

The draft genome of the Wisconsin Miniature Swine™, a valuable biomedical research tool

Alex C. Veith ^{1,2} Jennifer J. Meudt,³ Jamie L. Reichert ⁴ Jennifer M. Frank,^{4,6} Derek M. Pavelec,¹ Bridget Ladell,¹ James Speers,¹ Molly Zeller ¹ Taeyoung Shin,^{5,6} Joshua R. Hyman,¹ Christopher A. Bradfield,^{1,2} Charles M. Konsitzke,¹ Dhanansayan Shanmuganayagam ^{3,5,*†} C. Dustin Rubinstein ^{1,*†} Mark E. Berres ^{1,*†}

¹University of Wisconsin Biotechnology Center, University of Wisconsin-Madison, Madison, WI 53706, USA

²McArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, USA

³Biomedical and Genomic Research Group, Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

⁴Swine Research and Teaching Center, Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

⁵Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, USA

⁶Present address: Center for Biomedical Swine Research and Innovation, University of Wisconsin-Madison, Madison, WI 53706, USA

*Corresponding author: Department of Animal and Dairy Sciences, University of Wisconsin-Madison, 1675 Observatory Drive, Madison, WI 53706, USA. Email: dshanmug@wisc.edu; *Corresponding author: University of Wisconsin Biotechnology Center, University of Wisconsin-Madison, 425 Henry Mall Madison, WI 53706, USA. Email: crubinstein@wisc.edu; *Corresponding author: University of Wisconsin Biotechnology Center, University of Wisconsin-Madison, 425 Henry Mall Madison, WI 53706, USA. Email: meberres@wisc.edu

†These authors have contributed equally to this work and share senior authorship.

Porcine biomedical models have emerged as valuable tools in biomedical research due to their physiological, anatomical, metabolic, immunological, and genetic similarities to humans. As a result, they offer greater relevance for translational studies than rodent models. Moreover, compared to nonhuman primates, porcine models are more cost-effective, easier to manipulate genetically, and raise fewer ethical concerns. However, the conventional breeds of swine most commonly used in research have rapid growth rates, which lead to logistical challenges such as increased space requirements, making them impractical as biomedical models. The Wisconsin Miniature Swine™ (WMS™) was developed to address these shortcomings. The WMS™ porcine model grows slower, reaching and maintaining human sizes at adulthood. The model was also specifically designed to possess more human-like physiology that allows for easy modeling of comorbidities like obesity and metabolic syndrome that affect a large portion of the human population affected by chronic diseases. Thus, WMS™ is an ideal porcine gene editing platform for modeling complex multifactorial diseases. Here, we present the first draft genome assembly representative of the WMS™ line. The primary assembly was generated with ~20x coverage of long reads from Oxford Nanopore Technologies and independently error-corrected using 23x Pacific Biosciences reads. Arima Genomics Hi-C data were used to improve contiguity. Largely congruent with the existing *Sus scrofa* genome, we also show the utility of WMS™ as a model through comparisons between 2 WMS™ genes and human homologs. Finally, we show the utility of genotyping by sequencing across WMS™ herds. The WMS™ genome generated here is highly complete and supports investigators utilizing WMS™ in biomedical research.

Keywords: *Sus Scrofa*; Wisconsin Miniature Swine™; WMS™; genome; NF1; Genome Assembly

Introduction

Porcine models are becoming indispensable in biomedical research due to their remarkable anatomical, physiological, and genomic similarities to humans (Schomberg et al. 2016; Lunney et al. 2021; Meyerholz et al. 2024). These similarities are particularly pronounced in key systems such as the cardiovascular, gastrointestinal, and immune systems, making pigs highly translational models for studying human diseases and developing novel therapies (Meurens et al. 2012; Lelovas et al. 2014; Gerdtts et al. 2015; Pabst 2020). Beyond systemic similarities, pigs also provide an optimal platform for cutting-edge imaging modalities due to their size and physiology, which closely mimic those of humans. Advanced imaging techniques, such as MRI and CT, have been effectively adapted and validated in porcine models, enabling high-resolution studies that bridge preclinical

findings to clinical applications (Raval et al. 2022; Costine-Bartell et al. 2023). Moreover, swine models are increasingly being used to study aging and associated comorbidities, providing new opportunities to explore therapeutic strategies for age-related conditions (Schachtschneider et al. 2021; Lu et al. 2023). Another transformative area of research where pigs have gained prominence is xenotransplantation. The anatomical and physiological compatibility of porcine organs with human systems, combined with advances in genetic engineering to reduce immunological rejection, have positioned swine as the leading source for xenogeneic organ transplantation (Anand et al. 2023; Ali et al. 2024; Peterson et al. 2024). The recent successful transplantation of porcine kidneys and hearts into humans underscores the critical role of swine in addressing the global organ shortage crisis (Griffith et al. 2022; Porrett et al. 2022; Moazami et al. 2023).

Received on 24 January 2025; accepted on 18 March 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Porcine models complement and effectively address limitations of other animal models, such as rodents and nonhuman primates (NHP), while offering unique advantages for biomedical research. Despite their high translational relevance, NHP models are constrained by limited availability, high costs, and shifting public perception regarding their use in research (Carvalho et al. 2018; Grimm 2023). Swine, in contrast, are more accessible, less expensive to maintain, and face fewer ethical barriers, making them a practical and scalable alternative for many research applications. The physiological and anatomical similarities between swine and humans provide a level of translational fidelity that rodent models often cannot achieve. Rodents, while valuable for genetic studies and early stage research, have shown high failure rates in translating preclinical findings to human clinical trials, particularly in areas such as cardiovascular, immune, and metabolic diseases (Gutierrez et al. 2015; Frangogiannis 2022).

The advent of CRISPR-Cas9 genome editing has significantly enhanced the utility of porcine models, allowing researchers to create precise genetic modifications that replicate human disease states. This capability, once a hallmark advantage of murine models, now positions swine as a powerful tool for studying genetic disorders, drug development, and regenerative medicine (Perleberg et al. 2018). To this end, we recently generated several porcine models of neurofibromatosis type 1 (NF1), a phenotypically complex genetic disorder in humans (Rubinstein et al. 2021). The NF1^{a31/+} swine model formed neurofibromas similar to those seen in human patients and provided novel insight into the fibroblast populations and shed light on a potential role for cancer-associated fibroblasts on the recruitment of Schwann cells to produce neurofibromas (McLean et al. 2023). Here, porcine models possess a clear advantage over NHP, which, despite their translational competence, are impractical for the generation of genetic models by genome editing due to their long reproductive cycles and limited number of offspring. Rhesus macaques, the most used NHP in the United States, typically reach sexual maturity at 3–5 years of age and give birth to 1 offspring per year after a gestation time of ~165 days (Wolfensohn 2010; Campbell et al. 2011; Cooper et al. 2022; National Academies of Sciences 2023). In contrast, swine reach sexual maturity at 5–6 months of age and can give birth to 1–2 litters of 8–12+ piglets/litter per year with a gestation time of ~114 days; many minipig breeds give birth to 4–8 piglets/litter.

