



Department
for Environment
Food & Rural Affairs



Enabling genome enhanced precision farming by building on traceability - a key step towards a sustainable and prosperous future for the UK livestock production industry

April 2021

Produced by members of the Defra expert committee on
Farm Animal Genetic Resources (FAnGR)

Lead Authors

Huw E. Jones and Will Haresign

Contents

<i>Executive Summary</i>	iii
<i>Introduction</i>	1
<i>Current status and use of genomic technologies in animal production</i>	2
<i>Barriers to wider use on farms and along supply chains</i>	3
<i>Starting with a more standard genomics approach</i>	4
<i>Real-time access to information and reduced testing costs</i>	5
<i>Building trust and buy-in through linking to traceability</i>	6
<i>Opportunities to lower costs through collaboration</i>	7
<i>Not a substitute for effective data collection</i>	7
<i>Developing a platform for more innovation and entrepreneurship</i>	8
<i>Generating data to develop predictive models</i>	8
<i>Developing a marketplace for easily accessible test results</i>	9
<i>Blockchain and smart contracts, a potential method to develop and maintain trust despite increasing complexity</i>	9
<i>Opportunities not threats</i>	10
<i>Building alongside international developments and progress in other sectors</i>	10
<i>Conclusions</i>	11
Appendix 1. List of Abbreviations Used	13

Executive summary

Over the last decade there has been a growing belief that harnessing recent rapid advances in detection, data capture, analysis and transfer methods can bring benefits across agricultural systems. By developing and adopting precision approaches, both on-farms (precision agriculture) and further along supply-chains, it is expected that benefits such as substantial reductions in environmental impact, resource and antibiotic usage will be possible, while also improving resistance to pests and diseases, animal welfare and the overall efficiency of production.

While many benefits are already being captured through developments in Agri-tech, most advances have been made in relative isolation. There is little doubt that to capture the true potential benefits of precision agriculture will require a more integrated approach across a wide range of disciplines, ideally guided by a common overall vision so that any core infrastructure required can be developed effectively and in a timely manner.

This paper discusses recent developments in animal genetics and genomics, how the available tools are currently used, and the potential key role the technology could play in the future as part of a more integrated system to enable development of advanced precision agriculture approaches in livestock production.

Over the last 20 years, dramatic advances have been made in developing genomic tools along with our understanding of how the content of genomes impact characteristics of interest in various species. High quality core genomic resources are now available for most livestock species, and the cost of applying many of the technologies available is continually reducing. This in turn is facilitating a rapid on-going growth in knowledge about the effect of different genome variations.

The use of genomic information is already having a substantial impact on the rate of improvement in livestock breeding. However, currently there are significant barriers to more routine use of the technology to help tailor production approaches and health treatments based on the animal's genetic make-up. Most of these barriers are logistical rather technology limitations, and could potentially be overcome through development of a more coordinated industry wide approach, at least for some species.

In the cattle sector, many of the existing barriers could be overcome by developing a UK wide, cloud based, whole genome sequence database that was linked to the existing animal traceability systems, where each animal is DNA sampled only once during its lifetime (within a few days of birth). Such a system could potentially bring a range of benefits, such as providing:

- Managed, real-time, cost effective, access to any genetic test of interest through-out its life, and a strong platform to enabling developing a wide range of new tests of interest
- Added value rather than replace existing genetic improvement programs, and help expand access to genetic information across the cattle sector to increase the impact of its use
- A valuable resource for research of national interest when linked to other data sources, including an ability to better understand the nature of genetic diversity across the UK's wealth of locally adapted breeds
- A platform upon which a culture of innovation and entrepreneurship could be fostered across the UK cattle farming sector which could help develop genetic tests of UK and global interest

By linking to an established traceability system this would allow the owners of the animal to directly manage permissions to access the whole genome information. Enabling a level of owner control could play a key role in building trust in the system across the supply chain, help increase the quality of data being collected, and thus the level of positive impact that could be achieved across the industry.

Introduction

The pressure to reduce the environmental impact and the routine use of antibiotics and health treatments in animal farming systems is growing. Indeed, a number of clear targets are now in place for the UK, such as net zero carbon by 2050, and for reducing antibiotics use.

