Final report

Predicting age of livestock from DNA samples

Project code:	L.GEN.1808
Prepared by:	Ben Hayes University of Queensland
Date published:	16th April 2021

PUBLISHED BY
Meat & Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

Accurately recording age of animals is a major challenge in the Northern Beef industry, where animals are mustered infrequently. This limits the uptake of genetic evaluation and has adverse implications for herd management and meat quality. In this project, we have derived the world's first methylation clock to predict age in cattle from a hair sample, using Oxford Nanopore sequencing. With a reference population of 50 animals, the clock was able to predict birth year, for animals less than 10 years of age. In extensive conditions where bulls are with the herd year round, and combined with an annual pregnancy test, the clock could be used to identify those heifers that first fell pregnant as yearlings, and those that first fell pregnant as two or even three year olds, to enable culling to improve fertility. For meat quality, the ability to discriminate between animals that are in the range 6 and 10 years would have some value. These animals have very different meat quality, but currently there is no way to discriminate between them as they are all 8 tooth and have maximum ossification (pers comm Peter McGilchrist). To make the methylation clock more accurate (targeting birth month), we have sequenced an additional 50 animals with known ages with the Nanopore technology – adding this information to the reference population will increase the accuracy of the methylation clock to months rather than ~ 1 year. A newly released Nanopore technology called adaptive sequencing, which allows only informative sites in the genome to be targeted, should reduce the cost of obtaining the necessary sequence for the methylation clock for each animal by a factor of ~ 10. Finally, our vision is that the methylation clock will be integrated into "Crushside genotyping", such that producers obtain age, GEBV and parentage from a single assay.

Executive summary

Background

Accurately recording age of animals is a major challenge in the Northern Beef industry, where animals are mustered infrequently. This limits the uptake of genetic evaluation and has adverse implications for herd management and meat quality for Northern beef producers. In this project, we have derived a "methylation clock", such that beef producers can take a tail hair from their animals, and age can be predicted from the DNA in this sample.

Objectives

The two objectives in this project were:

1. Deliver a DNA test to predict age of animals, that uses the same sample as for taken for genotyping, and can be integrated into the workflows currently used to genotype animals on a large scale

2. Deliver DNA based predictions of ossification, that can be integrated into a prediction of MSA grade in live animals.

Objective 1. has been achieved. We can now take a tail hair sample from an animal, and predict it's age, using Nanopore sequencing to elucidate the methylation profile of the animal, and the information from the reference set, of animals of various ages and with Nanopore sequence derived methylation profiles.

Outcomes 2 has been partially achieved. We have collected samples from animals with a range of ossification scores and known birth dates, and have methylation profiles on these animals. The criteria and strategy used to select these samples is described in detail. However we have not yet analysed this data, as the focus has been on objective 1. Note that in the project proposal, it was anticipated that this analysis may not be completed in the project timeframe (2 years), and the milestone for this objective was "sequencing of samples for ossification analysis complete", which has been achieved.

Methodology

After evaluating a range of technologies to derive methylation profiles of animals, Nanopore sequencing was selected for deriving the methylation clock. Nanopore sequencing gave the most information (methylated sites) by an order of magnitude, and was cost competitive with the other technologies. With the advent of adaptive sequencing, Nanopore sequencing will be able to be deployed at a commercially acceptable cost.

A reference population of animals with known birth dates was then sequenced with the Nanopore technology. A methylation relationship matrix approach was used to predict age of validation sets of animals.

Results/key findings

Using Nanopore sequencing to derive methylation profiles (with 50 animals with known birthdates as reference population) a methylation clock was derived for predicting age. In an independent validation, age was predicted with an accuracy of 0.65. This was for both sites with 80% of animals called, and a much smaller set of sites that were shown to be predictive both in human and dog. The

average difference in predicted age and actual age for animals less than three years old was one year, for animals between 3 years and 10 years it was 1.5 years.

Benefits to industry

The methylation clock derived here has two immediate applications in the northern beef industry.

In extensive conditions where bulls are with the herd year round, and combined with an annual pregnancy test, it could be used to identify those heifers that first fell pregnant as yearlings, and those that first fell pregnant as two or even three year olds. This would allow selection of heifers that fell pregnant as yearlings, and culling of heifers that fell pregnant much later in life, improving the fertility of the herd over time.

For meat quality, the ability to discriminate between animals that are in the range 6 and 10 years would have some value. These animals have very different meat quality, but currently there is no way to discriminate between them as they are all 8 tooth and have maximum ossification (pers comm Peter McGilchrist).

As the reference population for deriving the methylation clock grows, the prediction of age will become more accurate. This will open up new opportunities for the northern beef industry. For example when we can accurately predict birth month, animals which obtained records, but otherwise have no birth date (reflecting the challenges of recording birth date in the Northern industry), could be entered into Breedplan, greatly improving the uptake of Breedplan in the North.

Future research and recommendations

To get the methylation clock into practise, there are at least two further activities to take place. We have sequenced an additional 50 animals with known ages with the Nanopore technology – adding this information to the reference population will increase the accuracy of the methylation clock to months rather than ~ 1 year. The reference population should continue to be expanded until birth month can be accurately predicted.

Secondly, adaptive sequencing, a new extension to the Nanopore technology that was developed in response to the COVID-19 pandemic, would allow only the human-dog methylation sites to be targeted, which would reduce the cost of obtaining the necessary sequence for the methylation clock for each animal by a factor of ~ 10.

