data is extremely rapid, taking less than 1 minute per plate.

5. Have removed all but one reference to “complex genotypes” from the manuscript, instead describing the imbalanced peak data that may arise from chimeras and additional polymorphisms in target sequences. The remaining reference to “complex genotypes” was required since both Dawson et al., the Institute of Animal Health (IAH) and DEFRA refer to them as such, see:

http://www2.defra.gov.uk/research/project_data/More.asp?I=SE0245&M=CFO&V=IAH

In detail, the changes made to the manuscript include:

- Additional reference included [23] detailing DEFRA’s description of “complex genotypes”
- Removed “and complex genotypes” from the results section of the abstract – [Page 2 of version 2]
- Changed the sentence: “Peak area ratios were employed to identify heterozygous samples possessing a complex genotype [23], comprising more of one allele than another (see below).” [Page 9] to: “Peak area ratios were employed to identify heterozygous samples that potentially possessed more than two copies of the PRNP gene [22,23].”
- Changed the title of the “Complex genotypes” results section to “Imbalanced melting peaks” [Page 14]. Major changes made to this section.
- Included the following sentence to detail the frequency of imbalanced melting peaks: “Our study of 8918 purified DNA samples was enriched for samples that had previously generated imbalanced peak data in MALDI-ToF analyses. Whilst 72 samples exhibited imbalanced MALDI peaks, only 43 heterozygous samples yielded imbalanced melting peaks.” [Page 14-15]
- Removed “generating ARR/ARR/ARQ and ARQ/ARH/ARH complex calls by MALDI and ARR/ARLQ and ARLQ/Q/ARH genotypes in HyBeacon tests.” [Page 16 of version 2]
- Changed: “As a result, fewer complex genotype calls were obtained in HyBeacon tests compared with hME and MALDI-ToF analysis of the same sample set.” to: “Fewer samples yielded imbalanced peak data in HyBeacon tests compared with hME and MALDI analysis of the same sample set, thereby reducing the number of samples requiring reanalysis.” [Page 18]
- Removed “complex genotypes, with” [Page 18 of version 2]
- Removed “Furthermore, samples possessing putative complex genotypes are identified by measuring the area ratio of heterozygous melting peaks.” [Page 20 of version 2]

We have not been able to state, as the reviewer suggests, that all abnormal results “can be resolved using tissue samples”. Both LGC and the other major supplier of scrapie genotyping services in the UK have observed “complex genotypes” with multiple tissue types. Since we cannot refer to our unpublished data, we have instead included the following sentence:

“Analyses employing alternate sample types, such as hair roots, ear tissue plugs and saliva, are expected to resolve a large proportion of the abnormal results obtained from blood.” [Page 14]

Samples that yielded imbalanced peak data were swapped, between LGC and the other supplier of scrapie genotyping services in the UK, for confirmation and “complex genotypes”
were reported to the National Scrapie Plan (NSP). Whilst the NSP continues to employ blood samples for PRNP analysis, genotyping methods must be capable of detecting putative “complex genotypes” through peak imbalances. Furthermore, analysis of tissue samples will not resolve the problem of imbalanced melting peaks arising from the presence of additional polymorphisms underlying primers and probes.

6. Have considerably reduced the number of figures in the manuscript as suggested, retaining only one example of melting peak data:
   - Removed the original figure 1. Retained the original figure 2 since this clearly presents melting peaks for A136 and V136 along with T137 and T136.
   - Removed figures 3-8

**Minor Points**

1. This section was not intended to be speculative, where “may be employed” refers to an alternate method that has been investigated within our laboratory. This section has been updated to avoid confusion:

   “Furthermore, since each of the HyBeacon tests employs the same primer pair, a single PCR plate can be prepared for each batch of samples, without inclusion of probes or Chill-Out liquid wax. Daughter plates prepared robotically post-amplification will mix a portion of the target amplicon with a specific HyBeacon probe. This approach considerably increases the number of samples that may be analysed each day.” [Page 20]

Have removed reference to “high-throughput” from the manuscript title, but have retained it within abstract and discussion sections. The method described in our manuscript was performed in a laboratory actively genotyping sheep samples. The robotic systems required for high-throughput analysis were employed for assay development and test validation. We are capable of analysing up to 1440 samples per day (i.e. in excess of 250,000 – 350,000 samples per year) with our current equipment, and consider this as high-throughput. Our capacity for sample analysis could also be increased if we acquired more robotic platforms or adopted strategies discussed in the manuscript. Whilst it is difficult to define what constitutes “high-throughput” analysis, many articles are still published with “high-throughput” and “high-throughput genotyping” in the title.