Despite the translational advantages of the species, conventional breeds of swine most commonly used in research exhibit rapid growth, reaching weights of 100 kg (220 lb.) by 4 months of age and continuing to grow to 249–306 kg (550–675 lb.) at full maturity (Schomberg et al. 2016). This rapid growth necessitates the use of prepubertal pigs in research to maintain manageable sizes. However, using juvenile pigs does not accurately model diseases prevalent in adult human populations, as their physiological and metabolic states differ significantly from adults. Furthermore, the accelerated growth rate of conventional swine (5–6 kg increase/week) is associated with rapid tissue remodeling, which does not accurately reflect the slower tissue turnover observed in adult humans. This discrepancy can complicate the translatability of research findings, particularly in studies involving tissue regeneration, wound healing, and chronic disease progression. Thus, miniature (“minipig”) breeds have received greater attention recently (Gutierrez et al. 2015). These breeds better grow slower and match human size at maturity while also reducing required husbandry resources. They also exhibit a greater physiological similarity to humans than conventional breeds. For example, higher similarity between minipig liver enzymes and those of

humans, indicating a greater translational value in pharmacokinetic studies, has been noted (Suenderhauf and Parrott 2013).

The Wisconsin Miniature SwineTM (WMSTM) was developed at the University of Wisconsin-Madison to provide the many advantages of minipigs, including age-related changes in genes associated with obesity and lipid metabolism comparable to humans (Fig. 1). The WMSTM breed reaches puberty at about 5–6 months of age with a body weight of ~27–32 kg (59–70 lb.) and can be maintained at human sizes [68–91 kg (150–200 lb.)] during adulthood (Schomberg et al. 2016). In contrast, while conventional breeds reach puberty at the same age, they typically weigh 110–135 kg (242–297 lb.). The WMSTM is derived from the once widely used Rapacz-Familial Hypercholesteremia (FH) mini swine, also developed at the university, which carried a clinically relevant mutation in the low-density lipoprotein receptor (LDL-R) associated with high circulating blood cholesterol (Rapacz et al. 1994; Hasler-Rapacz et al. 1998). The Rapacz-FH line, a hybrid of 37 pig breeds throughout Europe and North America, was bred to Guinea Hogs for 1 generation and interbred twice to produce unique WMSTM-WT (wildtype) and WMSTM-Familial Hypercholesteremia (WMSTM-FH) lines. WMSTM-WT was also previously published as WMSTM-N (normal) (Namous, Krueger, et al. 2023; Namous, Strillacci, et al. 2023). The WMSTM-FH line harbors the LDL-R mutation, while the WMSTM-WT does not (Schomberg et al. 2016). Thus, WMSTM-FH develops human-like complex atherosclerotic lesions in coronary arteries and other arterial beds that display necrotic core, calcification, and neovascularization (Namous, Strillacci, et al. 2023). Moreover, we have demonstrated that, like humans, the model develops neointimal restenosis in coronary arteries after stent implantation, unlike conventional swine breeds (Schomberg et al. 2016). Both lines have been invaluable in biomedical research (Lin et al. 2018; Miranpuri et al. 2018; Horton et al. 2020; Johnson et al. 2021; Schachtschneider et al. 2021; Wallat et al. 2021; Wuschner et al. 2021, 2022a, 2022b; Eix et al. 2022; Hanna et al. 2022; Johnson et al. 2022; Pirasteh et al. 2022; Lu et al. 2023; Flakus et al. 2023a, 2023b, 2023c; Namous, Krueger, et al. 2023; Namous, Strillacci, et al. 2023; De La Cruz et al. 2024). Here, we present the draft genome of the WMSTM to establish the genomic toolkit needed to broaden the utility of these lines and accelerate biomedical advancements using porcine models.

Materials and methods

Herd maintenance

The animal protocols related to this work were approved by the Institutional Animal Care and Use Committee at the University of Wisconsin-Madison in accordance with the *Guide for the Care and Use of Laboratory Animals* and the Animal Welfare Act. The WMSTM-WT and WMSTM-FH lines are housed at the University of Wisconsin Swine Research and Teaching Center (Arlington, WI), a specific pathogen-free facility. Both lines receive the same diet, formulated with essential macronutrients and micronutrients to meet their requirements for normal growth and development.

Sample preparation

A whole blood sample was collected from a representative adult male WMSTM-WT pig (“Benny,” Pig ID: 18-4274-3; Age: 1 yr, 6 months, 25 d) for the purpose of genome assembly. The boar used in this study was frequently used for breeding in the WMSTM-WT herd. Additional blood samples were obtained from both the WMSTM-WT and WMSTM-FH lines to conduct genotyping by sequencing (GBS) analysis. These collected samples were submitted to the University of Wisconsin-Madison Biotechnology

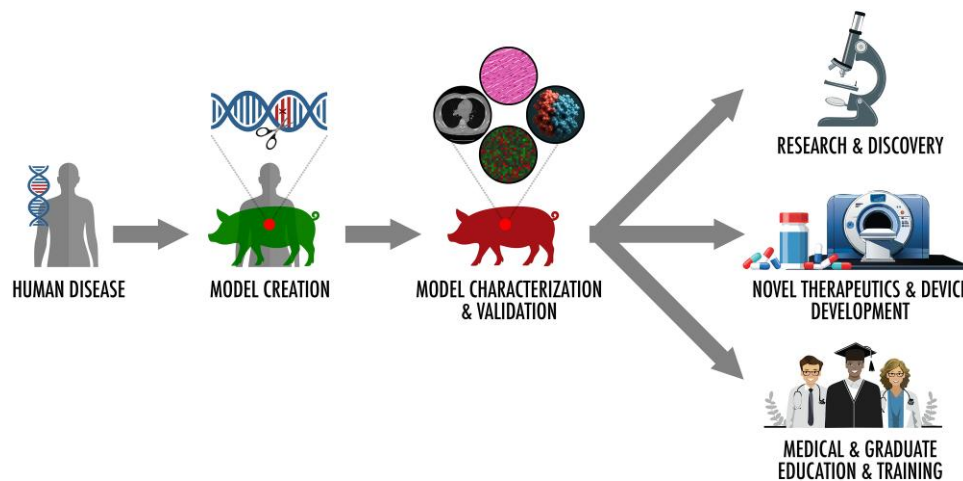


Fig. 1. The WMSTM, developed at UW-Madison, is an important translational platform for creation of genetic models. It has unique value in preclinical and translational research and as an education and training tool.