The methylation clock could be run as a stand alone assay from tail hair or an ear punch, by a lab like Neogen or Wetherbys for example. Ideally however the sites for used for the methylation clock will be integrated into the “Crushside genotyping” pipeline, such that producers receive back GEBV, parentage and age all in the one Crushside assay.

Table of contents

Abstract 2

Executive summary 3

1. Background 6

2. Objectives 7

   2.1 Deliver a DNA test to predict age of animals, that uses the same sample as for taken for genotyping, and can be integrated into the workflows currently used to genotype animals on a large scale 7

   2.2 Deliver DNA based predictions of ossification, that can be integrated into a prediction of MSA grade in live animals. 7

3. Methodology 7

   3.1 Which technology for deriving methylation profiles is most accurate, practical and cost-effective? 7

   3.2 Deriving the “methylation clock” for cattle 9

   3.3 Samples for ossification prediction 11

4. Results 11

   4.1 Which technology for deriving methylation profiles is most accurate, practical and cost-effective? 11

   4.2 Deriving the “methylation clock” for cattle 14

5. Conclusion 15

   5.1 Key findings 16

   5.2 Benefits to industry 17

6. Future research and recommendations 17

7. References 18

1. Background

If age of an animal could be accurately predicted from a DNA sample (eg. tail hair or ear punch), it would considerably simplify the implementation of genetic improvement programs. When combined with genotyping, the need for recording birth dates and mothering up could be avoided. In extensive production systems, such as the Northern Beef industry, this could allow BREEDPLAN to be adopted on a much larger scale, as the difficulty of recording birth dates is often given as a key barrier to wider adoption of BREEDPLAN. This would lead to considerably greater genetic gains in the Northern beef industry. In Merino sheep, recording birth dates and mothering up can sometimes be associated with lamb losses, so if age could be predicted from a DNA sample, these might be avoided.

In forensic science, there is a great deal of interest, and research, into predicting age from DNA (eg. Bocklandt et al. 2011, Bell et al 2012, Horvarth et al. 2013,. Vidaki et al. 2017). Early work in this area focussed on methods to measure telomere lengths (eg. Tsuji et al. 2002). However the assays required to measure telomere lengths are complex and expensive, with outcomes difficult to compare across labs (eg. O'Callaghan et al. 2011). More recent work has focussed on the extent and pattern of methylation of groups of genes, as this can be easily measured with next generation sequencing (eg bi-sulphide sequencing). While methylation at many genes varies with the environment, there are subsets of genes whose methylation pattern operates like a "biological clock", in both humans and mice (Horvarth 2013, Stubbs et al. 2017), and this information can be used to accurately predict age (eg. Horvarth 2013, Vidaki et al. 2017). It should be noted that methylation patterns at these genes are consistent across tissues, eg a tail hair sample or tissue sample would give a similar answer (eg Horvarth et al. 2013).

The value of DNA based age prediction comes down to it's accuracy – how precisely does it predict age? In humans, with a long lifespan, the accuracy is measured in years (eg +/- three years, Vidaki et al. 2017). This is clearly not adequate for our purposes. In mice however, with a shorter lifespan, predictions are much more accurate, with a recent large study reporting an error +/-3 weeks (Stubbs et al. 2017). If we could achieve a similar precision of age prediction, this would be suitable for use in genetic evaluation such as BREEDPLAN.

While the focus here is on age, methylation patterns in DNA samples will have other uses in livestock as well. Ossification is a key component of MSA, but is difficult to predict in live animals prior to slaughter because ossification reflects how hard an animals life has been – for example periods of poor nutrition increase ossification, over and above what would be expected for a animal of a give age. Methylation patterns at some genes do closely reflect the severity of the environment the individual has experienced (eg Horvarth 2013), so could be trialled to predict ossification. This could lead to a DNA test which predicts MSA grade accurately, reflecting both genetic and environmental factors (eg. Bos indicus content, genomic breeding values for marbling score, predictions of ossification, all from a single sample from a live animal taken by the producer). This information could ultimately be used for improved management and drafting of animals into different supply chains. For example, animals with high levels of predicted ossification could be drafted into lower value chains, with reduced expense of feeding.

2. Objectives

2.1 Deliver a DNA test to predict age of animals, that uses the same sample as for taken for genotyping, and can be integrated into the workflows currently used to genotype animals on a large scale

2.2 Deliver DNA based predictions of ossification, that can be integrated into a prediction of MSA grade in live animals.

Objective 2.1. has been achieved. We can now take a tail hair sample from an animal, and predict it's age, using Nanopore sequencing to elucidate the methylation profile of the animal, and the information from the reference set, of animals of various ages and with Nanopore sequence derived methylation profiles.

Outcomes 2.2 has been partially achieved. We have collected samples from animals with a range of ossification scores and known birth dates, and have methylation profiles on these animals. The criteria and strategy used to select these samples is described in detail. However we have not yet analysed this data, as the focus has been on objective 2.1 (In the project proposal, it was anticipated that this analysis may not be completed in the project timeframe (2 years), and the milestone for this objective was "sequencing of samples for ossification analysis complete", which has been achieved).