As a result, there is a continued and growing interest in exploring the appropriate use of new technologies to help find solutions. Many innovations of relevance to agriculture have been achieved over the last few decades within specific areas or disciplines, such as breeding, feeding/nutrition, recording and sensor technologies, or management practices. While substantial benefits are still likely to be achieved through innovation in each of these areas, there is a growing expectation that much more could be achieved by working across disciplines rather than in isolation, particularly given the development and adoption of precision agriculture approaches. Innovation in this space will also be supported by the rapidly growing capability to record and analyse large amounts of data from an increasing range of relevant scenarios, including research environments, on farms, and all along supply chains.

Over the last half century advances in breeding technologies and approaches have played a major role in dramatically increasing the efficiency of both plant and animal production world-wide. The rate of progress has accelerated over the last 20 years, driven in no small part by rapid advancements in the development of new genomic technologies (see New Technologies report ⁽¹⁾). Genomic tools are now routinely used in many commercial breeding programmes to aid selection decisions. While significant benefits have already been realised, we are still only at the start of the journey in terms of exploring and capturing value through their use, not only for breeding but also more widely across animal production supply chains particularly from farm of birth through to processors, but also through to customers by improving traceability.

The main research focus in recent years has been on developing underpinning genomic tools for a wide range of species that reduce costs of use, and many of the key resources needed are now available ⁽²⁾. For example, it took around six years at a cost of around \$50M to sequence the genome of the first cow (published in 2009), but today it is possible to sequence individual genomes for most livestock species to the same level of detail in only a few hours, and for a few hundred pounds or less (reagent costs). These costs are likely to continue to fall as further advancements are made. As a result, the primary focus for research is now shifting to developing a better understanding of the effects of various genome variants across a range of scenarios, such as in different production environments, and with different feeds, disease challenges and health treatments. Investigations are also not limited to the animals themselves, but also include a focus on genomic variations in the bacteria, viruses and other organisms that cause disease or play a key role in the digestion process (*e.g.*, affecting digestion efficiency or the production of emissions) and how these organisms are impacted by the genetics of the host animal itself.

Innovations deemed near impossible only a few years ago have already been enabled through the knowledge gained to date, such as genome editing of animals to confer increased resistance to specific diseases (see New Technologies report), and the development of on-farm diagnostic tests to determine the specific organism causing disease. As the level of knowledge continues to grow rapidly,

¹ A review of current and new technologies for both genetic improvement and breed conservation of UK farm animal genetic resources. FAnGR, Defra, June 2020

² Rexroad *et al.*, 2019 Genome to Phenome: Improving Animal Health, Production, and Well-Being – A New USDA Blueprint for Animal Genome Research 2018–2027. *Frontiers in Genetics*, vol 10

many more potential opportunities to capture benefits and value are expected, particularly if the appropriate infrastructure to support that innovation can be established.

In this paper, we will briefly review how genomic technologies are currently used in livestock farming and some of the barriers to their wider use. We will also present concepts and ideas that, if implemented, could establish a platform that would have a dramatic effect on the UK livestock industry's ability to harness and capture value from the opportunities being presented through the use of these technologies.

Current status and use of genomic technologies in animal production

The whole genome for an animal typically consists of around 3 billion linked units called nucleotides, which combine to form a sequence that includes thousands of protein coding units called genes (see New Technologies report for more detail). But within a breed or species variation between animals tends to occur in only a small proportion of these nucleotides, commonly referred to as Single Nucleotide Polymorphisms (SNPs).

Over the last few decades, the primary focus in genomic testing has been determining which variant was present at a relatively small number of specific SNP positions of interest in the genome, a process often referred to as genotyping. The specific SNPs of interest are placed on a holding plate called a chip, which is then used to conduct the genotyping test. The number of SNPs that can be included on individual chips has gradually increased over time from a few hundred or less (*e.g.*, tests for inherited diseases, such as BLAD disease or CVM syndrome in cattle), to a several hundred thousand (*e.g.*, high density SNP chips that are currently available for cattle: > 750k SNPs, and poultry: >500k SNPs⁽³⁾). As a result, the number of characteristics or traits that could be considered using a single genotyping test has also increased. Typically, however, the lower the number of SNPs that are of interest, the lower the cost of producing the test. This has resulted in the development of tailored tests or chips with a defined number of markers or SNPs chosen for specific purposes.