Center DNA Sequencing Facility for processing. High molecular weight (HMW) DNA (25 µg) was extracted from leukocytes using the Circulomics Nanobind CBB Big DNA kit (Pacific Biosciences, Menlo Park, CA, USA). Nucleic acid concentrations were quantified with the Quant-iTTM PicoGreenTM dsDNA and QubitTM dsDNA High Sensitivity kits (ThermoFisher Scientific, Waltham, MA, USA) used for Oxford Nanopore and Pacific biosciences (PacBio) runs, respectively. The concentrations (A260/A280) and purity (A260/A230) of DNA were determined using a NanoDropTM One spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Finally, the sizing and quality of diluted samples were assessed using the FemtoPulse System (Agilent Technologies, Madison, WI, USA). The average read size, between 40,000 and 508,400 bp, was 244,172 (bp).

Whole genome sequencing

Whole genome sequencing was performed using 2 third-generation sequencing platforms and a genome-wide proximity ligation (Hi-C) method, resulting in an error-corrected, chromosome-scale assembly of the WMSTM genome.

Oxford nanopore technologies sequencing

Oxford nanopore technologies (ONT) sequencing was utilized to generate the initial reads used for assembling the WMSTM draft genome. Sequencing libraries were prepared from 3 µg of input DNA using the SQK-LSK109 ligation sequencing kit (Oxford Nanopore Technologies, Oxford, UK) and quantified with the QubitTM dsDNA High Sensitivity Kit (ThermoFisher Scientific). The sequencing was performed on 5 MinION flow cells processed on a GridION instrument (Oxford Nanopore Technologies).

PacBio sequencing

PacBio sequencing was utilized to obtain highly accurate reads to error correct (“polish”) the assembly produced by the ONT reads. A PacBio HiFi SMRTBell library was prepared from 5 µg of input DNA according to PN 101-693-800 Version 01 (Pacific Biosciences, Menlo Park, CA, USA), with the following modifications: (1) shearing with gTUBES (Covaris, LLC, Woburn, MA, USA) and (2) size selection with BluePippin (Sage Sciences, Beverly, MA, USA). The quality of the library was determined by the FemtoPulse System (Agilent Technologies, Madison, WI, USA), and the QubitTM dsDNA High Sensitivity Kit (ThermoFisher Scientific, Waltham,

MA, USA) was used to quantify the library preparation. PacBio sequencing was performed on a Sequel I instrument (Pacific Biosciences, Menlo Park, CA, USA) with 4 PacBio 4 M SMRT cells.

Arima Hi-C sequencing

Hi-C sequencing was utilized to scaffold the error-corrected ONT draft genome assembly. Sample processing, crosslinking, and Hi-C library preparation were performed according to the manufacturer’s instructions for the Arima HiC v1 kit for Mammalian Blood (Material Part Number: A410030, Document Part Number: A160104 v01; Arima Genomics, Carlsbad, CA, USA) and the KAPA Hyper Prep Kit (Roche, Basel, CH) with the following modifications: during PBMC isolation, phosphate-buffered saline (PBS) was used for the first step as directed but subsequent steps used fetal bovine serum (FBS) and adapter ligation was increased from 15 min to 1 h. Library quality was assessed with an Agilent TapeStation (Agilent Technologies) and quantified with a QubitTM dsDNA High Sensitivity (ThermoFisher Scientific) kit. The resulting Illumina TruSeq library was sequenced as paired-end (2 × 150 bp) reads on a NovaSeq6000 (Illumina, Inc., San Diego, CA, USA).

Genome assembly

ONT reads meeting a minimum Q-value of 10 (phred scale) were size- and chimera-filtered (minimum length = 10,000 bp) to approximately 23× coverage with Filtlong v0.2.0 (<https://github.com/rrwick/Filtlong>) and Porechop v0.2.4 (<https://github.com/rrwick/Porechop>). Filtered reads were assembled with Flye v2.5, enforcing a 10,000 bp overlap minimum (Kolmogorov et al. 2019). Assembled sequences <10,000 bp were filtered from the draft assembly. Approximately 20× PacBio subread coverage was used with GCp v1.0.0-1807624 (<https://github.com/PacificBiosciences/gcpp>) to independently error-correct the ONT-based Flye assembly. The resulting error-corrected haploid (pseudohaploid) assembly was deduplicated (i.e. identification of syntenic elements, keeping only 1 element) with purge_haplotigs commit be7e44b (https://bitbucket.org/mroachawri/purge_haplotigs/src/master/) (Roach et al. 2018). Juicer v1.6.2 (<https://github.com/aidenlab/juicer>) and 3D-DNA v180114 (<https://github.com/aidenlab/3d-dna>) were used to scaffold the deduplicated, error-corrected assembly with the Hi-C data (Durand et al. 2016; Dudchenko et al. 2017). We assessed the quality of our genome assembly using BUSCO v5.7.0 with mammalia_odb10 (<https://busco.ezlab.org/>) and QUASTv5.0.2

(<https://quast.sourceforge.net/>) (Gurevich et al. 2013; Manni et al. 2021).

Gene identification

Putative genes in the WMSTM draft genome were identified using Liftoff v1.6.3 (<https://github.com/agshumate/Liftoff>) with default parameters against the reference *S. scrofa* Ensembl annotation for Sscrofa11.1 (The Swine Genome Sequencing Consortium, 2017, Genome assembly Sscrofa11.1, NCBI, GCF_000003025.6), genebuild 2022-02 (Shumate and Salzberg 2021).

Gene diversity analysis

The draft WMSTM genome and annotation files produced here were interrogated for the *NF1* (neurofibromin) and *AHR* (aryl hydrocarbon receptor) genes. For each putative WMSTM gene, we extracted the entire annotated gene sequence and aligned it to the *S. scrofa* reference coding sequence (CDS) using the align sequences feature in SnapGeneTM Version 7.2.1 (GSL Biotech, Boston, MA, USA). The reference sequence was selected based on Ensembl's canonical distinction. The aligned CDS from the WMSTM gene were translated and aligned to the respective translated CDS in *S. scrofa* as well as the canonical human CDS using SnapGeneTM software. The BLOSUM62 similarity matrix was used to delineate similar amino acids. The WMSTM and human amino acid sequences were then input into AlphaFold 3 for structural predictions using the same seed sequence (Abramson et al. 2024). The structural predictions from WMSTM and human proteins were aligned using PyMOL (The PyMOL Molecular Graphics System, Version 3.0.5 Schrödinger, LLC.)

Genotype by sequencing

Genetic variation across the WMSTM-WT and WMSTM-FH herds was assessed using genotype-by-sequencing (GBS), a method licensed to the UW-Madison Biotechnology Center by KeyGene N.V. (Wageningen, NL). GBS libraries were prepared as previously described with minimal modification (Elshire et al. 2011). Whole blood was collected from individuals across both herds, and DNA was isolated with a QIAamp 96 DNA QIAcube HT Kit. One hundred-fifty nanograms of purified DNA was then digested by the restriction enzymes *NsiI* and *MspI*. Barcoded adaptors compatible with Illumina sequencers were added to restricted-DNA by ligation using T4 ligase (New England Biolabs, Ipswich, Mass.). Three plates, each containing up to 96 adapter-ligated samples, were each pooled and amplified to generate a sufficient sequencing template. Adapter dimers were removed by SPRI bead purification. The quality and quantity of the finished libraries were assessed using the Agilent TapeStation (Agilent Technologies, Inc., Santa Clara, Calif.) and QubitTM dsDNA HS Assay Kit (Life Technologies, Grand Island, N.Y.), respectively. Libraries were sequenced on a NovaSeq6000 S2 platform (Illumina, Inc., San Diego).