3. Methodology

3.1 Which technology for deriving methylation profiles is most accurate, practical and cost-effective?

The first part of this project investigated the technologies that are available for deriving methylation profiles, applied to cattle. Assessment criteria were accuracy of the technology, practicality (could the technology be used in a high throughput setting, and potentially be integrated with genomic estimated breeding value pipelines) and cost-effectiveness There are several available technologies to study DNA methylation, however few of these have been applied in livestock. Based on an extensive review of particularly the recent literature, it was clear that new technologies for profiling DNA methylation patterns were rapidly emerging. So we elected to use four different technologies to profile methylation patterns, including targeted bisulphide sequencing. Based on reported coverage and accuracy of methylation detection (in human studies), we tested three technologies:

- Human Methylation EPIC array (> 850K methylation sites),
- Long-read Oxford Nanopore sequencing,
- Reduce representation bisulfite sequencing (RRBS),

Details of each platform were described below. Results from these four techniques will combined in order to get common CpG sites in cattle for further studies in this project.

3.1.1 Sample information

Sample information. Mostly tail hair and some liver samples were used in this first experiment of the project, Table 1. The liver samples were included because they represent extreme ages.

Table 1. Information of bovine samples used for DNA methylation project

Sample_ID	Type	Sex	Age
Liver	Liver	Female	10 years
Fetal Liver	Liver	Female	2 months
521986	Tail hair	Female	2.1 years
521969	Tail hair	Female	2.2 years
521954	Tail hair	Female	2.1 years
577500	Tail hair	Female	3.1 years
577579	Tail hair	Female	3.2 years
595180	Tail hair	Female	3.2 years

DNA from hair and liver samples were purified using the Gentra Puregene DNA kit according to manufacturer protocol (QIAGEN, Australia). DNA was resuspended in DNA hydration solution provided by QIAGEN, quantified using fluorometric quantification (Qubit 4.0, ThermoFisher Scientific) and nanodrop. Size of DNA samples was assessed by running on Pulsed-field Gel Electrophoresis system (PFGE).

3.1.2. Methylation platforms

Human Methylation EPIC array. Eight DNA samples were sent to AGRF for running Human Methylation EPIC array under project ID ILMLEPIC-15262 (PO number 3860008294). Briefly, DNA (500ng) was treated with sodium bisulphite using the EZ DNA Methylation kit (Zymo Research, CA, USA). DNA methylation was quantified using the Illumina HumanMethylationEPIC (EPIC) BeadChip (Illumina, CA, USA) run on an Illumina iScan System (Illumina, CA, USA) using the manufacturer's standard protocol.

Oxford nanopore sequencing. Samples were prepared for sequencing following the protocol in the genomic sequencing kit SQK-LSK109 (Oxford Nanopore Technologies; ONT). Briefly, approximately 4ug DNA was end-repaired and deoxyadenosine (dA)-tailed using the Ultra II end-repair module (New England BioLabs; NEB). Sequencing adapter (Oxford Nanopore Technologies; ONT) were ligated using blund/T4 ligase (NEB). Libraries from End-repair reaction and ligation steps were clean-up with AMPureXP beads (Beckman Coulter). After cleaning, DNA was resuspended in 12ul of Elution Buffer (Oxford Nanopore Technologies; ONT) before being combined with 37.5ul of sequencing buffer (SQB - Oxford Nanopore Technologies; ONT) and 25.5ul of loading bead and loaded on a MinION SpotON flow cell (FLO-MIN106). Sequencing was performed for 48h with a MinION sequencer.

Reduce representation bisulfite sequencing. The eight DNA samples were sent to Novogen (Hongkong) for reduce representation bisulfite sequencing under PO number 3860008269. Briefly, DNA was digested with the methylation insentive MspI enzyme (C^CGG). After end-repair, A-tailing, adapter-ligated, size-selected (40-220bp), treated with bisulfite (EZ DNA Methylation Gold Kit, Zymo Research) and PCR amplified, eight RRBS libraries were sequenced on an Illumina NovaSeq.

3.1.3. Data analysis

Human Methylation EPIC array. Raw IDAT files were processed with Illumina's GenomeStudio software v2011.1 and background normalised using negative control probes to generate methylation β-values which were used for all downstream analyses. We used MethylationEPIC_v-1-0_B2 manifest for processing EPIC data.

Oxford Nanopore sequencing. We used nanopolish v0.11.1 to detect base modifications. Briefly, all fastq files of each sample were merged into a single file. An index file that links read ids with their signal level data in the fast5 files was created using "nanopolish index –d path/to/fast5_files/ all_merge_output.fastq". We then aligned the basecalled reads to the bovine reference genome (ARS-UCD1.2 assembly) using "minimap2 –a –x map-ont –secondary=no reference.fa all_merge_output.fastq". "nanopolish call-methylation" command was used to detect methylated bases (in this case 5-methylcytosine in a CpG context). The output file provided information about the position of the CG nucleotide on the reference genome, the ID of the read and the log-likelihood ratio. A positive value in the log-lik-ratio column indicates support for methylation. Finally, we used "calculate_methylation_frequency.py" to calculate how often each reference position was methylated.

Reduce representation bisulfite sequencing. Bisulfite reads were aligned to the bovine genome (ARS-UCD1.2.). Briefly, adaptor sequences and poor quality bases were removed using Trimgalore (version v.0.6.2) in paired-end mode with default parameters. Bismark v0.15.0 was then used to align reads to bovine genome reference. At first, the bovine ARS-UCD1.2 genome was bisulfite converted and indexed to allow Bowtie alignments. Bismark for read alignment step used –rrbs –paired parameters. Count tables of the number of methylated and unmethylated bases sequenced at each CpG site in the genome were constructed using bismark_methylation_extractor with the parameters '-p –no_overlap –comprehensive –merge_non_CpG –bedgraph –counts –report –gzip.