The number of SNP chip based genetic tests available commercially for livestock species has been relatively limited to-date (discussed in more detail later), with the primary focus having been on developing tests to support more accurate selection for breeding. The most common commercial uses of tests currently used in livestock production are:

- Verification or assignment of parentage
- Determining Inherited disease status
- Increasing the accuracy of EBVs or PTAs for use in breeding programmes
- Traceability
- Breed verification

Within structured breeding programmes, the need to balance costs and level of genomic detail being generated has also led to the development of 'imputation methods'. The approach allows a higher level of genomic detail to be generated, even when using lower cost tests, by filling in gaps using more detailed genomic information collected on relatives. The process is essentially based around using more expensive methods (*e.g.*, whole genome sequencing or high density SNP chips) to test a small number of animals at the top of a breeding pyramid (*e.g.*, popular sires and their sons) and the cheaper tests with a smaller number of SNP's to genotype more animals lower down in the pyramid (*e.g.*, parents with few offspring and/or commercial production animals).

³ Kranis *et al.*, (2013) Development of a high density 600K SNP genotyping array for chicken. BMC Genomics, vol 14, Article no.: 59

Despite reducing costs, the use of tailored SNP chips limits the number of applications that specific chips can be used for, resulting in the need for separate samples and tests for different purposes, and this increases the complexity of testing systems that need to be developed. It also limits the scope to identify any new genome variants that might impact characteristics of interest. As knowledge about the effect of different SNP variations in different breeds or species continues to grow, there will be an increasing likelihood that new variants or genome regions not previously considered important will become of increasing interest. That is particularly true given that, for simplicity, nearly all the applications for genomic tests used to date have assumed that genes act in isolation, and effects are additive when more than one gene is considered together. In reality, we already know that in the majority of biological systems that is not the case, and that the size of some gene effects can also change over time.

It is important to note that for tests using a relatively small number of SNP's, hair samples (including the hair bulb) can be used as a source of DNA. However, other sample types such as blood or tissue (consisting of more DNA) are needed for tests incorporating a large number of SNP's. As the number of SNP's of interest increases, the potential welfare impact of sampling multiple times cannot be overlooked (see below).

Barriers to wider use on farms and along supply chains

Dramatic reductions in the costs of applying genomic technologies have been achieved in recent years. Although these costs are expected to continue to fall for some time, the largest barriers to their more routine use in farming systems are already most likely the result of logistical challenges as opposed to the actual cost of genotyping or sequencing DNA. For example, when a genomic test is required, it typically involves a sample being taken from an animal and sent to a lab for sequencing or genotyping. While the sequencing process itself may be very rapid once the sample arrives in the lab (typically a few hours), when transport, batching of samples to reduce costs and reporting times are factored in, the delay between sampling and results being available can be several weeks. Furthermore, different genomic tests are often carried out by separate service providers that typically use different approaches for their testing. In addition, there may also be an interest in using different genomic tests at different times within an animal's life.

While the prospect of using genomic tests more widely than just to aid breeding decisions has been discussed for a number of years (such as to support feeding or health treatment decisions), developments have been largely stifled for a number of reasons including: (i) low demand due to difficulties in factoring time delays into production cycles (see above), (ii) difficulties in justifying the relatively high cost of individual tests for use on production animals, and (iii) difficulties in accessing sufficient numbers of phenotypes to develop suitable prediction models for new tests, particularly for new characteristics of interest.

If it were possible to make test results available much more quickly, and to reduce the cost of individual genomic tests, then it could open up an opportunity to consider not only more widespread use for current applications, but also a range of potential new applications on-farms or in other parts of the supply chain that could help bring a wider range of possible benefits. In doing so it will be important to limit the number of times that an animal needs to be sampled. The ideal would be a single sample taken very early in its lifetime *i.e.*, at or close to birth.

If all these criteria could be met, possible additional applications that could be considered include:

- Breed characterisation

- Identifying new/emerging inherited diseases
- Mating optimisation – to reduce inbreeding, maximise expected performance for particular characteristics, or levels of heterosis when crossbreeding
- More accurate management of inbreeding/genomic diversity
- Increasing the accuracy of EBVs/PTAs – particularly for characteristics measured on crossbred or slaughter animals
- Increasing the accuracy of selection for characteristics with a low heritability (*e.g.*, longevity)
- Increasing accuracies for breeding for increased resistance to specific diseases or general immunity/robustness – particularly for pigs and poultry where elite breeding animals are maintained in high biosecurity units
- Generating indicators of relative genetic merit for various characteristics for animals that are not recorded nor closely related to recorded animals
- Determining the suitability of individual animals for specific management systems or markets
- Tailoring and adapting feeding plans at herd or individual animal levels to improve production efficiency and reduce environmental impact
- Designing tailored herd or flock health plans
- Determining the suitability of individual animals for specific disease treatment options - to maximise their effectiveness and reduce overall levels of use
- Helping to identify possible targets for genome editing by providing high level genome scans across breeds and populations to identify novel variants with positive effects on characteristics of interest (*e.g.*, resistance to diseases or ability to digest specific feed types in rare breeds).