The University of Wisconsin Bioinformatics Resource Center (UWBRCC) completed quality control, sequence alignment, and SNP calling on GBS data. Adapters, low-quality bases, and primers were trimmed from reads to obtain a Phred score of 20 via the trimming software Skewer (v0.2.2.b). Reads that were too short to be used were also discarded. (Jiang et al. 2014). The Tassel pipeline (version 2) was used to process sequence reads to single nucleotide polymorphism (SNP) genotypes (Glaubitz et al. 2014). Briefly, demultiplexed reads were aligned to the WMSTM draft genome using Bowtie2 (Langmead and Salzberg 2012). The Tassel pipeline was then used to identify potential tags in the aligned SAM files compared to the WMSTM reference genome, outputting a VCF file. Filtering was performed on VCF files to include only biallelic sites and a minor

allele frequency (MAF) ≥ 0.05 . SNP coverage and coverage per sample were then calculated. The population-wide inbreeding coefficient *F* was calculated with VCFtools (v.0.1.16) (Danecek et al. 2011).

Sample relationships were assessed using principal component analysis (PCA), and a bifurcating topology was generated using maximum likelihood. Briefly, for PCA analysis, SNPs were filtered by those with information at $\geq 60\%$ of samples, and samples with $\leq 80\%$ of pre-filtered single nucleotide polymorphisms (SNPs) were excluded. Beagle v4.1 was used to perform whole genome phasing and to impute missing data (Browning and Browning 2007, 2016). A mid-point rooted topology was created from the VCF data using RAxML and visualized with itol v6.9.1 (<https://itol.embl.de/>) (Stamatakis 2014; Huerta-Cepas et al. 2016; Ortiz 2019; Letunic and Bork 2024).

Results and discussion

Genome assembly

We anticipated generating a WMSTM genome assembly of approximately 2.5 Gb based on its expected similarity to *S. scrofa* (Sscrofa11.1). The reference genome assembly of Sscrofa11.1 was completed by adding Y chromosome sequences from other sources (GCA_900119615.2) because TJ Tabasco (Duroc 2–14) was female. The resulting reference genome sequence was termed Sscrofa11.1 and deposited in the public sequence databases (GCA_000003025.6), which is the reference used in this work (Warr et al. 2020). Oxford Nanopore (ONT) sequencing was chosen for the initial pass of the genome assembly due to its ability to generate extremely long reads and produce highly contiguous data (Lu et al. 2016). Sequencing data generated from 5 ONT MinION flow cells produced an input read length of approximately 48 Gb, achieving an expected coverage of about 20X. Using the Flye assembler, we generated a draft WMSTM assembly consisting of 1,215 contigs exhibiting a mean coverage of 17X.

One characteristic of the ONT sequencing platform is a relatively high error rate, which necessitates additional error correction (Delahaye and Nicolas 2021; Wang et al. 2021). While overlap layout consensus (OLC) tools are available to error-correct assemblies using the same reads input into the assembly, we opted to perform error correction independently with PacBio reads (Chen et al. 2021). We obtained a total read length input of 58 Gb from 4 SMRT cells (approximately 23 $\times$ coverage). Following error correction, the number of contigs was reduced by 16.6% from 1,215 to 1,013. The L50 of the error-corrected assembly was 22 contigs exhibiting an N50 of 36.751 MB (max contig length = 125.705 MB).

Regions of high heterozygosity can lead Flye, as well as other genome assemblers, to produce assemblies that exceed the expected haploid genome size. When allelic sequences reach a threshold level of nucleotide divergence, standard assembly algorithms may treat these regions as distinct sequences, generating separate contigs instead of a single, merged haplotype contig. This phenomenon reflects a limitation in the assembler's ability to reconcile high-divergence alleles within a unified assembly, which may lead to an over-representation of allelic regions in the final assembly output. We used Purge Haplotigs to mitigate this over-representation by identifying and removing redundant contigs derived from allelic variation, thereby refining the assembly to better reflect the true haploid genome structure better. Purge Haplotigs typically generates 2 distinct peaks in the k-mer distribution plot, corresponding to homozygous and heterozygous regions in assemblies containing syntenic haplotigs. However, when applied to the WMSTM draft genome, we observed only a single peak with no discernible shoulder peaks. This outcome may

reflect biologically low heterozygosity (i.e. identical alleles at most loci) or insufficient and/or uneven sequencing coverage, which could obscure the k-mer differences between homozygous and heterozygous regions. The software aggressively removed even large contigs when we applied Purge Haplotigs using permissive cutoffs. Given the existing low contig count, the N50 metric capturing the 20 expected chromosomes (MUMmer4 analysis), and the risk of false positives, we opted not to apply Purge Haplotigs to the draft assembly.

Hi-C data was incorporated to enhance the contiguity of the genome assembly, following the approach described elsewhere (Burton et al. 2013). We generated approximately 400 M 2 × 150 bp reads using Illumina sequencing. Systematic integration of the Hi-C data yielded a marked improvement in contiguity. However, comparison with the Sscrofa11.1 reference genome using MUMmer4 often revealed structural inconsistencies, including large- and small-scale mis-joins, sub-chromosomal inversions, and inter-chromosomal translocations (Marcais et al. 2018). These issues were likely exacerbated by regions of lower sequence coverage, particularly evident on the X and Y chromosomes, which had an average coverage of approximately 10x. While we expect some inherent structural variation in the WMSTM genome, the overall synteny is consistent with Sscrofa11.1, comprising 18 autosomes and the X and Y sex chromosomes. To address potential misassemblies, larger contigs generated with Hi-C data were manually reviewed. This involved aligning the original Oxford Nanopore reads to each contig for verification. Subsequently, CRAQ v1.0.9 (<https://github.com/JiaoLaboratory/CRAQ>) was used to detect large-scale structural errors, termed clip-based structural errors (CSEs) (Li et al. 2023).

Contigs were split at locations where significant clipping in the Oxford Nanopore reads indicated the presence of CSEs. In addition to errors introduced by the assembler, these structural inaccuracies may also result from incomplete DNA digestion during Hi-C preparation, erroneous ligation of DNA fragments, or misassignment of long-range contacts during scaffolding, which can misrepresent true genomic architecture. After manual inspection, JupiterPlot v1.1 (<https://github.com/JustinChu/JupiterPlot>) was used to create a syntenic comparison of WMSTM to the Sscrofa11.1 reference genome (Fig. 2). The final draft assembly was considered complete following manual curation, with the assembly statistics summarized in Table 1.