For each technology, the difference in DNA methylation profiles between individuals and between "old" versus "young" animals was assessed. The overlap in the profiles between technologies was also assessed, that is how many methylated sites were in common between the technologies.

3.2 Deriving the "methylation clock" for cattle

As described in results, after comparing the different technologies on the criteria of accuracy, practicality and likely cost in the near future, Nanopore sequencing was selected as the technology to profile a reference population for predicting age from methylation profiles.

3.2.1 Sample selection for the reference population

In selecting samples for the reference population, we tried to balance a number of criteria 1) inclusion of cattle of a wide range of ages, so the methylation clock could be used to predict age across the wide range, 2) a number of young cattle with age that differed only be a few weeks, so we had some power to predict small differences in age for young cattle as well 3) inclusion of cattle

across several breeds used in Northern Australia, so we had power to predict in these breeds. The samples taken after considering these criteria are given in Table 2.

Table 2. Samples for the reference population to predict age from methylation profiles.

Herd	Breed	Birth date (Years)	Independent samples	Repeat samples	Calves	Total samples
Herd A	Droughtmaster	2001 to 2019	38	10	0	48
Herd B	Brahman	2011 to 2013	2	0	0	2
Herd C	Tropcomposite	2016 & 2017	10	0	0	10
Herd D	Droughtmaster	2015 & 2016	10	0	0	10
Herd E	Brahman	2016 & 2020	0	20	10	30
Total			60	30	10	100

3.2.2 Sample preparation and sequencing

Genomic DNA was extracted using the Gentra Puregene Tissue Kit (Qiagen) according to the manufacturer's instructions with modifications. Briefly, 20-30 hair samples were lysed in 300 μl of Cell lysis solution (Gentra® Puregene® Tissue Kit) and 1.5 μl of Proteinase K solution (20mg/ml) for 5 hours at 55°C. RNA was then digested by addition of 1.5 μl of RNase A Solution, following 1-hour incubation at 37°C. Samples were placed on ice for 5 minutes after adding 100 μl Protein Precipitation Solution (Gentra® Puregene® Tissue Kit) and spun at 14000 x g for 3 minutes. 300 μl of Isopropanol was used to precipitate DNA. Samples were centrifuged at 14000 x g for 3 minutes. DNA pellets were washed in 300 μl of 70% ethanol, air-dried for 5 minutes and resuspended in 55 μl of DNA Hydration Solution (Gentra® Puregene® Tissue Kit). DNA concentrations were measured using the Qubit dsDNA Broad Range assay kit (Thermo Fisher Scientific). The purity of the extracted DNA was determined with the NanoDropND 1000 (v.3.5.2, Thermo Fisher Scientific), assessing the 260/280 nm and 260/230 nm ratios. The size of extracted DNA was examined using pulsed-field gel electrophoresis (Sage science, USA) with a 0.75% Seakem Gold agarose gel (Lonza, USA) in 0.5X Tris/Borate/EDTA (TBE) running buffer, run for 16 hours at 75 V. The gel was stained after the electrophoresis with SYBR Safe dye (10000x) and visualized using Quantity One analysis software (Bio-rad).

Extracted DNA samples were prepped using a ligation kit (SQK-LSK 109, Oxford Nanopore Technology) based on the manufacturer's instruction with some modifications. Starting with 6 – 10 μg of DNA produces enough prepped library to provide up to 4 library loads from a single prep. Samples were diluted at the clean-up points with nuclease-free water to prevent bead clumping during 0.4x AmPureXP purifications. End-prep reaction and ligation incubation times were increased to 30 minutes and 1 hour, respectively. Lastly, during the sequencing run (96 hours), the flow-cell was washed at least three times using the nuclease-flush kit (Oxford Nanopore) and then reloaded with the same prepped library. The use of DNase I clearing and flow-cell refueling helps to remove blocking DNA and increases the sequencing throughput.

Base calls were made from the raw current disruption data using GUPPY on the University of Queensland high performance computing infrastructure. F5c (Gamaarachchi et al., 2019) was used to implement GPU based methylation calling. Gene promoters were identified using CAGE-seq data to identify transcription start sites, and taking the 1000bp flanking regions. Methylation frequency was calculated using nanopolish. Then for 100 base pair windows across the whole genome, each window for each animal was called as methylated (1) or not methylated (0) if the average frequency of methylation for the sites in each window was greater than 0.5 or less than 0.5 respectively.

3.2.3 Data analysis

Using these calls, methylation relationship matricies were constructed among the animals. Two relationship matricies were considered:

1) Only using sites that called in at least 80% of animals (eg sequence coverage was sufficient) and there was variation in the calls (standard deviation of calls>=0.5)
2) Using sites close to genes reported in both humans and dogs to be predictive for methylation clocks in those species (Wang et al. 2020)

There were 56673 sites and 15667 sites for relationship matrix 1 and 2 respectively. The matricies were formed as Methylation Relationship Matrix (**MRM**) = **X'X**/(number of sites), where X is animal x number of sites matrix, which each element whether the animal is methylated at that site (1) or not (0).

Then this information was used in the model to predict age, where the model was

Age (years) = mean + herd + animal + error

Where mean and herd were fixed effects, and animal was a random effect assumed distributed as a normal distribution with mean 0 and variance **MRM**σ^2_m, where σ^2_m is the variance in age captured by the methylation profile. The methylation relationship matrix was built in GCTA (Yang et al. 2012) and the model was fitted, and variance components estimated, in ASREML.

