

# The development and application of genomic selection as a new breeding paradigm



André Eggen

Illumina Inc., San Diego, CA 92121

## Implications

- A significant gap exists between demand based on population growth and the current trajectory of yield; this is a point of leverage for genomics.
- Although traditional breeding methods have been effective in selecting for animals with easy-to-measure production traits, these methods have essentially “hit an asymptote,” and more difficult-to-measure (and often more important) traits cannot effectively be selected for using traditional methods.
- The race to sequence the first human genome, and, subsequently, the race to enable routine resequencing of tens, if not hundreds of thousands, of additional human genomes, has resulted in a 100 million fold decrease in DNA sequencing prices since 1990. Animal genome sequencing has benefitted from this.
- Sequencing and resequencing of economically important livestock species has resulted in the discovery of millions upon millions of single nucleotide polymorphisms. These single nucleotide polymorphisms are being deployed in massively parallel fashion on DNA microarrays, enabling genome-wide association studies to identify genotype-phenotype correlations for both simple and, more important, complex traits.
- Driven by ever-increasing reductions in the cost of measuring genetic variation, we are entering a new era in which the information from these genome-wide association studies will be utilized effectively in routine testing using genomic selection. Genomic selection holds promise for more widespread adoption than marker-assisted selection because it lacks the requirement of prior knowledge of alleles or marker positions of loci and the requirement that marker-assisted selection must be implemented within families.

**Key words:** gene mapping, genomic breeding, genomic selection, next-generation sequencing, single nucleotide polymorphism

## Introduction

In the past 50 years, the world has experienced an unprecedented increase in population growth. On the basis of current projections, the world

population will reach 9 billion people by 2030. Meeting the growing food need using fewer resources is therefore one of the greatest challenges that contemporary agriculture is facing. Recent estimations by the Food and Agriculture Organization of the United Nations (FAO, 2006) indicate that to meet the increasing demand, food production must double in the next 50 years. Expressed another way, agriculture will have to produce more food than in the last 10,000 years combined. When the year 2000 is used as a base, projections indicate an increase in global meat consumption of 68% and in global milk consumption of 57% by 2030 (Steinfeld and Gerber, 2010). The greater demand for food based on animal proteins together with the potential effects of climate change and water, nutrient, and energy scarcity will result in large productivity gaps. It is therefore critical to apply technical and scientific advancements systematically in feeding, nutrition, genetics, reproduction, animal health control, and general improvement of animal husbandry to fill the coming productivity gaps. The largest gains will come from innovations that accelerate agriculture productivity while reducing costs and limiting environmental impacts (Capper, 2011).

## Traditional Breeding Methods Are Successful but Limited

For centuries, animal breeders have very effectively manipulated the genomes of livestock species, making use of the fact that natural variations exist within a species, within a breed, and within a population. Traditional breeding has been done in the absence of molecular knowledge of the genes acting on a quantitative trait locus. Breeders have enhanced production traits in their herds by selecting superior individuals as progenitors for the next generations. These enhanced “breeding values” have been achieved by combining phenotypic recording of individual performance with genealogical information. In Holstein dairy cattle, milk production is still increasing by 110 kg per animal per year. In pig production, the kilograms of feed required to produce a kilogram of pork, known as feed conversion, is estimated to have decreased by 50% between the 1960s and 2005. Although these results illustrate powerful examples that can be achieved through traditional breeding methods, the efficiency of these traditional methods decreases when traits are difficult to measure, have a low heritability, or cannot be quickly, inexpensively, and correctly measured in a large number of animals. Such difficult-to-measure traits are often critically important; they include fertility, longevity, feed efficiency, and disease resistance. Selection for these traits must be achieved through genomic means.

## Big Expectations of Genomics

Over the past 2 decades, the rapid development of genomics has opened new paths to address the scientific basis of livestock biology and breeding, and has resulted in new production methods to achieve sustained increases in animal feed yields and long-term improvements in the efficiency of livestock production. A new era, the “genomics era,” promises to enable the objective prediction of consequences based on direct access to the full DNA sequence of many individuals, and therefore a renewed and more objective view of the genetic value of animals that is not limited to a few production traits.