2. N/A

3. Added a new results section describing “Assay performance” [Page 15].
   - Added “The performance of HyBeacon PRNP assays was assessed through determination of first time pass rates (FTPFR) and concordance with parallel homogeneous MassEXTEND™ and MALDI-ToF analysis.” [Page 15]
   - Moved “Samples that contributed to the FTPFR generated high quality peak data in 136, 154 and 171 assays. Samples were treated as “fails” if melting peaks were absent from any one of the tests or if software/analyst genotype calls were discordant with exported spreadsheet data. Samples were also excluded from FTPFR if they exhibited imbalanced melting peaks.” from the discussion section to page 15.
   - Clarified that “The average FTPFR for 384 well plates was 97.8% (standard deviation = 1.23) with purified DNA samples.” [Page 15]. Since the pass rate was determined using only samples (i.e. not controls), the pass rate per plate is identical to the pass rate per sample.
• Added “HyBeacon and MALDI-ToF genotype calls were 99.8% concordant. The 17 DNA samples that yielded incongruent calls were sequenced and demonstrated to possess leucine at codon 168, generating ARR/ARL\textsuperscript{168}Q and ARL\textsuperscript{168}Q/ARH genotypes in HyBeacon tests. These samples yielded imbalanced peaks in MALDI analyses and were reported as ARR/ARR/ARQ and ARQ/ARH/ARH respectively [22].” [Page 15]

• Moved “Analyses employing unpurified blood samples were extremely efficient yielding high quality melting peaks with the majority of samples, demonstrating first time pass rates between 85% and 95% (mean = 91.2%). Pass rates using unpurified blood were considerably more variable (standard deviation = 3.95) than PCR plates employing purified DNA samples.” from the discussion to page 16.

• Added “HyBeacon and MALDI-ToF data was 99.9% concordant, where the 4 DNA samples that yielded incongruent calls possessed leucine at codon 168. Blood samples yielded elevated numbers of weak positives with melting peaks of reduced height. Our study of 3663 unpurified blood samples was not enriched for samples that had previously generated imbalanced peak data in MALDI-ToF analyses. Whilst 7 samples yielded imbalanced MALDI peaks (with extracted DNA), 153 heterozygous samples exhibited imbalanced melting peaks in at least one assay.” [Page 16]

• Additional descriptive statistics included in the manuscript and described elsewhere in this cover letter.

4. 141 probe data now included in table 2.

Probe Tms were predicted using a combination of Primer Express software (version 3.0) and an in-house algorithm based on the Breslauer et al. (1986) paper, including a correction to account for the Tm reduction caused by the fluorescein dT monomers. We know that this may not be the best set of thermodynamic data to use, but it does provide us with an estimate of probe Tm as reliable as Primer3 and Primer Express. This paper is probably not the best place to discuss the design of HyBeacons and the prediction of probe Tm. The following sentence has, therefore, been amended:

“Probes exhibited $\Delta$Tms greater than 5°C, enabling reliable detection and differentiation of A$^{136}$, V$^{136}$, L$^{141}$, F$^{141}$, R$^{154}$, H$^{154}$, R$^{171}$, Q$^{171}$ and H$^{171}$ melting peaks using the LightTyper software.” [Page 10]

5. Additional information supplied concerning the Cypriot sheep samples that were employed to analyse T136 and K171 alleles:

Added “Identification of the T$^{136}$ melting peak was demonstrated using Cypriot sheep samples of ARQ/TRQ and TRQ/TRQ genotype. Differentiation of V$^{136}$ and T$^{136}$ peaks was demonstrated using a mixture of VRQ/VRQ and TRQ/TRQ DNA samples.” [Page 11]

Removed “Cypriot sheep samples that were heterozygous for alanine/threonine and valine/threonine were clearly identified through shifts in peak Tm.” [page 11 of version 2]

Added “Identification of the K$^{171}$ melting peak was demonstrated using Cypriot sheep samples of ARR/ARK, ARQ/ARK and ARK/ARK genotype. Differentiation of H$^{171}$ and K$^{171}$ peaks was demonstrated using a mixture of ARH/ARH and ARK/ARK DNA samples.” [Page 13]
We did not refuse to disclose the results for other genotypes requested by reviewer 1. Assay results obtained with the requested rare alleles (detailed in the discussion of version 1) were presented in table 3 of version 2. This information has been updated and is included as table 4 in the current version (see below).