To assess the accuracy of predicting age from the methylation profile, we used a cross validation strategy. Sets of 5 randomly chosen individuals had their phenotypes (ages) removed from the analysis, but they were still included in the MRM. This resulted in age effects being predicted for these animals. These age effects were then correlated with the actual age for these animals. The cross validation procedure was repeated 10 times until all animals had been dropped from the analysis but included in the validation. The resulting correlation of predicted age and actual age was taken as the accuracy of prediction of age.

3.3 Samples for ossification prediction

Working with three Northern Australian herds, we have obtained DNA samples have been obtained from 87 animals with complete MSA data, as well as birth dates (to enable chronological age and biological age to be disentangled). The animals cover a range of ages, from 18 months to 5 years. The animals were slaughtered in a reasonable sized contemporary groups (>5). 10 of these animals have been sequenced and methylation profiled using the Nanopore technology described above.

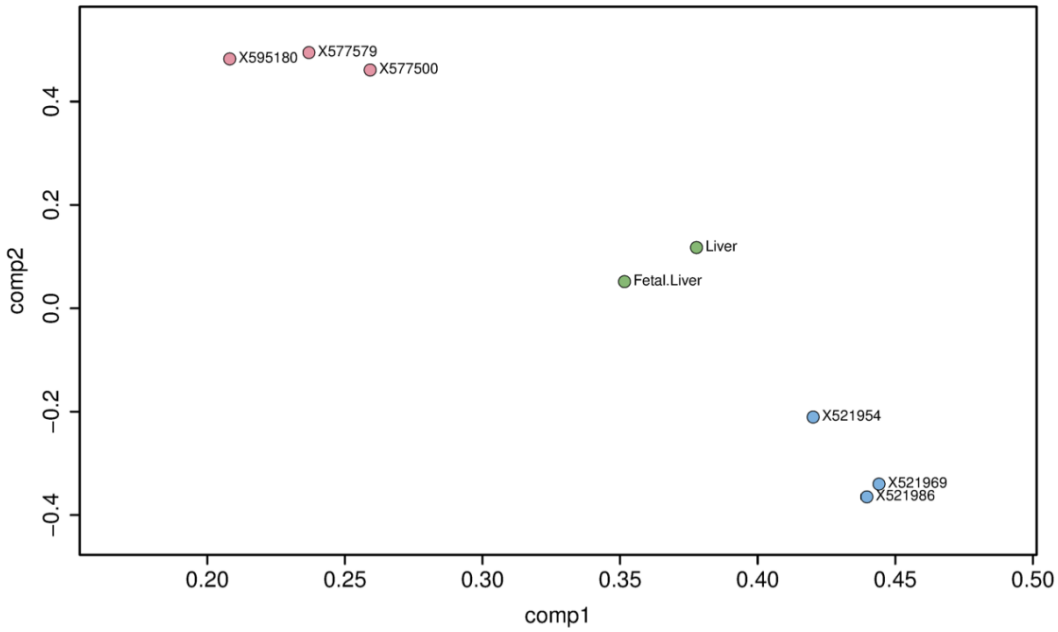
4. Results

4.1 Which technology for deriving methylation profiles is most accurate, practical and cost-effective?

Human Methylation EPIC array. As expected, the CpG sites detected in our bovine samples from Human methylationEPIC array were low, with between 85,000 and 17,000 genome sites detected as methylated. However, the relatively small number of sites were sufficient to clearly distinguish "old"

versus “young” animals (Figure 1) with a clear separation between the groups using principal component analysis.

Figure 1. PCA of Human Methylation EPIC data, showing separation of sample groups (young versus old animals, 2 year old versus 3 year old cows, red versus blue dots respectively) and sample type (hair versus liver)



DNA methylation profiles from Oxford Nanopore. The Nanopore technology detected a much larger number of methylated sites in the genome of each sample, ranging from 13.5 million to 20.4 million Table 3. This very large number of sites should enable very accurate predictions of age from this technology.

Table 3. Summary of Nanopore sequencing data for each sample and the call methylation status

Sample	Group	Data (Gb)	N50	Mapped reads (%)	Total C’ in CpG context (Million)
Fetal Liver	Young	6.6	12.5	94.35	18,7
Liver	Old	8.1	5.8	98.35	16,9
577579	Old	8.7	1.2	92.26	13,5
521954	Young	7.5	1.3	96.72	13,6
521969	Young	6.7	2,4	96.89	17,0
521986	Young	8.4	2.3	95.06	20,4

Reduced representation bisulfite sequencing. As to be expected for a technology that provides a methylation profile for only a “reduced part of the genome, the number of methylated sites which

make up these methylation profiles was considerably fewer with this method than with the Nanopore, Table 4.

Table 4. Summary of methylation calling and mapping efficiency for reduced representation bisulphite sequencing data.

Sample	Group	Mapping efficiency (% reads mapped)	Total C's in CpG context (Million)
Fetal Liver	Young	37	4,8
Liver	Old	35	5,4
521969	Young	30.2	5,1
521986	Young	32.1	4,6
577500	Old	33.9	4,8
577579	Old	35.1	5,3
595180	Old	31.1	5,0

About one quarter of the sites detected with the Oxford nanopore technology were detected with the reduced representation bisulphide sequencing data, even though the price per sample is approximately the same.

Comparison of DNA methylation profiles between technologies. We compared the DNA methylation profiles from the Nanopore sequencing and reduced representation bisulphite sequencing (RRBS). A relatively small proportion of the sites overlapped, reflecting difference in the technologies. When called methylated sites did overlap, there was good agreement in the degree of methylation, as evidenced by the high correlation, Table 5 .