One of the triggering factors for the development of this genomic era was the international project to sequence the human genome (the Human Genome Project). The goal of this project was to produce the first (de novo) full DNA sequence of a human being. Along with it came the development and implementation of new genomic tools, particularly improved DNA sequencing technologies and increased availability of high-throughput genotyping platforms. The price to sequence a single nucleotide of DNA has fallen 100 million fold since 1990. This is the equivalent of filling up your car with gas in 1998, waiting until 2011, and being able to drive to Jupiter and back twice. As Figure 1 illustrates, the rate of information coming from current-generation DNA sequencers is increasing exponentially, faster than the data from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), faster than Moore’s law, and faster than other comparable high-growth scenarios. Technological breakthroughs that have driven this cost decrease have also facilitated the production of whole-genome sequences for several animal species (Table 1).

## Genomics: Moving Animal Science to a New Dimension

From a scientific point of view, accelerated genome-scale measurement will have a profound impact. It will fuel comprehension of the basic structure and function of livestock genomes; help unravel the history of life, characterizing the cause of relatively simple phenotypes and genetic

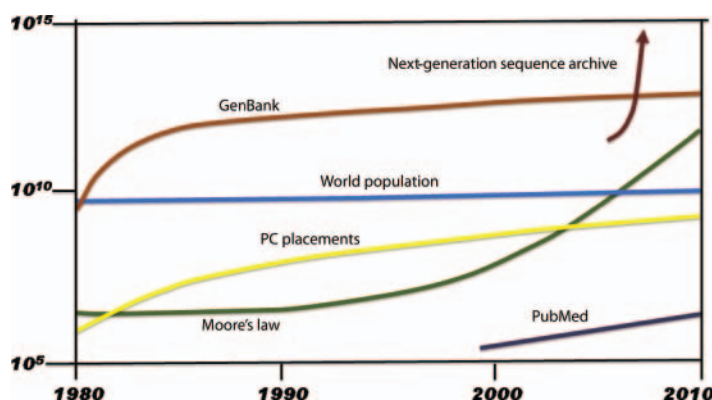
**Table 1. Summary of first sequenced genomes for animal species<sup>1</sup>**

Species	Genome size (assembly), Gb	Year
Chicken ( <i>Gallus gallus</i> )	1.05	2004
Dog ( <i>Canis familiaris</i> )	2.4	2003
Cattle ( <i>Bos taurus</i> )	2.91	2009
Horse ( <i>Equus caballus</i> )	2.47	2009
Pig ( <i>Sus scrofa</i> )	2.2	2009
Sheep ( <i>Ovis aries</i> )	2.78	2008
Cat ( <i>Felis catus</i> )	1.64	2006
Rabbit ( <i>Oryctolagus cuniculus</i> )	2.67	2009
Turkey ( <i>Meleagris gallopavo</i> )	1.08	2009
Dromedary ( <i>Camelus dromedarius</i> )	2.2	2011
Medaka ( <i>Oryzias latipes</i> )	0.7	2011
Honeybee ( <i>Apis mellifera</i> )	0.236	2011

<sup>1</sup>Modified from Fan et al. (2010).

diseases; and further explain the control of complex characteristic features. It should accelerate the discovery process and help close the gap between genetics and the traits that are observed, known as the “genotype/phenotype gap.”