To evaluate the quality of our assembly, we performed a BUSCO assessment of conserved single-copy orthologous sequences. The WMSTM draft assembly achieved a BUSCO score of 98.2% ([S:97.9%, D:0.3%], [F:1.0%, M:0.8%]), indicating that we have a highly complete assembly. Notably, the detection of only 27 duplicated BUSCOs out of 9,226 indicates a low prevalence of haplotigs, supporting the decision that haplotig purging was not strictly necessary. It remains possible that false collapses affected the number of identified duplications. The BUSCO assessment of the assembly generated here surpasses the BUSCO scores for existing *S. scrofa* reference genomes (Sscrofa11.1 = 93.8% and USMARCv1.0 = 93.1%) (Warr et al. 2020). Overall, our data indicates that the draft WMSTM genome is a highly complete genome assembly.

At the time of our analysis, Porechop was the only tool available that provided comprehensive functionality for both demultiplexing and adapter trimming. While newer tools such as Guppy (Guppy_barcode) have since improved and now supersede Porechop, they were still under development during the initial stages of our study. Specifically, Porechop outperformed early versions of Guppy in terms of accuracy and reliability for demultiplexing and trimming, as evidenced by our own internal benchmarks at the

time. The same is true for sequencing error rates, which have been reduced by refinements in chemistry and algorithms. Our analysis focused on ensuring consistency and reliability with the tools available at the time the study was conducted. While it is true that reanalyzing data with updated tools can provide refinements, our approach ensures the reproducibility of our methods as initially performed. Moreover, the tools used were state-of-the-art for their respective timeframes, reflecting the quality of the data generated and the robustness of our comparisons.

Annotation

The WMSTM line was created from a hybrid of approximately 37 *S. scrofa* breeds, predominately Guinea Hog (*Sus scrofa domesticus*), and therefore, we expected high identity with the reference *S. scrofa* genome. To analyze this, we employed LiftOff, an annotation lift-over tool that maps gene features from closely related species onto a target genome (Shumate and Salzberg 2021). When applied to WMSTM, the lift-over annotation resulted in a 97.76% match with Sscrofa11.1 annotations with an alignment coverage and sequence identity ≥50%. The gene features in the *S. scrofa* genome used in our analysis included 20,306 annotations, of which we identified 19,853 matches (96.4%). This number reflects similar gene counts identified in other swine models and indicates high concordance between the gene features of WMSTM and *S. scrofa*. Notably, 93% of the 453 genes that were not mapped by the LiftOff analysis have no assigned function. Our study has demonstrated that LiftOff provides robust and accurate annotations, meeting the objectives of this research by providing a high-quality functional annotation.

Comparison to *S. scrofa*

The WMSTM assembly generated here exhibited a strong identity with the Sscrofa11.1 genome build (Fig. 2). The cross-genome analysis identified several small structural variations between the WMSTM and *S. scrofa* genomes. There was also a prominent inversion on chromosome 9 that persisted throughout our analysis. This inversion was present with and without Hi-C assembly information, and our data strongly suggests that this is a *bona fide* difference from Sscrofa11.1. It is likely that structural variations exist due to both assembly errors and true genetic variation. Additional work will be needed to validate the WMSTM genome, including potential structural variations (Yang et al. 2024).

Gene-level comparison

To illustrate the utility of WMSTM as a model for human disease or discovery, we compared sequence variation between WMSTM, *S. scrofa*, and human NF1 and AHR genes. One active area of interest to our group is the use of swine models in NF1 (Rubinstein et al. 2021; McLean et al. 2023). In our previous research, we utilized larger swine breeds, which limited the number of animals we could maintain. To demonstrate the utility of the WMSTM as a biomedical research platform, we compared the gene annotations of the NF1 and Ahr genes WMSTM to the *S. scrofa* and human reference transcripts to determine their similarity.

The annotation lift-over identified a ~265kb-long putative NF1 gene in the WMSTM genome that matched the *S. scrofa* NF1 gene (ENSSSCG00000017748). The Ensembl canonical CDS for the NF1 gene CDS (ENSSSCT000000093848.2) produces a 2,839-long amino acid sequence when translated. The ENSSSCT000000093848.2 was successfully aligned to the putative WMSTM NF1 gene to produce the WMSTM NF1 CDS, encoding a 2,840-long amino acid. Local alignment of the amino acid sequences showed a sequence identity match of 99.96%, differing by a single amino acid insertion (Pro771_Ala772insThr) in the WMSTM NF1 sequence.

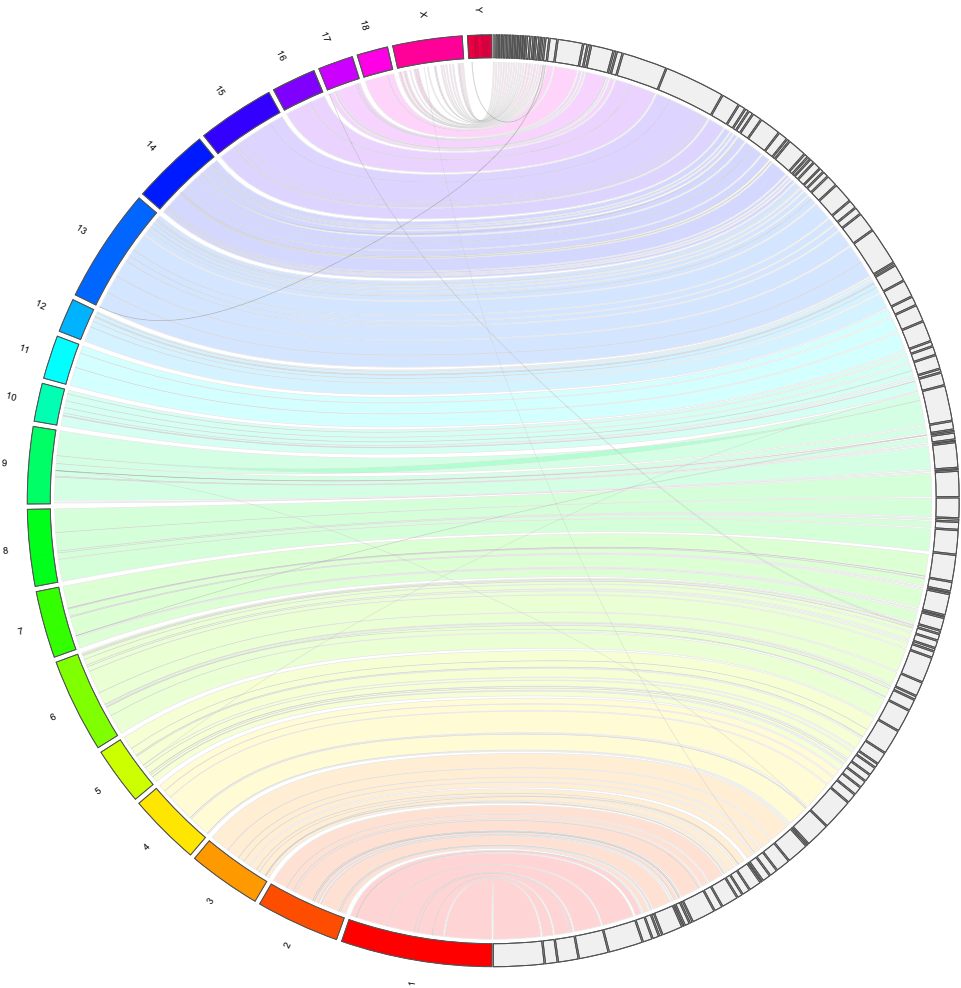


Fig. 2. Visualization of WMSTM and *S. scrofa* syntentic identity. The *S. scrofa* scaffold is represented on the left hemisphere compared to the WMSTM scaffold, which is depicted on the right hemisphere. The Circos-based plot links the reference chromosomes to the scaffolds generated in the WMSTM assembly, with each chromosome uniquely shaded. We see strong similarity between the 2 genomes but do note structural differences.