Table 5. Extent of common CpG sites between Nanopore and reduced representation bisulphite sequencing (RRBS) as well as the correlation of degree of methylation at these sites.

Sample	**Sample group**	**Nanopore vs RRBS**	
		Common CpG sites	**Pearson correlation**
Fetal Liver	Young	145545	0.74
Liver	Old	115435	0.87
521986	Young	133698	0.83
521969	Young	124963	0.82
577579	Old	99543	0.81

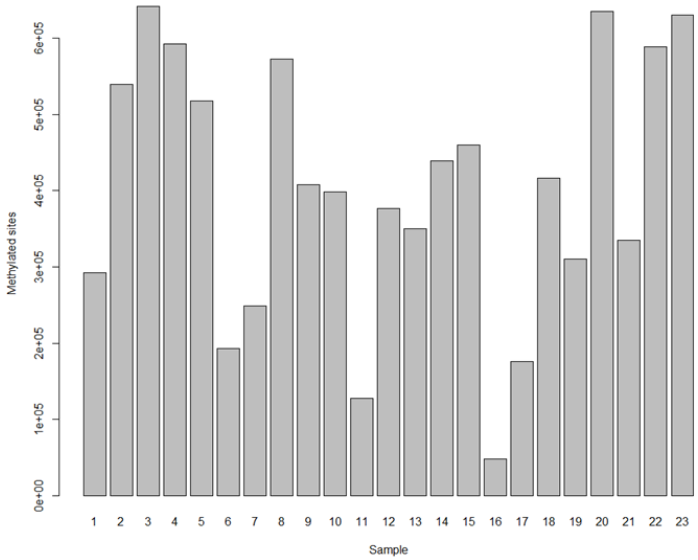
Given the very large number of methylated sites generated by the Oxford Nanopore technology (up to 20.4 million methylated sites in some samples), and the comparable accuracy to reduced representation whole bisulfite sequencing, as well as comparable or cheaper cost than the

HumanEpic Array, Oxford Nanopore was selected as the technology for the next step, deriving the "methylation clock" for cattle. An additional point in favour of Nanopore is the soon to be released capacity for "adaptive sequencing", which allows specific genome sites to be sequenced only. This could reduce the cost of the technology to ~ $15 per sample.

4.2 Deriving the "methylation clock" for cattle

All 100 samples in Table 2, the reference population for predicting age from methylation profiles, have been Nanopore sequenced. However due to time constraints, only 56 samples have been called for methylation profiles thus far. Sites called across the genome across samples was quite variable, Figure 2.

Figure 2. Methylated sites called across samples with Nanopore sequencing.



In later sequencing runs, the number of sites called became more consistent as we became more adept at using this new technology.

Predictions of age using methylation profiles was made:

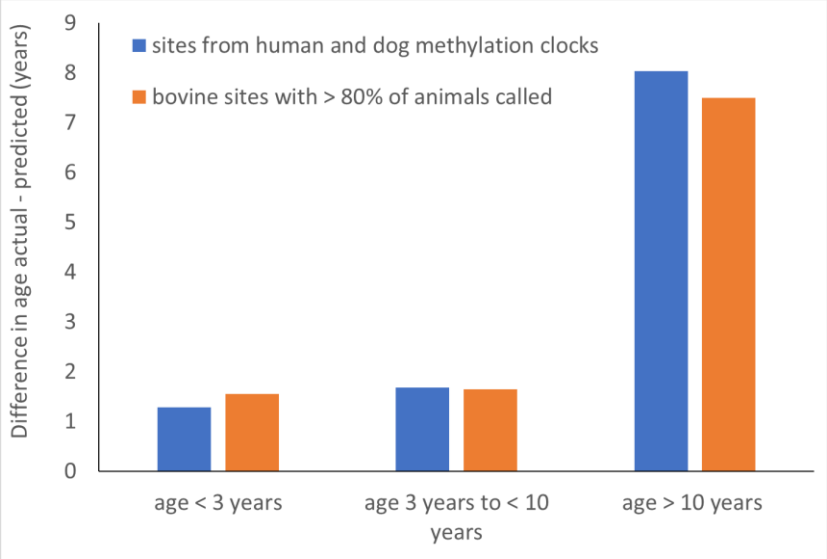
1) Only using sites that called in at least 80% of animals (eg sequence coverage was sufficient) and there was variation in the calls (standard deviation of calls>=0.5)
2) Using sites close to genes reported in both humans and dogs to be predictive for methylation clocks in those species (Wang et al. 2020)

There were 56673 sites and 15667 sites 1 and 2 respectively, and as described in methods these sites were used to build a methylation relationship matrix (MRM) to enable predictions of age into a validation set (age for individuals in the validation set was never used in the reference set).

The correlation of predicted age and actual age in the validation data sets was **0.65** (moderate to high) regardless of which class of sites was used to build the GRM.

The difference in predicted age and actual age was ~ 1 year for animals age less than three years, about ~ 1.5 years for animals age 3-10 years, but was 8 years from animals 10 years and older, Figure 3. This is most likely because we had very few animals in the reference population older than 10 years. The result could also reflect slower methylation at greater ages (eg a non-linear rate of methylation), as was observed in dogs (Wang et al. 2020). In any case, animals older than 10 years represent a very small class in industry and are the least likely to be candidates for ageing.

Figure 3. Average difference in age predicted from the methylation clock and actual age for three age classes of animals, and two methods of selecting methylated sites for inclusion in the clock.