Starting with a more standard genomics approach

The tendency to use genotyping with SNP's specially selected based on data from recorded populations, as opposed to whole genome sequencing, over recent years was largely driven by cost considerations. However, the cost of sequencing whole genomes has dramatically reduced over the last decade and continues to fall. With appropriate development of the right supporting infrastructure and a standard approach, the costs of whole genome sequencing (as opposed to genotyping) may already be at a level where it can be considered as a viable option, at least for some high value species with fairly long production lives such as cattle. It would also allow the generation of more detailed genomic information than is routinely collected in farming systems today.

The approach most commonly used for whole genome sequencing is called shotgun sequencing. This involves cutting multiple copies of the genome into many small fragments (of various lengths) and sequencing the individual DNA fragments. As the fragments are sequenced in a random order, some parts of the genome may get sequenced many times and others a smaller number of times. The sequenced fragments are then arranged, in a jigsaw like manner, against a template or scaffold called the reference genome. These high-quality reference genome sequences are species specific and are now available for the majority of livestock species, having been generated in recent years using far more advanced, precise and costly sequencing methods.

When arranging the jigsaw of short fragments against the reference sequence, each single attempt to map the entire genome (typically referred to as 1x coverage) may have gaps in it. However, by overlaying multiple genome coverage attempts, these gaps reduce in both number and size. When a high level of detail is required, it is typical to perform sufficient sequencing to allow around 30x coverage of the genome as this has been shown to ensure that virtually no gaps remain.

The depth of genome coverage being targeted has a large effect on sequencing cost. While a coverage of 30x is more ideal for an individual animal, when a large number of related animals within a population have their whole genome sequenced and imputation is also used, then a large amount of whole genome detail can be generated using relatively low depths of sequence coverage (for example, 1 to 5x). Indeed, as the number of completely unrelated animals is typically low within most livestock populations, sufficient genomic detail for most applications of interest could potentially be generated with a depth of coverage of as little as 1 to 2x or less. While the accuracy of imputation would be expected to increase over time as more related animals are sequenced, in establishing a new initiative, it would be helpful to sequence a few animals, such as highly used sires, to a higher depth in order to provide a strong starting platform. For some dairy bulls in particular, this information may already have been generated for use in breeding schemes.

By focusing on developing the whole genome sequence for each animal, the choice of which density SNP chip to use becomes irrelevant as information on all potential SNPs of interest would be generated. By adopting this approach, it also provides a high level of future proofing as our knowledge and understanding of the role of different SNPs continues to evolve and develop.

Although sequencing costs have significantly reduced over the last 10 years, the commercial cost of sequencing animals to a depth of 1 to 2x remains prohibitive for many end-users when considered in isolation. However, it is important to recognise that one important barrier to reducing sequencing costs is throughput. If a high number of animals were sequenced to a standard depth coverage of 1 to 2x as part of a structured national program, individual sequencing costs in the region of low tens of pounds may already be within touching distance. This potentially opens the door for more routine use of the information that could be generated, provided it could be accessed relatively quickly and easily, and sequencing costs discounted across a number of applications.

Real-time access to information and reduced testing costs

Rapid access to information is a benefit most people are now familiar with, and even expect, as they browse the internet on their smart phones or computers. This is made possible through information being stored in cloud-based storage systems. Many on-farm recording systems are already based around cloud storage, and given the rapid rate of on-going development, the possibility of accessing information on at least part of an animal's genome in real-time would not be difficult to imagine if full genome sequences were stored securely in a cloud based system.

When considering whole genome sequencing, it is important to remember that, within most livestock species, variations occur in only a small number of locations (SNPs), and this number reduces even further when variation is considered on a within-breed or on-farm population basis, thereby providing opportunities to minimise storage costs by storing only the variation. In addition, for most of the applications suggested previously, rapid access to genomic information would likely only be of interest for animals that are still alive or those whose ova or sperm have been conserved. This could also help further reduce overall storage costs, by allowing data for dead animals to be archived using lower cost storage options.

