Advances in porcine genomics and proteomics—a toolbox for developing the pig as a model organism for molecular biomedical research

Emøke Bendixen, Marianne Danielsen, Knud Larsen and Christian Bendixen

Abstract
Our current knowledge of human biology is often based on studying a wide range of animal species. In particular, for understanding human diseases, the development of adequate animal models is of immediate importance. Although genetic strains and transgenic animal model organisms like fruit fly (Drosophila), zebrafish and rodents are highly informative about the function of single genes and proteins, these organisms do not always closely reflect human biology, and alternative animal models are thus in great demand. The pig is a non-primate mammal that closely resembles man in anatomy, physiology and genetics. Pigs, although not easily kept for laboratory research, are, however, readily available for biomedical research through the large scale industrial production of pigs produced for human consumption. Recent research has facilitated the biological experimentation with pigs, and helped develop the pig into a novel model organism for biomedical research. This toolbox includes the near completion of the pig genome, catalogues of genes and genetic variation in pigs, extensive characterization of pig proteomes and transcriptomes, as well as the development of transgenic disease models. The aim of this review is to highlight the current progress of these ongoing areas of research, which are mandatory for successful development of biomedical pig models that are in demand for understanding human biology in health and disease.

Keywords: pig; biomedical models; genomics; proteomics; transcriptomics; transgenics

INTRODUCTION
To further the understanding of human diseases, and for the development and validation of new therapeutics, the use of adequate animal models is of immediate importance. Though inbred and transgenic strains of classical model organisms like fruit fly (Drosophila), zebrafish and rodents have proven informative regarding the functions of single genes and proteins, it has become increasingly clear that many of these organisms insufficiently reflect human biology leading to a demand for alternative models. Even the small mammalian models of human diseases often do not faithfully mimic the human conditions. Examples are plentiful, e.g. rodent models of Parkinson’s disease (PD) which do not recapitulate human pathophysiology [1–3].

Although not classically thought of as an obvious model organism, the pig has recently become increasingly relevant as a model organism for biomedical research. This is mainly due to the fact that the
anatomy, genetics and physiology of pigs reflect human biology more closely than the classic animal models. In addition, pigs are easy to breed, they produce large litters, are available in a large variety of genotypes and phenotypes, and provide easy access to biopsies and post-mortem samples. Miniature pigs have already proven particularly valuable in biomedical research [4].

The size of pigs obviously makes them less suited for keeping and studying in laboratory facilities than the rodent model, but the agricultural industry produces 1 billion pigs every year globally [5]. Thus, industrial pig production represents a valuable resource from which experimental animals can be selected due to the following considerations: (i) the health and biological phenotypes of animals that are selected as breeding stock are monitored in great detail, (ii) the breeding stocks are genetically well characterized [6], (iii) extreme phenotypes have been created through domestication and genetic selection and (iv) industrial pig production systems can provide exceptionally large populations of pigs, which makes them a very strong resource for finding and describing rare genetic and phenotypic variations, in particular for muscle growth, fat deposition and metabolism [7]. Making full use of this exceptionally large biological resource for developing useful biomedical animal models has so far been hampered by the relative lack of information about the molecular biology of the pig, which does not currently match the vast amount of biological information available on classic model organisms like drosophila [8] and rodents [9].

However, within the past decade, this situation has changed at a fast pace, and much recent effort has been devoted to characterizing the structural and functional genomics of pigs. Of particular importance are the near completion of the pig genome sequence [10] and the detailed genomic mapping and description of genetic variation in pig populations which have been developed. A wide range of gene expression and proteome studies have been presented that greatly contribute to characterizing these animals at the molecular level, which is needed to establish adequate biomedical models.

Furthermore, the fields of transgenics and cloning technologies in the pig have been under development, and substantial progress has been reported within the past decade. The collection of these technologies creates a strong toolbox for future biological experimentation with pigs. These technologies are of great importance for developing porcine model animals suited for biomedical research. Such porcine models may in particular be well suited for finding better diagnostic markers, and for the development of new therapeutic agents and examination of their efficiencies.