Results from using either sites called in 80% of our samples, with a standard deviation of methylation scores>0.5, or sites common in human and dog methylation clocks, both gave similar accuracy of predicting age, Figure 2. If adaptive sequencing is used, which is very likely the next step in deriving a cost effective methylation clock, using the sites common in human and dog methylation clocks is particularly attractive because only ~15k sites have to be assayed.

5. Conclusion

In this project, we have derived the world's first methylation clock for cattle, using Oxford Nanopore sequencing. With a reference population of 50 animals, the clock was able to accurately predict birth year, for animals less than 10 years of age. The predictions were made from a tail hair sample, which is relatively easy to obtain.

Predicting birth year in Northern Australian cattle has some utility. For example, in extensive conditions where bulls are with the herd year round, and combined with an annual pregnancy test, it could be used to identify those heifers that first fell pregnant as yearlings, and those that first fell pregnant as two or even three year olds.

For meat quality, the ability to discriminate between animals that are in the range 6 and 10 years would have some value. These animals have very different meat quality, but currently there is no

way to discriminate between them as they are all 8 tooth and have maximum ossification (pers comm Peter McGilchrist).

To get the methylation clock into practise, there are at least two further activities to take place. We have sequenced an additional 50 animals with known ages with the Nanopore technology – adding this information to the reference population will increase the accuracy of the methylation clock to months rather than ~ 1 year. Secondly, adaptive sequencing, a new extension to the Nanopore technology that was developed in response to the COVID-19 pandemic, would allow only the human-dog methylation sites to be targeted, which would reduce the cost of obtaining the necessary sequence for the methylation clock for each animal by a factor of ~ 10.

Finally, our vision is that the methylation clock will be integrated into "Crushside genotyping" (eg MLA Scholarship "Crushside genotyping" Harrison Lamb).

5.1 Key findings

Nanopore sequencing was able to detect vastly more methylated sites per animal (20.4 million) than either the HumanEpi array, or reduced representation bisulfite sequencing

The accuracy of detecting methylated sites was similar for Nanopore sequencing and reduced representation bisulfite sequencing

Using Nanopore sequencing to derive methylation profiles (with 50 animals with known birthdates as a reference population) a methylation clock was derived for predicting age.

In an independent validation, age was predicted with an accuracy of 0.65. This was for both sites with 80% of animals called, and a much smaller set of sites that were shown to be predictive both in human and dog

The average difference in predicted age and actual age for animals less than three years old was one year, for animals between 3 years and 10 years it was 1.5 years.

The Nanopore sequence generated in this project was used in other activities as well, exploiting the ability of the long Nanopore reads to identify structural variants (long sections of the genome that are duplicated or deleted). For example it was used to confirm the polled allele in Brahman cattle is largely celtic (eg from Angus/British breed cattle) in origin (Lamb et al. 2020). This information has been used to refine the SNP test used for polled in Brahman cattle (Randhawa et al. 2020).

The Nanopore sequence was also opportunistically used to confirm and characterise a structural variant in Bos indicus cattle that causes dark colouration of the head, shoulders and rump, Figure 4. Working with Brazilian collaborators (who were working on Nelore cattle), we confirmed this structural variant was present in Brahman cattle (Trigo et al. 2021).

Figure 4. A. Bull with no copies of the agouti signaling protein gene (ASIP) deletion, B. Bull with double copy of the deletion at the agouti signaling protein gene (ASIP) deletion.

A.	B.
	

5.2 Benefits to industry

The methylation clock derived here has two immediate applications in the northern beef industry.

In extensive conditions where bulls are with the herd year round, and combined with an annual pregnancy test, it could be used to identify those heifers that first fell pregnant as yearlings, and those that first fell pregnant as two or even three year olds. This would allow selection of heifers that fell pregnant as yearlings, and culling of heifers that fell pregnant much later in life, improving the fertility of the herd over time.

For meat quality, the ability to discriminate between animals that are in the range 6 and 10 years would have some value. These animals have very different meat quality, but currently there is no way to discriminate between them as they are all 8 tooth and have maximum ossification (pers comm Peter McGilchrist).

As the reference population for deriving the methylation clock grows, the prediction of age will become more accurate. This will open up new opportunities for the northern beef industry. For example when we can accurately predict birth month, animals which obtained records, but otherwise have no birth date (reflecting the challenges of recording birth date in the Northern industry), these animals could be entered into Breedplan, greatly improving the uptake of Breedplan in the North.

6. Future research and recommendations

To get the methylation clock into practise, there are at least two further activities to take place. We have sequenced an additional 50 animals with known ages with the Nanopore technology – adding this information to the reference population will increase the accuracy of the methylation clock to months rather than ~ 1 year. The reference population should continue to be expanded until birth month can be accurately predicted.

Secondly, adaptive sequencing, a new extension to the Nanopore technology that was developed in response to the COVID-19 pandemic, would allow only the human-dog methylation sites to be

targeted, which would reduce the cost of obtaining the necessary sequence for the methylation clock for each animal by a factor of ~ 10.

The methylation clock could be run as a stand alone assay from tail hair or an ear punch, by a lab like Neogen or Wetherbys for example. Ideally however the sites for used for the methylation clock will be integrated into the "Crushside genotyping" pipeline, such that producers receive back GEBV, parentage and age all in the one rapid, Crushside assay.