Once available in a cloud database, information on genome variations at sites known to affect specific characteristics of interest could be downloaded for use in genetic evaluations to support existing breeding schemes. Results for any genetic test could also be generated by test providers interrogating the cloud based whole genomes and reporting the status at defined SNP positions that had been identified from previous research to have a significant effect on the characteristic of interest in the target production environment. The results could also be reported in a similar manner to existing genomic tests. Once developed, most genomic tests for specific characteristics tend to focus only on

variation at a relatively small number of genome positions, which should further simplify interrogations and open the possibility of real-time decision support even from mobile devices. The test results could then be used for stand-alone decision support (*e.g.*, traceability or inherited disease status) or potentially combined with other data sources to provide even more informed and integrated decision support (*e.g.*, optimum feed strategies or the specific parasite an animal is infected with).

Having the whole genomes available from an early stage in the animal's life for interrogation on a cloud based system could not only eliminate the need to sample an animal multiple times but also reduce the cost to access and time to results for different individual SNP-based tests.

Building trust and buy-in through linking to traceability

One of the largest barriers to the effective sharing of data along a supply chain is trust. Within animal production there is a general concern, particularly among producers, that too much freedom to access specific farm data could result in the data being used to disadvantage stakeholders in early parts of the supply chain. Such concerns are particularly apparent within supply chains that are highly fragmented, as is the case for beef cattle and sheep in the UK (see New Technologies report). The development and maintenance of trust with all data providers along the supply chain would be vital for the success of any data sharing initiatives. Failure to do so, particularly with farmers, would likely reduce, (i) the quantity and quality of the data being collected, (ii) the level of benefits that could be generated, and (iii) the level of buy-in that could be expected for any similar future initiatives.

Where individual animal traceability schemes are well established, as is the case for cattle in the UK, this offers a potential means to help address at least some of those concerns. By establishing a national program for sequencing and linking the information to the traceability record, it could provide a means for the current owner of an animal to grant permissions as appropriate for genome access to specific third parties. By linking to the traceability record it could also provide a relatively easy means to determine which animals are still alive and/or which have genetic components (semen, ova) conserved for potential breeding use.

If the establishment and maintenance of such a program was at least part financed using public funding (either through subsidising data collection or establishing a public-private partnership, or publicly owned organisation to managed and run the activity), it could also offer a valuable source of data for addressing future research challenges that were in the national interest.

Although establishment of such a system could be of benefit in most livestock production systems, it will be most relevant for species where recording of individual identities is already routinely recorded or is planned. In the first instance, it would also be more likely to be successful if established for high value animals, particularly for those with relatively long expected production lives, such as cattle, where there would likely be a higher number of instances during the animal's life when knowledge about its genetic make-up may be of interest to different parties.

Taking DNA samples using Tissue Sampling Units (TSU's) close to birth at the time that animals are tagged for official registration for traceability presents an ideal sampling point, as the animal's genome sequence could then be available for interrogation throughout most of its life, allowing the greatest opportunities for potential use and benefits to be captured.

Opportunities to lower costs through collaboration

Particularly for cattle, tissue sampling using TSUs at or close to birth is not a new concept as it is already practiced as part of some diseases eradication programmes, such as part of the Bovine Viral Diarrhoea (BVD) eradication programmes. Whether the same TSU samples could be used to test for presence of the BVD virus and still generate enough DNA for sequencing is not yet clear, but, if possible, it could offer a clear opportunity for potential cost and labour savings.

Sample collection for DNA extraction and subsequent genotyping using a bovine 50k SNP chip on a large number of animals from commercial farms has also already been undertaken as part of the Beef Efficiency Scheme in Scotland. The experience gained in establishing and running that programme may also provide valuable lessons and insights into how a UK-wide genome sequencing programme could be effectively rolled out and managed.

Not a substitute for effective data collection

While establishment of a UK-wide genomic database could be a key step towards meeting many of the goals that have been set for livestock production in the UK, it will have little impact unless it is associated with systems for effective data collection and collation across the UK industry (as mentioned in the New Technologies report). The greatest benefit would be achieved if both approaches were developed in parallel, with the means for effective data transfer or joint interrogation between both.