The aim of this review is to summarize the recent progress in characterizing structural and functional genomics and proteomics in pigs, and highlight the importance of these fields for efficient progress in developing porcine models for biomedical research. A short overview of key technologies and terminology in the fields of transgenics, genomics and proteomics are summarized in the terminology boxes [1–3].

CURRENT PROGRESS IN PIG GENOMICS

The size and composition of the porcine genome is comparable to that of humans. It is expected to comprise ~2.7 billion basepairs [11], and both gene content and sequence are highly conserved relative to the human genome. Comparison of genomes encompasses sequence similarity, gene location, the length and number of coding regions within genes, as well as the amount of non-coding DNA in each genome. The extent of synteny between pig and man as exemplified by alignment of human chromosome 3 and pig chromosome 13, indicates almost identical gene contents, although many gene-order differences are observed [12].

Studies providing a low resolution picture of genomewide syntenic relationships between pig and man has been published based upon physical maps or linkage maps [13, 14]. Preliminary comparative sequence analysis have been performed [15, 16], but detailed comparative analyses are still hampered by the lack of complete annotation and assembly of the porcine genome. Never the less, genomic comparisons between the pig and human show more structural resemblance than, for example, mouse and human [11, 13, 17].

The need to characterize the genetic basis of economically important production traits in pig has been a strong driver for the fast advance of the contemporary tools and resources of structural and functional genomics in pigs. These include the sequencing of the porcine genome, characterization of genetic markers with high density, as well the wide range
**Box 1: Proteomics terminology**

**Mass spectrometer (MS):** instrument for accurate measurement of mass-to-charge ratios of charged molecules (ions). In proteomics, peptide ions (often derived from trypsin digestion of proteins) or less frequently protein ions are analyzed.

**MS/MS spectra:** generated in tandem mass spectrometers, which are able to select a specific peptide from a complex mixture of peptides, and fragment it into a series of product ions. The measurement of the peptide fragments is called a “fragment ion spectrum” or an MS/MS spectrum and contains (partial) peptide sequence information.

**LC-MS/MS:** Commonly applied proteomics approach based on LC separation of peptides (derived from trypsin digested proteins) followed by peptide analysis in a mass spectrometer capable of acquiring MS/MS spectra.

**Mapping-proteomics:** Qualitative proteome analyses are usually termed “mapping” or “cataloging” proteomics. This is a descriptive and not quantitative analysis, which is an important fundation for defining tissue types, and achieving information about which subsets and amounts of proteins that are normally expressed in different tissues.

**Targeted proteomics:** proteomics approach focused on preselected proteins or peptides, e.g. SRM.

**SRM:** (selected reaction monitoring) [54] a targeted proteomics approach, where only specific preselected peptides are analyzed. SRM requires preexisting knowledge about protein specific (proteotypic) peptides in the sample and how these peptides perform in the mass spectrometer. The benefit is that once an SRM method has been built, the specific peptides of interest can be quickly and consistently analyzed in many samples.

**Quantitative proteomics:** comparative and differential analysis of proteomes at different conditions, challenges or disease states. In 2DE, quantitation is based on spot densitometry, often in combination with fluorescent dyes (DIGE) [130]. In LC-MS/MS platforms, quantitation is performed by either metabolic (SILAC) [51] or chemical labeling (ICAT, iTRAQ) [52, 53] or by internal standard (AQUA, QconCAT) [55, 131] approaches.

**Relative quantitation:** quantitative proteomics which refers to comparative analyses, where proteome expressions are compared across e. g. different disease states, and information about relative quantitative changes in protein expression can be achieved as fold changes, not absolute values.

**Absolute quantitation:** is still very difficult to achieve at a panoramic proteome level, although much progress has been reported lately. Relies on the use of spiked-in internal standards, such as AQUA and QconCAT [55, 131].

---

**Box 2: Genome terminology**

**Bacterial artificial chromosome (BAC):** A DNA construct which is based on a functional fertility plasmid from *E. coli*. A BAC can harbor a large (150-300 kbp) insert from any organism. BACs have proved useful in the sequencing of many genomes.

**Expressed sequence tag (EST):** A unique DNA sequence within a coding region of a gene. An EST is a sequence-tagged site derived from cDNA. ESTs have applications for identification of full-length genes and mapping of genes.