Failure to detect a rare polymorphism may not be a real drawback if, for example, a very small number of sheep are mistakenly slaughtered or employed for breeding. However, a rare SNP in one population may not be rare in another. For example, whilst not required for studies of UK flocks, analysis of Cypriot sheep populations may require detection of T137 and K171.

6. The reviewer’s original comment referred to the acceptance criteria employed for identification of imbalanced heterozygotes (page 14 of version 1), based on peak area ratio. The reviewer’s comment for version 2 now appears to refer to the acceptance criteria employed for peak identification (table 2), based on peak Tm. We will address both concerns here.

To avoid further confusion we have modified the column “Acceptance criteria” of table 2 to “Tm acceptance criteria” and have added a new table (table 3) to detail peak area ratio data.

Modified the following section: “Acceptance criteria for normal, balanced, heterozygotes were determined using the area ratios obtained from 4560 purified DNA samples (i.e. twelve 384 well plates, excluding negative controls). Acceptance criteria for peak area ratios (table 3) were determined using only A\text{136}/\text{V}\text{136}, R\text{154}/H\text{154}, R\text{171}/Q\text{171}, R\text{171}/H\text{171} and Q\text{171}/H\text{171} genotypes, excluding data generated in the presence of additional rare polymorphisms (e.g. R\text{171}/L\text{168}Q\text{171}). Samples that exhibit imbalanced melting peaks and area ratios that are outside of defined limits must be repeated for confirmation.” [Page 14]

The method employed to determine peak area acceptance criteria (i.e. mean ± 3 standard deviations) is now discussed in the description of table 3: “4560 purified DNA samples were employed to determine the acceptance criteria for peak area ratios. The peak area ratios of heterozygous samples were calculated by dividing the area under the mismatched melting peak by the area under the matched melting peak. N is the number of heterozygous samples and SD is the standard deviation. Acceptance criteria for normal balanced heterozygotes were determined as the mean area ratio ± 3 standard deviations. Samples exhibiting imbalanced melting peaks generated area ratios outside of defined limits.” [Page 25].

The acceptance criteria employed for peak detection and identification (table 2) were generally based on the mean ± 2°C, rather than mean ± 3 standard deviations. We have made a couple of changes to table 2 based on validation data:

- Have corrected the lower limit for H154 to 54.6°C. All H154 Tms were greater than this value.
- Have not changed the upper limit of H154 as 6 of 278 samples were between 58.6°C and 58.7°C. Changing the upper limit for this peak would reduce the overall pass rate to 97.7%.
- Have corrected the upper limit for R154 to 51.9°C. All R154 Tms were lower than this value.
- The acceptance criteria for Q171 were narrower than ±2°C due to the neighbouring L168 polymorphism. Acceptance criteria for Q171 are still greater than ±3 standard deviations.
Removed: “where sample-to-sample variation was typically less than ±1°C. Genotypes were generally assigned if peak Tms deviated from expected values by less than ±2°C (table 2).“ [Page 10 of version 2]

Added: “The major PRNP alleles were typically identified if peak Tms deviated from expected values by less than ±2°C. The acceptance criteria employed for the detection and identification of Q171 melting peaks were narrower than the mean Tm ± 2°C due to the neighbouring L168 polymorphism. In each case, the acceptance criteria for the major PRNP alleles exceeded the mean ± 3 standard deviations, ensuring robust detection of melting peaks.” [Page 10]

The description for table 2 has been expanded to: “The mean Tms for each probe melting peak are presented, where N is the number of peaks investigated and SD is the standard deviation. The peak Tms for the major PRNP alleles were obtained from six 384 well plates prepared on different working days, thereby accounting for sample-to-sample and run-to-run sources of variation. Variant alleles exhibiting uncommon genotypes or additional SNPs are italicised. Acceptance criteria for these rare polymorphisms were estimated from the few samples analysed, minimising the potential for miscalling melting peaks.” [Page 25]

7. N/A

8. N/A

9. No additional changes made to the document. We have suggested a variety of methods to neutralise unwanted polymorphisms and didn’t intend describing all possible methods available. Whilst universal bases do increase the cost of probe synthesis, they only increase the cost of analysis by less than £0.003 per sample. Furthermore, in the absence of programs that thermodynamically account for “bulging”, it may be easier to neutralise unwanted SNPs using base analogues.