As has been discussed previously (see New Technologies reports), there are clear benefits that could be accrued through establishment of an appropriate platform to support more effective data sharing and integration across livestock sectors and building upon the number of successes that have already been achieved to date. Examples include generating EBV's for resistance to bTB in cattle, by combining results from government coordinated bTB testing with data collected through BCMS and data collected through performance recording on-farms and through milk recording data collected through three commercial companies (NMR, CIS and Dale Farm). Similarly, EBVs for commercial carcass traits in purebred beef cattle have been generated by combining data collected in slaughterhouses on commercial animals with data collected through BCMS and on-farm performance recording of purebred animals. There are also two other excellent examples of data collected through centrally coordinated industry based commercial farm progeny testing being combined with data collected through performance recording on purebred animals. These are the Ram Compare project and the Beef Feed Efficiency Programme. Detailed analysis of these combined data has resulted in significant refinements to the selection indices used by pedigree breeders of dairy cattle, beef cattle and terminal sire sheep breeds in the UK. All are excellent examples of the benefits that can accrue when data from different sources are combined for evaluation and analysis (see the New Technologies report for more details).

It is important to emphasise that establishment of a UK-wide DNA database would not reduce the benefits that could be generated through more effective data sharing and integration, but its establishment could further increase the impact and benefits that could be generated, and thus provide added incentives for engagement and collaboration from a range of stakeholders across the industry.

Access to data from commercial farms would also be of particular interest in the pig and poultry breeding sectors for improving resistance to endemic diseases. Within both, elite breeding animals are maintained in very high health status environments, and therefore variation in the response of animals due to their genetic make-up when under challenges from various diseases is rarely observed.

The ability within their genetic evaluation systems to routinely use data from commercial production units where disease challenges are more common could generate real added value. If the data could also be combined with genomic information on the commercial animals, the potential benefit would be even higher and could transform the impact of breeding schemes in terms of generating animals even better suited to perform at a high level in commercial farm environments.

Even with effective DNA collection and sequencing, the success of any genome database program would depend heavily on the collection and availability of relevant high-quality phenotypic records. The availability of such data would be important both for developing relevant prediction models for any new test, and potentially for more routine decision support so that DNA information could be combined with other sources of data to increase the accuracy of the decision being made. For example, the accuracy of deciding how best to feed an animal to optimise efficiency and reduce environmental impact would likely increase if, in addition to information on age and genome content, access was also available to information on one or more of; current and future management conditions, health and feeding history, feed options, weather patterns or temperature, and in the case of ruminants the bacterial profile in the animal's rumen. As rapid advancements in Agri-tech continue to be made, more and more data sources are likely to become available.

Developing a platform for more innovation and entrepreneurship

Generating data to develop predictive models

Even with the availability of whole genomes for individual animals, there would still be a research need to develop additional prediction models to generate new genomic tests for novel traits of interest and importance.

The need for data to develop prediction models for various new genomic tests may present opportunities for some farms or farming groups to offer and operate on-farm data collection and recording services to generate data for test developers for specific measures of interest. Provided data of sufficient quality can be collected, the use of data from commercial farms as opposed to just research farms, could offer a means to effectively reduce the timeline to market as it could, if structured appropriately, incorporate a commercial farm validation step. The availability of such services, along with easier access to genome sequences for the animals of interest, may also provide a means for more companies to become engaged in the development of tests, thereby increasing the number of potential applications and opportunities to generate benefits.

There are already a number of successful examples of data being collected on commercial farms in the UK to support on-going research and development relevant to commercial producers, such as the Beef Feed Efficiency Programme and The Ram Compare project (see above and the New Technologies report). Both were managed and coordinated by Levy Boards (led by AHDB) and the lessons gained from establishing and running these programmes could offer valuable insights into how such approaches could be refined and strengthened so that high quality data can be generated more routinely, including that needed for further developments based on genomic data.

Providing permission to third parties to link to high quality records may also offer an opportunity for revenue generation for farmers, as well as a potential source of revenue that could help contribute towards the costs of maintaining the overall programme, while also providing a win-win scenario for many actors along the supply chain.

Developing a marketplace for easily accessible test results

As mentioned previously, the ability to access whole genome information in real time for each animal from a cloud based system could have a dramatic effect on both reducing the cost of accessing results for individual genome tests and the ease of their use in many situations. Rather than high up-front costs for users such as farmers or vets for each test, once the programme was established a smaller cost per interrogation for a genomic test could be used, with revenue generation opportunities for farmers and scheme providers during any test development phase, and for scheme providers and test developers once the test had been developed and marketed.