**Quantitative trait locus (QTL):** A sequence of DNA associated with a particular phenotypic trait. QTLs can be used to identify candidate genes underlying a trait and in combination with gene expression profiling (eQTLs) reveal cis and trans controlling elements.

**Restriction fragment length polymorphism (RFLP):** DNA fragments created by digestion with a specific restriction endonuclease. The DNA fragment pattern, polymorphism, is characteristic of the particular DNA molecule. RFLP can be used in parentage determination and in DNA fingerprinting to identify individuals.

**Single nucleotide polymorphism (SNP):** A variation in the DNA of individuals, specifically a variation of one nucleotide base. SNPs can occur in both coding and non-coding regions of a gene. An SNP in coding regions can lead to amino acid substitutions and SNPs in non-coding regions can affect binding of transcription factors and gene splicing.

---

**Box 3: Transgenesis terminology**

**Somatic cell nuclear transfer (SCNT):** A technique where the nucleus of a somatic cell is removed and inserted into the enucleated egg cell. After insertion into the egg, the somatic cell nucleus is reprogrammed by the host cell. The resulting blastocyst is transferred to a surrogate animal giving birth to cloned offspring.

**Sperm-mediated gene transfer (SMGT):** SMGT is a technique based on the intrinsic ability of sperm cells to bind and uptake exogenous DNA and transfer it to oocytes by fertilization.

**Lentivirus mediated gene transfer (LVMT):** Lentivirus (belongs to the Retroviridae family) can be used as a tool to introduce a gene product into in vitro systems or animal models by infection of a zygote or any other tissue. The most efficient technique for generation of transgenic animals.
of platforms and assays that has been developed specifically for porcine transcriptome analyses.

Genetic maps
Linkage maps are essential tools for locating genes and quantitative trait loci (QTLs). The first porcine linkage map covering all 18 autosomes of the pig was published in 1995 [18], followed by a large map containing ~1200 markers [19]. These maps were primarily constructed on the basis of anonymous microsatellites and restriction fragment length polymorphism (RFLP) markers [18–20]. Human-pig comparative maps based on ESTs exclusively [21] as well as ESTs in combination with bacterial artificial chromosome (BAC) end sequences [22] have been constructed using the INRA-Minnesota porcine Radiation Hybrid panel, ImRH7000 [23]. Moreover, a high quality BAC map of the pig genome has been developed [24].

Recently, single nucleotide polymorphism (SNP)-based linkage maps of all 18 porcine autosomes were constructed based on primarily gene-associated SNPs. The total length of the sex-averaged whole porcine autosome was estimated to 1711.8 cM resulting in an average SNP spacing of 3.94 cM [14]. Studies revealing the frequency and size distribution of structural genetic variation in pigs using Nimblegen arrays have also recently been published [25].

Efforts by an international porcine SNP-consortium have resulted in the de novo identification of more than 375,000 SNPs [26, 27] of which 60,000 have been included in a commercial genotyping tool, PorcineSNP60 Beadchip, which is commercially available (Illumina Inc.).

Genome sequence and cDNA resources
The porcine genome comprises 18 autosomes and the X and Y sex chromosomes. The genome size is approximated to 2.7–2.9 Gb. Sequencing of the porcine genome has progressed in several stages during the past decade.

From 2001 to 2004, Danish and Chinese researchers completed a pilot ‘whole genome shotgun’ sequencing project resulting in 0.66× coverage of the porcine genome [16]. In 2003, ‘The Swine Genome Sequencing Consortium’ (SGSC) was formed to coordinate international efforts for sequencing the pig genome. Using a high quality physical map of the pig genome [24] as a template for sequencing, the Wellcome Trust Sanger Institute has almost completed clone-based sequencing of the pig genome. The first version of the ‘high-coverage Sscrofa9 assembly’ for chromosomes 1–18 and the X chromosome was released in April 2009.