With the production of whole-genome sequences for the major livestock species at a fraction of the cost of the Human Genome Project, the comparison of sequences from several individuals of different breeds with a reference sequence resulted in an almost inexhaustible source of genetic markers, primarily polymorphisms in the form of single nucleotide polymorphisms (SNP). With the release of the first draft of the chicken genome sequence (International Chicken Genome Sequencing Consortium, 2004), the chicken community was the first animal community not only to have access to millions of SNP, but also to be organized in a freely accessible database, the Chicken Variation Database (ChickVD; <http://chicken.genomics.org.cn/>). Similar sequencing and resequencing efforts in several other livestock species resulted in the discovery of hundreds of thousands of SNP covering the entire genome (Ramos et al., 2009). These genetic markers were subsequently deposited in a publicly accessible database maintained by the National Center for Biotechnology Information, the dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>). This database currently contains SNP information for more than 86 organisms and a total of more than 50 million SNP. Another major technological breakthrough has been the development and constant improvement of DNA array technology, which allows for the inexpensive measurement of SNP within a given sample. The success of these DNA arrays resides in the fact that they present a strong parallel-processing capacity, tremendous miniaturization, and a remarkable ability to be automated. Although first used for gene expression studies, these DNA arrays proved very useful for the development of whole-genome SNP panels for many species, including several agriculturally relevant species (Table 2). With DNA arrays, hundreds of thousands of SNP can be screened in parallel for the cost of a few hundred dollars, allowing scientists to perform genome-wide association studies that simply would have been out of reach with traditional (i.e., microsatellite) markers. Over the past years, several studies have been published dem-



**Figure 1.** The rate of information coming from current-generation DNA sequencers is increasing exponentially, faster than the data from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), faster than Moore’s law, and faster than other comparable high-growth scenarios ([http://www.ted.com/talks/lang/eng/richard\\_resnick\\_welcome\\_to\\_the\\_genomic\\_revolution.html](http://www.ted.com/talks/lang/eng/richard_resnick_welcome_to_the_genomic_revolution.html), courtesy of TEDxBoston 2011).

**Table 2. Currently available whole-genome single nucleotide polymorphism (SNP) chips developed for important agricultural species**

Species	Identification <sup>1</sup>	Classification	Provider <sup>2</sup>	Consortium	SNP, no.
Potato	Potato	Public	Illumina	SolCAP	8,303
Tomato	Tomato	Public	Illumina	SolCAP	7,720
Apple	Apple	Public	Illumina	RosBREED	8,788
Peach	Peach	Public	Illumina	RosBREED	8,144
Cherry	Cherry	Public	Illumina	RosBREED	5,696
Maize	MaizeSNP50	Commercial	Illumina	Commercial	56,110
Rice	Rice 44K	Commercial	Affymetrix	Commercial	44,100
Chicken	Chicken	Private: public sale	Illumina	Cobb Vantress-Hendrix-USDA	57,636
Cat	Feline	Private: public sale	Illumina	Morris Animal Foundation	62,897
Horse	Equine	Private: public sale	Illumina	Neogen (GeneSeek)	65,157
Sheep	Ovine	Private: public sale	Illumina	AgResearch	5,409
Cattle	BovineHD	Commercial	Illumina	Various	777,962
Cattle	BovineSNP50v2	Commercial	Illumina	Various	54,609
Cattle	BOS 1	Commercial	Affymetrix	Various	648,000
Sheep	OvineSNP50	Commercial	Illumina	Various	52,241
Cattle	BovineLD	Commercial	Illumina	Various	6,909
Pig	PorcineSNP60	Commercial	Illumina	Various	62,163
Dog	CanineHD	Commercial	Illumina	Various	173,662

<sup>1</sup>HD = high density; LD = low density.<sup>2</sup>Illumina Inc., San Diego, CA; Affymetrix, Santa Clara, CA.

onstrating the effectiveness of using low-cost, whole-genome SNP arrays (e.g., BovineSNP50, OvineSNP50, EquineSNP50, PorcineSNP60)

- For the fine-scale mapping of inherited defects when testing only a modest number of cases and controls (Becker et al., 2010; Brooks et al., 2010; Meyers et al., 2010).
- For the mapping of quantitative trait loci and DNA regions involved in the genetic mechanism of complex phenotypes, to discover and estimate marker effects (Fortes et al., 2010; Minozzi et al., 2010; Pryce et al., 2010).
- For the characterization of population structure, to better document and understand the history of our livestock populations (Decker et al., 2009; Kijas et al., 2009; Gautier et al., 2010).