Table 1. Final assembly information.

Total length (bp)	2,517,610,705
Total length > 50,000 bp	2,511,032,328
# Contigs	1,022
# Contigs > 50,000 bp	729
Largest contig (bp)	108,923,112
N50 (bp)/L50	35,546,377/25
N90 (bp)/L90	2,592,705/102
auN (bp)	36,668,315.6
GC	42.15%
# N's per 100 kbp	0.19

Comparison of the WMSTM NF1 amino acid sequence to the human homolog (ENST00000358273.9) revealed a 99.12% (2817/2842) identity match and an additional 10 positions in which they shared similar amino acids (99.47% similarity match). Structural prediction and alignments show a strong similarity, reflected by a root mean square deviation of 2.114 Å (Fig. 3a). The potential structural differences tended to be localized to areas that are predicted with low confidence. This was expected, as NF1 has been implicated in regulating tumorigenesis, and others have demonstrated strong sequence similarities between human and murine homologs (Bernards et al. 1993; Jouhilahti et al. 2011).

In contrast, the AHR is a ligand-dependent transcription factor that is highly conserved yet exhibits significant sequence variation

between species (Hahn 2002; Hahn et al. 2017). We have recently described a “xenokine” model in which AHR ligands regulate physiology (Avilla et al. 2020). Swine have been used to isolate endogenous AHR ligands, and we hypothesize that WMSTM may be a useful model to study the role of AHR in physiology (Song et al. 2002). The annotation lift-over identified a ~45 kb-long putative AHR gene in the WMSTM genome that corresponds with the *S. scrofa* AHR gene (ENSSSCG00000030484). Alignment of the *S. scrofa* canonical AHR CDS (ENSSSCT00000031031.4) to the WMSTM genome produced a CDS encoding an 844-long AHR amino acid that shared a 98.35% identity match to the *S. scrofa* AHR amino acid sequence. This level of variability is consistent with previous findings between *Ahr* alleles between different lines of mice (Poland et al. 1994). To this end, the WMSTM AHR amino acid sequence shared an 81.66% identity match to the human AHR (699/856 positions), with an additional 61 positions sharing similar amino acids (88.79%). The predicted structures of both the WMSTM and human AHR proteins were aligned, as shown in Fig. 3b.

Porcine models hold substantial translational potential due to their remarkable physiological and anatomical similarities to humans, coupled with their susceptibility to precise genetic modifications (Lunney et al. 2021; Rubinstein et al. 2021). The high-quality draft WMSTM genome presented here is a valuable resource for comparative genomic analyses and the informed design of targeted

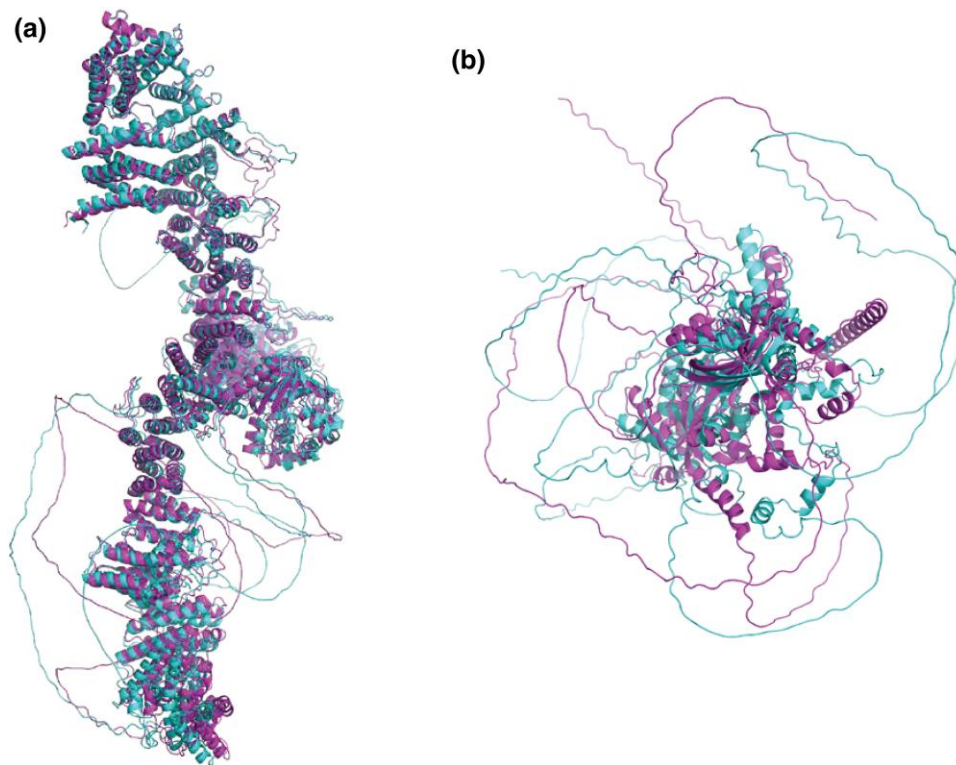


Fig. 3. Structural comparison of WMSTM and human NF1 and AHR proteins. a) Alignment of the predicted structures of the WMSTM (magenta) and human (cyan) NF1 proteins. We see strong structural similarity between both proteins, strongly supporting porcine models, and particularly WMSTM, as a platform for NF1 research. b) Alignment of the predicted AHR structures of WMSTM (magenta) and human (cyan). 360-degree movie clips are included in the supplemental files. Relative sizes of NF1 and AHR proteins are not depicted to scale.

mutations for modeling human diseases. Our analysis demonstrates that the WMSTM NF1 allele exhibits significant sequence and likely functional homology to the human ortholog. This finding emphasizes the utility of the WMSTM genome as a robust platform for studying NF1-related pathologies, allowing greater access to the molecular and phenotypic mechanisms underlying NF1 disorders in a physiologically relevant model system. Moreover, the availability of this genomic resource is anticipated to facilitate broader comparative genomic studies across species and accelerate the development of innovative therapeutic strategies.