Gene expression analyses in pigs
Gene expression analysis is often essential in the process of establishing and validating animal models as well as being an important tool for subsequent research. The rapid progression of functional genomics has introduced novel technologies capable of generating quantitative measurements of gene expression in pigs. Within the last decade microarray technology has been the popular choice to achieve quantitative assessment of gene expression levels for thousands of genes simultaneously [28]. Many research groups have been focusing on tissue specific or otherwise specialized microarrays, made by printing cDNA fragments expressed in the tissue of interest, e.g. [29–31]. However, construction of genome wide arrays requires access to sequences or cDNA clones expressed in many tissues and developmental stages in order to obtain adequate gene representation. As part of the Sino–Danish Pig Genome Sequencing Project, a large EST sequence and cDNA clone resource have been established and analyzed [15, 32]. This resource comprised more than ~1 million cDNA clones from 96 different cDNA libraries and was used to develop a microarray platform that was applied in a wide range of studies, including gene expression in testes [33], liver [34] and in diseased lung tissue [35] and to establish a catalogue of the global expression of genes across a collection of 23 different tissues [36]. A long-oligonucleotide array platform has also been established for pigs [37]. To conduct microarray experiments in pigs several commercial arrays are available. Agilent Technologies offers an expression array with 43,803 60-mer oligoprobes (http://www.chem.agilent.com) and Affymetrix offers a GeneChip Porcine Genome Array (http://www.affymetrix.com) with 23,937 probesets which are able to interrogate the expression of more than 20,000 transcripts in pigs.

Recent developments in sequencing technologies [38, 39] have enabled massively parallel sequencing of cDNA fragments, and the frequency of sequence tags resulting from a particular gene has been shown to provide reliable measures of gene expression. These methods are still in their infancy and only few studies concerning porcine gene expression
have so far been reported. Hornshøj et al. [40] reported the use of 454-sequencing to study gene expression in cardiac and skeletal muscle. However, it is to be anticipated that sequencing-based technologies will soon replace microarray technology, mainly because sequence-based transcript analyses provide much additional information, including SNP-information, alternative splice variants, as well as the ability to identify and interrogate novel transcripts. These data are an important part of the picture when aiming to characterize the function of genes and their roles in health and disease.

**CURRENT PROGRESS IN PIG PROTEOMICS**

Pathogenesis is often associated with changes in expression, modification or stability of cellular proteins. Proteomics is aimed at analyzing the complex networks of proteins expressed in cells, tissue samples and body fluids, and thereby achieving a ‘snapshot’ of the protein-status of a complex biological sample. Comparative proteome analyses are often used as a step towards more detailed understanding of molecular mechanisms behind biological processes or a diseases mechanisms, to identify biomarkers that can be useful for early diagnosis [41] or for the development and examination of new therapeutics.

When describing the function of genes, proteomics is a very important complementary analysis to transcriptome studies, partly because proteins are the ultimate product of genes, but maybe most importantly because many cellular mechanisms depend on post-translational modification of proteins [42] and specific protein interactions [43]. These events are not directly reflected by mRNA expression.

Moreover, it has been estimated that a mammalian genome of 22,000–25,000 genes probably expresses more than 1,000,000 different proteins, when post-translational modification and splice variants are taken into account [44]. In order to study as many as possible of these proteins in a single display, advanced methods have been developed that allow protein extraction, separation and identification of such complex biological samples [45, 46].

**Current technologies**

For the past 20 years, 2D gel electrophoresis has been the cornerstone of proteome analysis [47]. The method has the advantage of presenting a global image of the analyzed proteome, and is well suited for investigating specific post-translational modifications, e.g. glycosylations [48]. However, today LC–MS (liquid chromatography–mass spectrometry) based strategies have become more commonly used for global proteome studies, as the integration of separation and identification of proteins allow both automation and greater speed of analysis. The number of proteins which can be identified in a proteomics experiment varies immensely, depending on sample type, fractionation methods, instrumentation and the amount of genome data available. However, under optimized conditions between 3 and 5000 unique proteins can currently be identified from a sample [49, 50].

Quantitative analysis can also be applied in high through-put LC–MS mode by the use of protein or peptide labeling techniques, including SILAC [51], iTRAQ [52] and ICAT [53].

The latest generation of proteome quantitation technology is based on a targeted proteomics approach [54] known as SRM (selected reaction monitoring) in combination with spiked-in standards [55]. SRM is based on analyzing two to three preselected peptides [56], which uniquely represent a specific protein (i.e. proteotypic peptides).