Additional Major points in version 2:

1. The data originally presented in table 3 (table 4 in version 3) was derived from LightCycler analyses of duplicate samples. We had performed melting curve analysis with the major PRNP allele oligonucleotides for comparison but did not include these in the original table. Following the comments of the reviewer, we have subsequently repeated the analysis of all oligonucleotides using 100nM of probe and the LightTyper instrument for a more direct comparison with PCR data. Target concentrations between 100nM and 500nM were evaluated. 200nM of oligonucleotide target was found to yield Tms most comparable with PCR data. Six replicate analyses were performed for each target oligonucleotide. Mean Tm and standard deviation data is presented in table 4. The nucleotides representing PRNP polymorphisms are now highlighted. The following was added to the methods section:

“Melting curve analysis with complementary oligonucleotides

The affect of rare polymorphisms on probe Tm was predicted using complementary oligonucleotides. 100nM of each probe was hybridised to 200nM of oligonucleotide homologue in QIAGEN PCR buffer and a total of 3mM MgCl₂. Microseal white 384 well PCR plates (Bio-Rad) were heated to 95°C for 1 minute and cooled to 20°C for 1 minute, prior to melting curve analysis using a LightTyper instrument, heating samples from 35°C to
75°C, at 0.1°C per second. Peak Tm data was determined using LightTyper software version 1.1.” [Page 10]

Moved the sentence: “The affect of glycine at codon 170 (D170G; GAT>GGT) was investigated using two DNA samples previously genotyped as ARR/AH G170Q using a combination of hME, MALDI-ToF and sequencing analysis.” from the discussion to the results section (see below).

Added the following to the results section:

**“Predicting the affect of rare sequence variants**

In addition to the T^{136}, T^{137}, L^{168}, K^{171} and E^{175} polymorphisms described above, additional published polymorphisms in codons 138, 151, 154, 170 and 172 are covered by the diagnostic probes [14,18]. The affect of glycine at codon 170 (D170G; GAT>GGT) was investigated using two DNA samples previously genotyped as ARR/AH G170Q using a combination of hME, MALDI-ToF and sequencing analysis. The G^{170} polymorphism reduced the Tm of the 171R6 melting peak to 40.0°C and is not expected to compromise the accuracy of PRNP genotypes. In the absence of suitable DNA samples, the affects of other rare polymorphisms were predicted by hybridisation of probes to complementary oligonucleotide sequences (table 4). Whilst melting peak Tms obtained with oligonucleotide and PCR targets are expected to vary (e.g. due to length and concentration differences), probe hybridisation to 200nM of oligonucleotide homologue yielded melting temperatures that were comparable with PCR amplified target sequences.” [Page 16]

**Literature**

1. Reference [14] of version 2 removed. Have not been able to locate a suitable reference for the SNaPShot or SNP-IT tag array systems employed by Orchid Cellmark for scrapie genotyping.
2. Reference [17] of version 2 removed as no longer available.
3. Corrected issue number of version 2 reference [25].
4. Corrected volume and page numbers of the version 2 reference [26].

**Quality of English**

We have not been able to identify any further problems with the quality of written English now that the scientific terminology is considered acceptable (minor point 7).

**Additional revisions**

Added the following sentences to detail the number of samples found to possess T136 and K171 polymorphisms:
“The presence of $T^{137}$ in 3 of the 8918 DNA samples was confirmed through sequencing.” [Page 11]

“The $L^{168}$ polymorphism was detected in 17 purified DNA samples and 4 of the unpurified bloods.” [Page 13]

The 8918 purified DNA samples were analysed only with 136, 154 and 171 assays. The following sentence was added to the methods section detailing the evaluation of the 141 probe with a reduced sample set:

“The 141C3 probe was evaluated using a further 400 DNA samples, again comparing genotype calls with MALDI-ToF data.” [page 7]