Genotyping by sequencing

GBS is a cost-effective and high-quality approach commonly used in agricultural settings to assess genetic variation at the population level (Wang et al. 2020). GBS is a reduced-representation sequencing approach, leveraging restriction-enzyme accessible, variant-harboring targets to define population structure and conduct genome-wide association analyses (Elshire et al. 2011; Scheben et al. 2017; Hess et al. 2020).

For our GBS analysis, we collected DNA from 57 individual animals in the WMSTM herd and 113 individual animals in the WMSTM-FH herds, representing the entire herd population at that time (170 total). We generated an average of 2.5 M demultiplexed reads per sample, which provided 186,903 total SNPs, with 92,510 meeting our filtering criteria. Ordination analysis of biallelic SNPs revealed robust segregation of unique subpopulations, with the first 3 principal components explaining 65% of the variance in the dataset. Notably, a single principal component clearly distinguished between WMSTM-WT and WMSTM-FH samples (Fig. 4a). This result was expected, as these herds were selectively bred to retain the FH allele and were maintained separately, reflecting their genetic

divergence. A PCA biplot of all samples further highlighted the primary differences between the WMSTM-WT and WMSTM-FH herds. To explore potential subpopulations within the WMSTM-FH herd, we reanalyzed this group independently using PCA in conjunction with a k-means silhouette procedure to define groups (Fig. 4b). Without the confounding effect of FH allelic discrimination, this analysis provided finer resolution and revealed distinct subpopulations within the WMSTM-FH herd. Lastly, we constructed a bifurcating topology with a maximum likelihood approach to detail the relationships among all samples (Fig. 4c).

The FIS (F) of 170 individuals based on filtered GBS data was 0.38. At this level, it highlights a substantial deficit in heterozygosity, indicating moderate to high inbreeding within the population. In other words, there is a significant deficit of heterozygous individuals in the WMSTM herd compared to what is expected under Hardy-Weinberg equilibrium. This suggests that individuals are more genetically similar than expected, which could be due to (1) Inbreeding and/or (2) population substructure (Wahlund effect; a mixed population made up of subpopulations can show reduced heterozygosity when treated as one). Our data indicate that both issues are likely to be present in the WMSTM herd. FIS in domestic swine, especially commercial lines, is likely to be much lower, in the range of 0.1 to 0.3 (Tang et al. 2013). We anticipate that inbred populations used in research or specialized breeding programs exhibit higher F values (>0.30) (Gomez-Raya et al. 2015).

The application of GBS in managing the WMSTM herd offers 2 significant advantages. First, as a newly established line, the WMSTM herds (WMSTM-WT and WMSTM-FH) likely exhibit some magnitude of founder effect. GBS allows for a more comprehensive assessment of genetic diversity, which will enhance our understanding of population-level dynamics. Our findings indicate

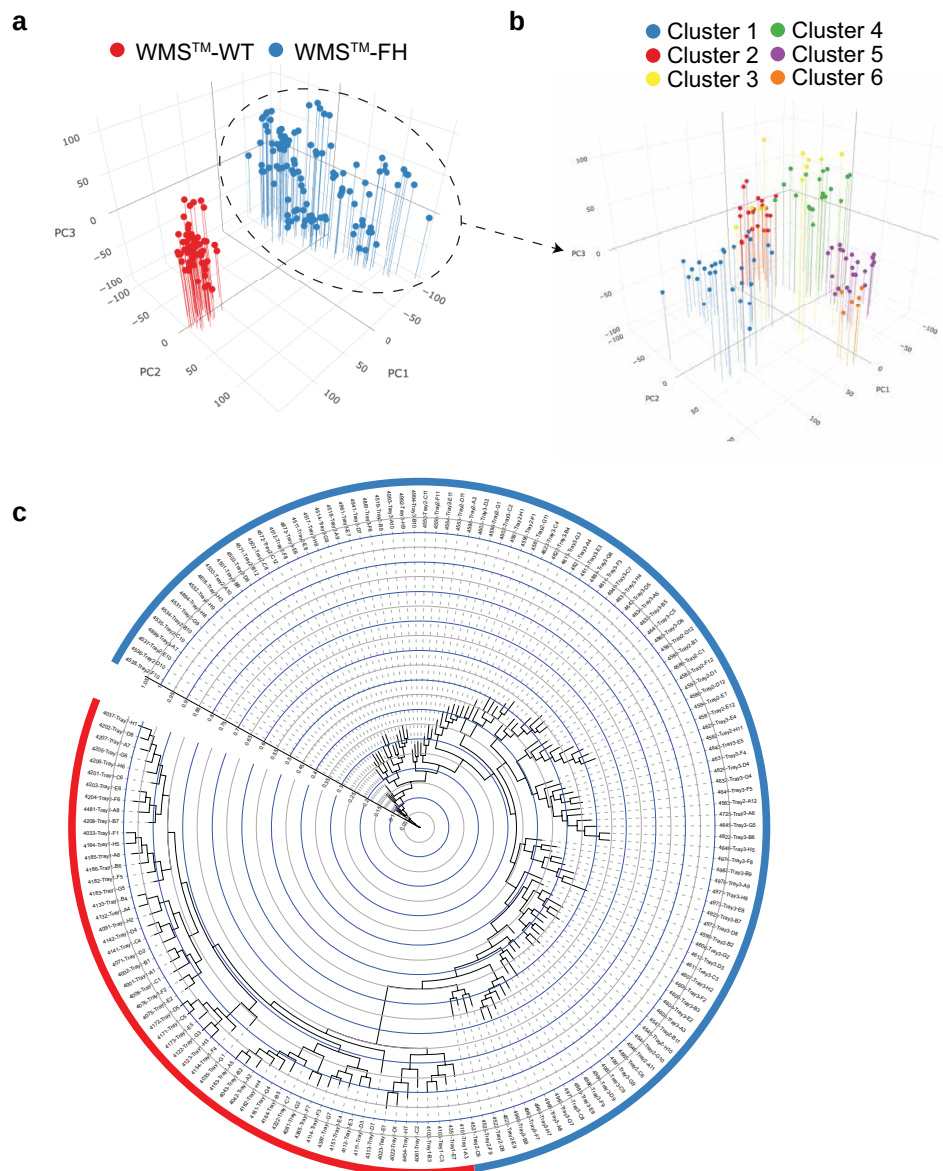


Fig. 4. Population dynamics of WMSTM. a) Principal component analysis of the WMSTM-WT (red) and WMSTM-FH (blue) herds. There is a clear separation between the WMSTM-WT and WMSTM-FH herds, likely due to the association with the FH allele. b) To provide improved resolution of a single herd, the WMSTM-FH herd was reanalyzed separately. Removing the effect caused by the FH allele allows for finer segregation of subpopulations, which are colored independently. c) GBS phylogenetic tree indicating the relationship between all individuals across the WMSTM-WT and WMSTM-FH herds. Samples from each herd were identified by the outer color band (color from panel a).

that, although the WMSTM animals display overall genetic similarity, there is evidence of structured genetic diversity within both the WMSTM-WT and WMSTM-FH herds as well as within the WMSTM-FH herd. Second, GBS helps associate specific traits with genetic loci. While inbred animal models have been a staple in biomedical research for decades, growing evidence suggests that outbred models may offer superior translational relevance (French et al. 2015; Tuttle et al. 2018). The observed genetic segregation between the WMSTM-FH and WMSTM-FH samples underscores the feasibility of using GBS to detect and track genetic variation related to phenotypes across a population, advancing the utility of outbred models in research.