This strategy allows rapid serial analyses of many specific proteins, and thus larger numbers of biological replicates to be analyzed for proteome changes. Targeted proteomics is therefore a promising approach for overcoming some of the current limitations in proteomics, namely the investigation of proteomics variation in large populations and sample sets [57] and analysis of a wider dynamic range of proteins in a sample [58].

**Proteome data repositories**

LC–MS analyses produce enormous amounts of protein and peptide data, which have even greater value if collected in searchable data repositories [59, 60]. Some of the major databases include Pride (http://www.ebi.ac.uk/pride), PeptideAtlas (http://www.peptideatlas.org/repository), GPMDB (Global proteome Machine Database (http://gpmdb.thegpm.org/) and the NCBI-peptidome (http://www.ncbi.nlm.nih.gov/projects/peptidome). For current reviews of contents and use of these repositories, see Riffl and Eng [60] and Mead et al. [61]. These repositories largely cover human proteomes and most classic model organisms like yeast, mouse and drosophila, but so far only the PeptideAtlas covers
the pig proteome well. The pig proteome database as presented at the PeptideAtlas repository currently covers 20 pig tissues, and contains more than 15,000 peptides, representing more than 4000 unique proteins.

Data repositories are an important tool when building targeted proteomics methods [62] as knowledge of the peptide’s physiochemical properties is required. However, once constructed, targeted proteomics methods are more accurate, faster and can handle increasing numbers of biological replicates. These data collections are also an important resource for developing methods for absolute rather than relative measurements of proteins [63].

Proteome maps and biomarkers

Mapping and cataloguing the entire proteomes of tissues and body fluids is fundamental for characterizing the subsets of proteins that are normally expressed in the different tissues and cell types. Extensive knowledge of proteomes also greatly facilitates the search for biomarkers, which is an important aim in biomedical sciences. Numerous individual proteome maps of human tissues and body fluids have been presented through the past two decades [64]. However, proteome maps of many porcine tissues have not yet been well characterized, although this field is progressing rapidly. The pig proteomics field was initially closely related to characterizing the biological traits related to pig production, such as meat quality traits [65–67] and pig health [68–70]. In particular gastrointestinal health and lung infections (pneumonia), which cause high mortality rates in pig production, have been studied [71–74].

However, the most recent literature reflects that the majority of pig proteome studies currently has a clinical scope although it must be emphasized that the achieved data often have a dual relevance, namely for both biomedical research and the farm animal industry. Most prominently these areas include models on reproduction, embryology, host-pathogen interactions and vaccine development. With the increasing availability of pig models, the knowledge about the pig proteome is expected to grow rapidly in near future.

PORCINE BIOMEDICAL MODELS

Already existing pig models of human diseases encompass a broad array of biomedical topics such as: heart physiology, reproductive function, skin physiology, transplantation, gut physiology and nutrition, brain function, biomechanical models, tissue engineering, respiratory function and infectious disease models (for review see Lunney 2007) [75]. Pigs have for a long time been recognized as excellent disease models in a broad variety of diseases including dermatology, diabetes, eye disease, degenerative disease and skeletal growth. Information has been derived from pig biomedical models investigating heart physiology, e.g. ventricular hypertrophy [76], surgery and endoscopic techniques, and diabetes. Valuable data have been obtained about skin cancer by the study of a pig melanoma model, the Melanoma-bearing Libechov mini-pig, and during the past 20 years a great effort has been put into research of the use of the pig in xenotransplantation.

Natural and induced models

Porcine models have already been developed for studying human pathophysiology, including the fields of obesity, diabetes, cancer, female reproductive health, cardiovascular disease and infectious diseases. Both natural and induced models have been used. Two examples of natural pig models are (i) a model of Schmid metaphyseal chondrodysplasia (SMCD), a mild skeletal disorder associated with dwarfism, caused by a mutation in the COL10A1 gene [77] and (ii) a model of hypercholesterolemia identified through a mutation in the lipo-protein receptor (LDLR) [78].