Once the tests were developed, it may also be possible for the platform to support a marketplace similar in many ways to that existing for the provision of application (apps) for smart phones on platforms such as the Apple apps store or Google Play if tests were offered in isolation or could be provided through existing on-farm software providers. However, the success of such an approach would likely depend on an effective means to verify the usefulness of specific tests offered for different circumstances also being established, to build both trust and demand.

As the global interest in precision agriculture continues to develop, establishment of a UK wide animal genomics database, along with effective data collection systems, and promotion of the wide range of environments, management systems and breeds that are already available, could help firmly establish the UK as a destination of choice for any new Animal Agri-tech companies wishing to establish a base and grow.

Blockchain and smart contracts, a potential method to develop and maintain trust despite increasing complexity

Over the last 5 years there has been a growing interest in the use of blockchain within the agricultural industry. Blockchain is essentially an electronic record that cannot be altered and continues to extend for an individual or unit as more relevant data is collected. The main objective is to improve traceability and reduce the opportunities for fraudulent acts to be carried out, thereby increasing the level of trust that can be generated along a supply chain ⁽⁴⁾. While still at a relatively early stage in terms of determining whether they can be used effectively in agriculture, some trial usage has been carried out or is currently underway within organisations such as the Food Standards Agency (FSA) ⁽⁵⁾, and a number of commercial companies with agricultural interests are now exploring potential applications (for example, within the cattle industry in Northern Ireland ⁽⁶⁾ and more widely across the industry ⁽⁷⁾), including the benefit of combining their use with genomic information ⁽⁸⁾.

A smart contract is an electronic method for automatically controlling or documenting any relevant event or action according to the terms of a contract or agreement, and interest in potential application in Agriculture has been growing in the last few years. It is increasingly being considered as a possible

⁴ Xiong *et al.*, 2020 Blockchain Technology for Agriculture: Applications and Rationale. *Frontiers in Blockchain*, 21 February (<https://doi.org/10.3389/fbloc.2020.00007>)

⁵ <https://www.food.gov.uk/news-alerts/news/fsa-trials-first-use-of-blockchain>

⁶ <https://www.nsf.org/news/new-consortium-uses-blockchain-technology-to-protect-northern-ireland-agriculture-products-in-world-markets>

⁷ <https://www.startus-insights.com/innovators-guide/8-blockchain-startups-disrupting-the-agricultural-industry/>

⁸ <https://www.ledgerinsights.com/blockchain-food-traceability-ripe-neogen-genomics/>

mechanism that could be used jointly with blockchain technology to help manage contract obligations, including any pre-agreed access to data, and any financial transactions that may be due as a result ⁽⁴⁾.

As more progress is made in developing precision agriculture approaches, the level of data capture, use and complexity in supply chains will increase, as will interest in developing and adopting approaches and methods to help manage the increased complexity. Whether the use of blockchain and smart contracts becomes commonplace across the Agriculture industry remains to be seen, but recent developments suggest that interest in their use is growing. However, in combination they could present an effective means to manage access to genome information for individual animals (held in a wide UK database) and other data sources along with any resulting financial transactions as appropriate. Developing such a capability could help establish a strong and effective platform for building trust and for supporting on-going developments and growth of the industry.

Opportunities not threats

When a new concept is being proposed, it is only natural for existing groups or companies already working in the space to view it, at least initially, as a potential threat. It is therefore important to point out that establishment of a UK wide animal genomics database on its own would not result in significant disruption for many of the existing approaches being used in the livestock production industry. For example, genetic evaluation for recorded populations would still be needed, but the database would offer a means to dramatically increase the impact that they could deliver across the industry, particularly for traits that are currently difficult to improve because either they have a low heritability or are difficult or expensive to measure (*e.g.*, require specialist equipment), or only appear sporadically - as is the case for some diseases. Similarly, management of genetics for specific breeds, that is currently mostly done by breed societies, will still be an important activity, but such a database could enable societies to more easily carry out their work through much easier and lower cost access to detailed genomic information on the populations being managed and conserved.

There is little doubt that establishment of such a database could result in some disruption for existing developers of stand-alone genomic tests. However, it could offer an ideal opportunity for them to adapt their business model and take advantage of an opportunity to lower their development and manufacturing costs. It could also dramatically increase the pool of potential customers both in the UK (through all farmers having easier access to whole genome information for all their animals) and in other countries (through marketing of test results developed in the UK more easily than would have been possible in the past).