7. References

Bell J.T., Tsai P.C., Yang T.P., Pidsley R., Nisbet J., Glass D., Mangino M., Zhai G., Zhang F., Valdes A., Shin S.Y., Dempster E.L., Murray R.M., Grundberg E., Hedman A.K., Nica A., Small K.S., Mu T.C., Dermitzakis E.T., McCarthy M.I., Mill J., Spector T.D., Deloukas P. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 2012;8(4):e1002629.

Bocklandt S., Lin W., Sehl M.E., Sanchez F.J., Sinsheimer J.S., Horvath S., Vilain E. Epigenetic predictor of age. PLoS One. 2011;6(6):1–6.

Endelman, J (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4, 250-255.

Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA, Goddard ME. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2012 Jul;95(7):4114-29.

Gamaarachchi, H, Lam, CW, Jayatilaka, G, Samarakoon, H, Simpson, JT, Smith, MA, Parameswaran, S (2019) GPU Accelerated Adaptive Banded Event Alignment for Rapid Comparative Nanopore Signal Analysis. bioRxiv 756122.

Hannum, G, Guinney, J, Zhao, L, Zhang, L, Hughes, G, Sadda, S, Klotzle, B, Bibikova, M, Fan, J-B, Gao, Y, Deconde, R, Chen, M, Rajapakse, I, Friend, S, Ideker, T, Zhang, K (2013) Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Molecular Cell 49, 359-367.

Horvath, S (2013) DNA methylation age of human tissues and cell types. Genome Biology 14, 3156.

Lamb HJ, Ross EM, Nguyen LT, Lyons RE, Moore SS, Hayes BJ. Characterization of the poll allele in Brahman cattle using long-read Oxford Nanopore sequencing. J Anim Sci. 2020 May 1;98(5):skaa127.

O'Callaghan NJ, Fenech M. A quantitative PCR method for measuring absolute telomere length. Biological Procedures Online. 2011;13:3.

Petkovich, DA, Podolskiy, DI, Lobanov, AV, Lee, S-G, Miller, RA, Gladyshev, VN (2017) Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. Cell Metabolism 25, 954-960.e6.

Randhawa IAS, Burns BM, McGowan MR, Porto-Neto LR, Hayes BJ, Ferretti R, Schutt KM, Lyons RE. Optimized Genetic Testing for Polledness in Multiple Breeds of Cattle. G3 (Bethesda). 2020 Feb 6;10(2):539-544. doi: 10.1534/g3.119.400866.

Stubbs, TM, Bonder, MJ, Stark, A-K, Krueger, F, Bolland, D, Butcher, G, Chandra, T, Clark, SJ, Corcoran, A, Eckersley-Maslin, M, Field, L, Frising, UC, Gilbert, C, Guedes, J, Hernando-Herraez, I, Houseley, J, Kemp, F, MacQueen, A, von Meyenn, F, Okkenhaug, K, Reik, W, Rhoades, M, Santbergen, MJC, Stark, A-K, Stebegg, M, Stubbs, TM, Veldhoen, M, von Meyenn, F, Stegle, O, Reik, W, Team, BIAC (2017) Multi-tissue DNA methylation age predictor in mouse. Genome Biology 18, 68.

Thompson, MJ, vonHoldt, B, Horvath, S, Pellegrini, M (2017) An epigenetic aging clock for dogs and wolves. Aging 9, 1055-1068.

Trigo, BB. Adam Utsunomiya, Alvaro A A D Fortunato, Marco Milanesi, Rafaela, Harrison Lamb, Loan Nguyen, Elizabeth M Ross, Ben Hayes, Rômulo C M Padula, Thayla S Sussai, Ludmilla B Zavarez, , Rafael S Cipriano, Maria M T Caminhas, Flavia L Lopes, Cassiano Pelle, Tosso Leeb, Danika Bannasch, Derek Bickhart, Timothy P L Smith, Tad S Sonstegard, José F Garcia, Yuri T Utsunomiya. Variants at the ASIP locus contribute to coat color darkening in Nellore cattle. Genetics Selection Evolution. In press.

Tsuji A., Ishiko A., Takasaki T., Ikeda N. Estimating age of humans based on telomere shortening. Forensic Sci. Int. 2002;126:197–199.

Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. Forensic Science International Genetics. 2017;28:225-236.

Wang T, Ma J, Hogan AN, Fong S, Licon K, Tsui B, Kreisberg JF, Adams PD, Carvunis AR, Bannasch DL, Ostrander EA, Ideker T. Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome. Cell Syst. 2020 Aug 26;11(2):176-185.e6. doi: 10.1016/j.cels.2020.06.006.

Wang, T, Tsui, B, Kreisberg, JF, Robertson, NA, Gross, AM, Yu, MK, Carter, H, Brown-Borg, HM, Adams, PD, Ideker, T (2017) Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. Genome Biology 18, 57.

Weidner, CI, Lin, Q, Koch, CM, Eisele, L, Beier, F, Ziegler, P, Bauerschlag, DO, Jöckel, K-H, Erbel, R, Mühleisen, TW, Zenke, M, Brümmendorf, TH, Wagner, W (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biology 15, R24.

Wu, X, Chen, W, Lin, F, Huang, Q, Zhong, J, Gao, H, Song, Y, Liang, H (2019) DNA methylation profile is a quantitative measure of biological aging in children. Aging (Albany NY) 11, 10031-10051.

Ziller MJ, Elena K. Stamenova, Hongcang Gu, Andreas Gnirke and Alexander Meissner. 2016. Targeted bisulfite sequencing of the dynamic DNA methylome. Epigenetics & Chromatin 2016 9:55