An important example of an induced model is the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced Parkinsonism. An accidental finding demonstrated that this dopaminergic neurotoxin induces PD in humans [79] and other primates [80–82]. Pigs have also been included in these studies and a Gottingen mini-pig represents a non-primate model of MPTP-induced Parkinsonism syndromes being reversible and lasting at least months [83]. The porcine, primate and mice MPTP-based PD models have resolved important aspects of the pathophysiology of PD [84–87].

Transgenic pigs as disease models

Due to the recent technological advances in cloning and transgenics, the application of genetically modified and cloned pigs in translational research has increased [88, 89]. Since the first cloned animal, the sheep Dolly [90], somatic cell nuclear transfer (SCNT) has been successfully implemented in a
large variety of animals [91–96]. Cloned animals have identical genomes and are therefore of potentially great benefit to biological sciences [97]. Cloning of pigs for use in biomedical research has recently received much interest [75, 98]. Cloned animals in particular are invaluable for studying diseases and biological traits in which the interplay of genes and environmental effects are complexly intertwined, including severe lifestyle-related disorders like obesity [99], depression [100] and cardiovascular disorders [101].

Transgenic pigs have been generated by different techniques such as microinjection, sperm-mediated gene transfer (SMGT) [102, 103], virus-mediated gene transfer [104, 105] and nuclear transfer and cloning [106, 107]. The first transgenic pig made as a model of human disease was a pig retinitis pigmentosa model [108]. The pig model recapitulated retinal cone cell regression as seen in human patients and has proven valuable in further studies of the disease.

Another example of transgenic pig models is the cystic fibrosis pig model. This model is overexpressing mutant forms of the CF transmembrane conductance regulator (CFTR) gene [109] and could be very important for the future studies of the pathophysiology of cystic fibrosis. Development of a transgenic pig model of diabetes has been attempted by Umeyama et al. [110]. The MODY3 diabetes pig model was generated by cloning of somatic cells transfected with a mutant form of human hepatocyte nuclear factor (HNF)-1alpha gene. This model recapitulates a variety of the pathophysiological characteristics of diabetes such as high blood glucose level, abnormal formation of Langerhans islets and poor insulin secretion.

To study the cardiovascular regulation of endothelial cell nitric oxide synthase (eNOS) in cardiovascular systems transgenic pigs overexpressing the eNOS gene have been created [111]. Vascular function, vascular structure and homeostasis are thought to be partially regulated by nitric oxide released by eNOS. The eNOS pigs were generated by nuclear transfer and may prove to be useful as models in the study of eNOS function in regulating muscle metabolism and the cardio respiratory system.

Using lentiviral mediated gene transfer transgenic pigs expressing glucose-dependent insulinotropic polypeptide receptor (GIPR) have been generated [112]. The purpose was to study the so-called incretin effect, the phenomenon that an oral glucose load elicits a higher insulin response than does the intravenous glucose load. This particular transgenic pig model may be used in the study of the roles of GIP/GIPR system and the importance in type 2 diabetes.

Production of transgenic mini-pigs (Göttingen) overexpressing the porcine huntingtin gene has also been constructed in order to mimic Huntington’s disease in humans [113].

Transgenic pigs with reporter genes have also been generated by several groups. Genetically engineered pigs expressing single or multiple fluorescent cell markers such as green (EGFP), blue (EBFP) and red fluorescent protein (DsRed) have been developed [102]. Such transgenic pigs are highly valuable in research requiring tracking of transplanted cells or tissues.

Another application of pigs in biomedical research is the development of genetically modified pigs for xenotransplantation research. The human organ donor shortage represents a persistent worldwide problem. The solution to this problem could be xenotransplantation, i.e. the use of non-human transplants into humans. Effective xenotransplantation is totally dependent on avoidance of organ rejection. In order to solve this problem it is desirable to humanize pig organs and this can likely be done by genetic modifications. As the first step to humanize pig organs transgenic knockout pigs defective in the galactosyltransferase gene have been established [114–116]. Using SCNT transgenic pigs with multiple genetic modifications, i.e. more than one gene knocked out, have been created [117, 118].