Building alongside international developments and progress in other sectors

The feasibility of using genomic information to support the development of more effective precision farming approaches and methods is of global interest, and the potential benefits of international collaboration to help overcome challenges, particularly in relation to further development of underpinning genomic tools, should not be overlooked ⁽⁹⁾. However, as few countries have well established traceability systems for livestock, this could present an opportunity where the UK could take an international lead.

⁹ Rexroad *et al.*, 2019 Genome to Phenome: Improving Animal Health, Production, and Well-Being – A New USDA Blueprint for Animal Genome Research 2018–2027. *Frontiers in Genetics*, vol 10

Many of the underpinning approaches suggested above are already in development for use in maintaining and improving human health. They are also likely to be developed for use to support the health of companion animals over time, given the financial size of the global market. While each industry will have their own specific structure and challenges to address, it is important to recognise that the parallel developments may offer opportunities to learn from (at least in some areas) and adapt approaches appropriately for use in the livestock industry. Therefore, active sharing of ideas and maintaining a mutual awareness of on-going developments in these different sectors should be encouraged. As well as underpinning tools, one other area of common interest that is currently gathering momentum in the UK in terms of developing technical capability is in developing understanding in relation to microbiomes. This area is likely to play a key future role in health and efficiency of feed use in livestock ⁽¹⁰⁾ and is an area where detailed knowledge of the host animal's genome would be a real asset.

Particularly if the data were to be used for research, there are examples of recent or existing projects that have used whole genome information that may provide useful insights or lessons to consider when establishing a larger programme, such as the UK Biobank ⁽¹¹⁾ (human health) and the 1000 bulls genomes project ⁽¹²⁾ (cattle breeding).

With regard to underpinning genomic technologies our main focus in this paper has been on using shotgun sequencing and fragments assembly to generate a representation of the whole genome for individuals. While highly effective in characterising many forms of genomic information, it is acknowledged that the approach (as currently used) does have limitations, for example in accurately characterising copy number variants or detecting epigenetic effects. However, these challenges are not unique to the field of livestock genetics and is an area where future collaboration with other sectors to determine the most appropriate tools and approaches to use to generate more insights and clarity will be important.

Conclusions

Over the last two decades the cost of generating genomic information for individuals has reduced dramatically. As a result, it is now possible to envisage how taking account of the exact genetics an animal carries could become a frequently used tool in farming systems. It could be used not only to increase the accuracy of breeding decisions, but also as a key innovation component for driving the development of effective precision farming methods and approaches that could help bring benefits across a range of target areas, such as reducing environmental impact whilst improving animal health, welfare and overall production efficiency.

The costs associated with genomic technologies are expected to continue to fall. However, by developing a suitable infrastructure and approach now, they may already be at a level where they could feasibly be considered for more routine use for species such as cattle, and in time for other livestock species.

The development of a UK wide whole genome sequence database for livestock, linked to traceability and other effective data collection and collation systems, could play an important role in driving innovation of relevance to animal production, both in the UK and globally. By using such an approach,

¹⁰ <https://ktn-uk.org/news/ktns-microbiome-innovation-network-launches-the-microbiome-strategy-roadmap/>

¹¹ <https://www.ukbiobank.ac.uk/>

¹² <http://www.1000bullgenomes.com/>

the potential benefits from using genomic information could be made much more accessible to a wide range of farmers and other stakeholders across livestock sectors, not just those actively engaged in producing or using performance-recorded breeding stock. A national database of animal genomes linked to other national databases will create a national asset that will have value for decades.

Appendix 1. List of Abbreviations Used

Abbreviation	Definition
AHDB	Agriculture and Horticulture Development Board
BCMS	British Cattle Movement Service
BLAD	Bovine Leukocyte Adhesion Deficiency
bTB	Bovine Tuberculosis
BVD	Bovine Viral Diarrhoea
CIS	Cattle Information Service
CVM	Complex Vertebral Malformation
DNA	Deoxyribonucleic acid
EBV	Estimated Breeding Value
FSA	Food Standards Agency
NMR	National Milk Records
PTA	Predicted Transmitting Ability
SNP	Single Nucleotide Polymorphism
TSU	Tissue Sampling Unit