The availability of whole-genome SNP panels is therefore bolstering the search for mutations underlying genetic variation in simple and complex traits and revolutionizing the speed at which gene regions and specific genes are being discovered. With the use of whole-genome SNP panels, the traditional approach, namely, positional cloning with a whole-genome scan to map the region of interest, followed by a time-consuming fine-mapping step, is fully replaced with an efficient and cost-effective genotyping step using a whole-genome SNP array containing 50,000 to 70,000 SNP.

## Genomics: A Paradigm Shift in Animal Breeding (Toward the Genome-Assisted Barnyard)

The biggest (r)evolution is taking place in the application of genomics to the design and implementation of livestock breeding programs, promising gains across the value chain. For breeders, breeding organizations, and members of the livestock industry, genomics is expected to increase the efficiency and productivity of animal breeding, whereas for consumers and the processing sector, it should enhance security and the quality of animal products. New insights into the growth, nutrition, health, and protection of animals are expected, enabling a better understanding of the molecular mechanisms of traits of interest. Therefore, genomics proposes further opportunities to improve selection accuracy while decreasing the costs, reducing generation intervals, and exploiting new sources of polymorphisms (Dekkers, 2004).

### From Marker-Assisted Selection to Genomic Selection

Beginning in the 1990s, breeders used gene marker technology in the form of marker-assisted selection (MAS) to remove deleterious gene alleles (such as halothane in pigs or fish taint in brown-shelled chicken eggs) or to select favorable conditions based on some marker information (such as the Polled condition in cattle). The limitation of MAS is that it

requires prior knowledge of gene alleles or markers that are associated with the traits of interest together with quantitative estimates of these associations in the specific population. It must therefore be implemented within families. Furthermore, MAS explains only a limited part of the genetic differences between individuals. However, with the availability of cost-effective whole-genome SNP panels for the major livestock species, genomic selection tools are now leading the way to a paradigm shift in animal breeding. With sufficient genetic markers, one can follow the segregation of the entire genome and not merely a set of specific regions of interest, moving from MAS to genomic selection. Parental relationships are no longer required to explain similar performances in animals because, with a dense set of SNP, similar performances can now be explained by the fact that animals are sharing identical chromosome fragments.