**Pig models for human neurodegenerative diseases**

The high resemblance between the central nervous systems of humans and pigs makes the pig an ideal model organism for studying human neurodegenerative diseases. For amyotrophic lateral sclerosis (ALS), PD and Alzheimer’s disease (AD), the pig may represent a model superior to the model presently available. Numerous models for ALS and PD have been established in a wide range of species such as *C. elegans* (nematode), zebrafish, fruit fly and in the rodents. However, these models do not recapitulate clinical and pathological characteristics of ALS and PD sufficiently. Large animals, including pigs and non-human primates in neuroscience, enable the use of conventional clinical brain imagers and the
direct use and testing of surgical procedures and equipment from the human clinic. The higher complexity of the large animal brain allows a more direct translation to human brain function in health and disease.

Large animal neuroscience can be used to complement information obtained from basic studies with small animals by constituting an intermediate research system, bridging small animal CNS research to the human CNS. Large animal models may, accordingly, be of high value in the field of translational neuroscience.

Our research group have isolated and characterized several pig homologues of human genes associated with human neurodegenerative diseases such as PD, AD, ALS and dystonia. This includes the molecular characterization of porcine amyloid precursor protein (APP) [119] and Presenilin1 and Presenilin2 [120] which both are genes whose products are essential components in AD. Also, we have cloned and characterized genes associated with familial forms of PD such as Dj-1 [121], Parkin/PARK2 [122] and LRRK2 [123] and TorsinA which is known to be associated with dystonia [124]. In addition, we have studied the porcine synuclein gene family alpha, beta and gamma-synucleins, in particular alpha-synuclein, which is one of the most prominent genes associated with familial Parkinsonism [125–127]. The porcine alpha-synuclein gene is currently used in combination with SMGT to create a pig model of PD. As the step to introduce lentivirus-mediated gene transfer to generate a mini-pig model of PD we have cloned and characterized porcine Synapsin I [128]. This protein is neuron-specifically expressed in humans and rats, and therefore we aimed to isolate the promoter of porcine Synapsin I. Using nuclear transfer and handmade cloning Kragh et al. [129] attempted to develop a porcine (Göttingen miniature pig) model of AD. The potential AD model is based on overexpression of a mutated version of the APP gene. Both mutant APP transcript and protein have been detected in the transgenic pigs, but clinical symptoms of AD are still awaited.

The current pharmacotherapy of PD and AD is symptomatic and does not prevent the progressive loss of dopaminergic neurons. The aim for the future is to develop neurorestorative therapies that could arrest the degenerative process or even better regenerate injured neurons.

**FUTURE DIRECTIONS—WHAT’S NEXT?**

The progressive advance of functional genomics and proteomics in pigs as discussed in this paper is of great importance for developing and describing porcine models used in biomedical sciences, because these technologies allow a more detailed understanding of the complex molecular mechanisms that control the biology and pathology of pig models. The term ‘systems biology’ is commonly used to address the need to integrate disciplines and technologies in order to characterize biological systems. In particular, the integration of genome, transcript and proteome data is currently relied on for extracting knowledge about complex biological systems, like the impact of a genetic variant, or the appearance of health and disease traits of humans and animals.

Although many shortcomings and challenges remain, before integration of cross-disciplinary data becomes streamlined and easily accessible, the current progress in bioinformatics and data integration technologies promises better data mining to come. Also the fast pace by which transcriptome and proteome methods are developing holds great promise that the current shortcomings can be overcome. This includes increasing the throughput of technologies, so that much larger sets of biological variation (i.e. populations) can be studied. For proteomics, a major issue is also increasing sensitivity, in order to include also low abundant proteins in the analyses.

With the current awareness of the usefulness of porcine biomedical models, the tools and technologies discussed in this review will be indispensable for extending our knowledge on porcine molecular mechanisms, and thereby fully exploit the benefits of using pig models for investigating human disease mechanisms.

**Key Points**

- The pig closely reflects genetics and physiology of man.
- Porcine models are useful in biomedical research.
- This review gives a state-of-the-art in porcine genomics, transcriptomics and proteomics.
- Examples of pig models established through transgenic technology are given.
References

9. Shaw DR. Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. Curr Protoc Bioinformatics 2009 Chapter 1:Unit1.