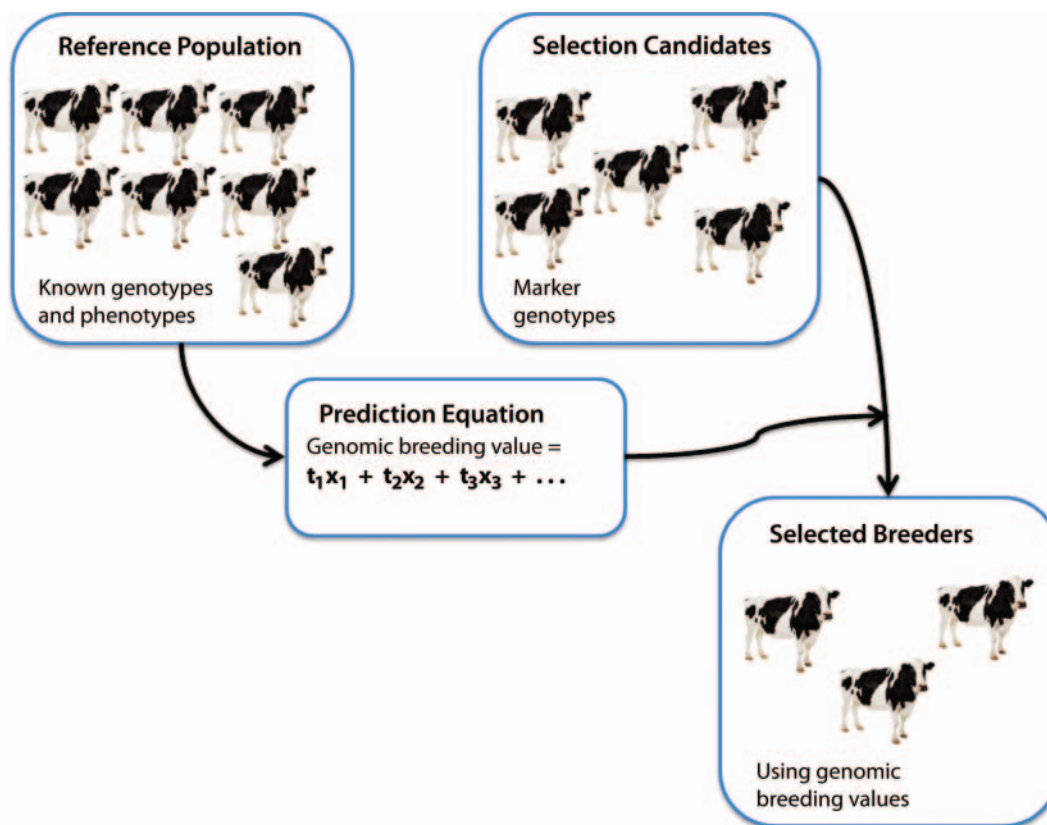
### Principle of Genomic Selection

Genomic selection was first described by Meuwissen et al. (2001) and is based on the fundamental principle that information from a large number of markers could be used to estimate breeding values without having a precise knowledge of where specific genes are located on the genome. With tens of thousands of SNP, well chosen to be representative of the entire genome, it is expected that there will always be an SNP in close proximity to a particular gene or DNA fragment of interest; the existing linkage disequilibrium between one (or several) SNP and a causal

mutation will be substantial and can then be used to explain a significant fraction of the variation of the observed trait. The first step in the genomic selection process is therefore access to a large group of animals, either a reference or training population with accurate phenotypes for the trait(s). This population should also be genotyped using a whole-genome SNP array. The resulting data will then serve as a reference to develop a statistical model estimating the effect of each SNP with the trait(s) of interest. The result is a predictive equation to calculate a genomic estimated breeding value (GEBV). After a validation step, the genomic breeding value of new animals can be computed using the prediction equation, based on their genotypes from the SNP array and in the absence of any accurate phenotypes for these animals (Figure 2 and Table 3). The accuracy of the GEBV depends of the size of the population and the heritability of the trait to be considered. In the chicken, for example, González-Recio et al. (2009) referenced a 4-fold increase in GEBV accuracy over the parent average for feed conversion efficiency.

### Implementation of Genomic Selection

Genomic selection builds on existing breeding programs in which the collection of pedigree information together with phenotypic data is already routine; it provides a new level of information that can be integrated into the decision-making process to identify and select the most promising animals. The principal advantages of genomic selection are that it can be



**Figure 2.** A reference population of animals is scored for key production traits and genotyped using a commercial or custom single nucleotide polymorphism array. The genotypes are represented by the variable  $x$ , with values 0, 1, 2 (homozygous, heterozygous, or alternate homozygous). A prediction equation is generated, combining all the marker genotypes with their effects to compute a genomic estimated breeding value for each animal. This prediction equation can be applied to a group of animals that have not been phenotyped, breeding values can be estimated, and the best animals can be selected for breeding. Adapted from Goddard and Hayes (2009) by permission from Macmillan Publishers Ltd: Nature Reviews Genetics, © 2009.



**Table 3. Example of a simplified calculation of the genomic breeding value with 4 single nucleotide polymorphisms (SNP) and estimated effects (allele A vs. B) of +8, +4, +2, and –6 for SNP 1, 2, 3, and 4, respectively**

Animal	SNP 1		SNP 2		SNP 3		SNP 4		Genomic breeding value
	Genotype	Value	Genotype	Value	Genotype	Value	Genotype	Value	
1	AA	8	BB	–4	AA	2	AA	–6	0
2	AA	8	AA	4	BB	–2	AB	0	10
3	AB	0	AB	0	AB	0	BB	6	6
4	AB	0	AB	0	AB	0	AA	–6	–6
5	BB	–8	AA	4	AA	2	AA	–6	–8
6	BB	–8	BB	–4	BB	–2	AB	0	–14

implemented very early in life, is not sex limited, and can be extended to any traits that are recorded in a reference population. It especially provides for difficult-to-improve traits and better selection accuracy while reducing the generation interval, thereby increasing the intensity of selection. It explains a much greater proportion of the genetic variance than MAS and, unlike MAS, is not limited to specific families.

As early as 2006, Schaeffer (2006) showed that using genomic selection, the genetic gain per year could be doubled in dairy cattle, with a potential to reduce costs for proving bulls by more than 90%. Rather than cattle going through a long and expensive progeny test with the recording of phenotypic information on large numbers of daughters, accurate GEBV could be calculated through a cost-effective genotyping step.

In cattle, more than 15 countries are now using genomic breeding values on the national level and have successfully passed the international GEBV test organized by Interbull (<http://www.interbull.org/>). The evident benefits observed in dairy cattle can also be translated to other species, and the broader use of genomic approaches, or even genomic selection, promises to be one of the next major advances in breeding programs for all animal species. As cost-effective genomic tools continue to be developed for livestock, crop, and aquaculture species, the corresponding breeding industries will be able to make selection decisions sooner, improve traits that are difficult to address with traditional breeding methods, and provide consumers with higher quality, safer foods while reducing the impact of breeding on the environment and ensuring its long-term sustainability.

Applications of genomic approaches in the field extend well beyond breeding because genomic tools can also provide pertinent and accurate information on animal identification, validation of parental relationships (parentage), and traceability. Genomic approaches can be applied to the herd management process in optimizing mating and reducing inbreeding.

### Next Wave of Discoveries and New Approaches

Higher density SNP arrays with several hundred thousand SNP are already being developed in several livestock, crop, and companion animal species. In cattle, the success of genomic selection is being extended by combining several breeds to increase the size of the reference population and perform cross-breed evaluations. This approach should prove very beneficial for breeds with a limited number of individuals or phenotypic records, or for species for which cross-breeding is an effective tool in the breeding process.

As the cost of sequencing continues to decrease and access to the whole-genome sequence for specific individuals becomes affordable, one of the next steps will be to include whole-genome sequence data in routine genetic evaluations. According to a simulation presented by Meuwissen and Goddard (2010), a 40% gain in accuracy in predicting genetic values could be achieved by using sequencing data instead of data from 30,000

SNP arrays alone. Furthermore, by using whole-genome sequencing data, the prediction of genetic value was able to remain accurate even when the training and evaluation data were 10 generations apart: observed accuracies were similar to those in which the test and training data came from the same generation. According to the authors, “these results suggest that with a combination of genome sequence data, large sample sizes, and a statistical method that detects the polymorphisms that are informative . . . , high accuracy [in genetic/genomic prediction] is attainable” (Meuwissen and Goddard, 2010, p. 630).

However, genomic selection today still treats the genome as a “black box.” It is not necessary to understand what is inside the black box to make effective selection decisions. From this point of view, it is therefore not much different from the broadly used and accepted best linear unbiased prediction method. Genomic selection could be further improved by integrating pertinent biological information and using efficient methodologies to get from the knowledge of a statistically associated marker locus to a functional gene variation. This would move from simply associating sequence variation with distinct phenotypes to a true understanding of the biology of the animal that makes those variations significant, effectively shedding light on the black box.

### Opportunities for Developing Countries

In developed countries, phenotypes and pedigrees have been recorded for certain species, such as dairy cattle, for more than 100 years. Progeny testing has been implemented for nearly 50 years. Developing countries are often limited by the absence of programs that record phenotypes on pedigreed animals and the lack of evaluation or national testing programs to assess the genetic value of germplasms. Genomic approaches should help in identifying critical populations for preservation together with some local well-adapted breeds that could be further utilized to breed valuable animals through a combination of selection and cross-breeding. Of course, as with genomics, you can manage only what you can measure, and collecting a minimum number of phenotypes in the field will remain one of the critical and challenging steps to further deployment of genomic selection in developing countries.

### Conclusion and Perspectives for the Future

The ability to investigate the genome, the transcriptome, the epigenome, and the metagenome of any species by high-throughput sequencing methods is opening a new world of possibilities. Further reduction in sequencing costs will continue to drive broader acceptance of new approaches and their implementation for the benefit of animal research, the breeding industry, and consumers. All economically impactful agricul-

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.