Conclusions

In this study, we present the draft genome of the WMSTM, a minipig line developed as a novel tool for biomedical training

and translational research. Our assembly achieved high completeness, with a BUSCO score exceeding that of the current *S. scrofa* reference. We also successfully annotated the genome. As expected, the WMSTM genome exhibits a strong syntenic identity with the *S. scrofa* genome. To illustrate its research utility, we compared the WMSTM and human *NF1* and *AHR* genes, underscoring the potential of WMSTM as a model for studying *NF1* disease mechanisms and advancing precision therapeutics. Additionally, our population-level analysis of the WMSTM herd demonstrates clear genetic structuring between WMSTM-WT and WMSTM-FH individuals, establishing a foundation for genetic monitoring and trait association within this unique animal model.

While this study focuses on the genome of WMSTM, we recognize that additional analysis will be needed to assess genetic and epigenetic drivers of disease. For example, WMSTM inherently possess DNA methylation patterns of accelerated biological aging, including increased susceptibility to cardiovascular disease, that

parallels epigenetic signatures in humans associated with body mass index (Schachtschneider *et al.* 2021). These findings suggest that the WMS™ model may serve as a translational bridge for studying disease mechanisms relevant to human health. Future studies will extend this work by incorporating comparative analysis of immune-related gene variants and cancer-associated mutations. Such investigations could enhance the utility of this genome assembly for the broader biomedical community, particularly in the context of genome editing applications and disease modeling. By exploring these avenues, we hope to provide further insights into the genetic underpinnings of chronic diseases and contribute to the development of improved translational models for aging and disease research.

Data availability

The WGS has been deposited at GenBank under the accession JBJJOT000000000.1. The version described in this paper is JBJJOT010000000 under NCBI BioProject PRJNA1190099. The sequencing data used to assemble the genome as well as the GBS data are also associated with NCBI BioProject PRJNA1190099. The unannotated genome, GFF3 annotation, and supplemental data files are available at figshare at <https://doi.org/10.6084/m9.figshare.27904893.v1>.

Acknowledgments

The authors utilized the University of Wisconsin-Madison Biotechnology Center (UWBC)'s DNA Sequencing Facility (Research Resource Identifier—RRID:SCR_017759) to extract HMW DNA and generate and sequence PacBio, ONT, and Hi-C libraries. The UWBC Bioinformatics Resource Center (BRC) Core Facility (RRID:SCR_017799) led the assembly and data analysis. The UWBC Advanced Genome Editing Laboratory (RRID:SCR_021070) was supported by the UWCCC Support Grant (P30CA014520). The authors also thank Molly Nellen and Brontë Rosenbluh for their assistance in processing GBS samples.

Funding

This work was supported by funding from the Neurofibromatosis Network, Links for Lauren, NF North Central, NF Central Plains, NF Northeast, NF Michigan, Cheri Stewart, and NF Team. Additional funding was provided by the National Institutes of Health NCI P30CA014520 and NIEHS R35ES028377.

Conflicts of interest

DS and JLR are inventors of WMS™.

Literature cited

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.* 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 630(8016):493–500. doi:10.1038/s41586-024-07487-w.
- Ali A, Kemter E, Wolf E. 2024. Advances in organ and tissue xenotransplantation. *Annu Rev Anim Biosci*. 12:369–390. doi:10.1146/annurev-animal-021122-102606.
- Anand RP, Layer JV, Heja D, Hirose T, Lassiter G, Firl DJ, Paragas VB, Akkad A, Chhangawala S, Colvin RB, *et al.* 2023. Design and testing of a humanized porcine donor for xenotransplantation. *Nature*. 622(7982):393–401. doi:10.1038/s41586-023-06594-4.
- Avilla MN, Malecki KMC, Hahn ME, Wilson RH, Bradfield CA. 2020. The ah receptor: adaptive metabolism, ligand diversity, and the xenokine model. *Chem Res Toxicol*. 33(4):860–879. doi:10.1021/acs.chemrestox.9b00476.
- Bernards A, Snijders AJ, Hannigan GE, Murthy AE, Gusella JF. 1993. Mouse neurofibromatosis type 1 cDNA sequence reveals high degree of conservation of both coding and non-coding mRNA segments. *Hum Mol Genet*. 2(6):645–650. doi:10.1093/hmg/2.6.645.
- Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 81(5):1084–1097. doi:10.1086/521987.
- Browning BL, Browning SR. 2016. Genotype imputation with millions of reference samples. *Am J Hum Genet*. 98(1):116–126. doi:10.1016/j.ajhg.2015.11.020.
- Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*. 31(12):1119–1125. doi:10.1038/nbt.2727.
- Campbell C, Fuentes A, MacKinnon K, Bearder S, Stumpf R. 2011. *Primates in Perspective*. 2nd ed. Oxford University Press.
- Carvalho C, Gaspar A, Knight A, Vicente L. 2018. Ethical and scientific pitfalls concerning laboratory research with non-human primates, and possible solutions. *Animals (Basel)*. 9(1):12. doi:10.3390/ani9010012.
- Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, *et al.* 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun*. 12(1):60. doi:10.1038/s41467-020-20236-7.
- Cooper EB, Brent LJN, Snyder-Mackler N, Singh M, Sengupta A, Khatiwada S, Malaivijitnond S, Qi Hai Z, Higham JP. 2022. The rhesus macaque as a success story of the anthropocene. *Elife*. 11:e78169. doi:10.7554/eLife.78169.
- Costine-Bartell BA, Martinez-Ramirez L, Normoyle K, Stinson T, Staley KJ, Lillis KP. 2023. 2-photon imaging of fluorescent proteins in living swine. *Sci Rep*. 13(1):14158. doi:10.1038/s41598-023-40638-z.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.
- De La Cruz KF, Townsend EC, Alex Cheong JZ, Salamzade R, Liu A, Sandstrom S, Davila E, Huang L, Xu KH, Wu SY, *et al.* 2024. The porcine skin microbiome exhibits broad fungal antagonism. *Fungal Genet Biol*. 173:103898. doi:10.1016/j.fgb.2024.103898.
- Delahaye C, Nicolas J. 2021. Sequencing DNA with nanopores: troubles and biases. *PLoS One*. 16(10):e0257521. doi:10.1371/journal.pone.0257521.
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, *et al.* 2017. De novo assembly of the aedes aegypti genome using hi-C yields chromosome-length scaffolds. *Science*. 356(6333):92–95. doi:10.1126/science.aal3327.
- Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 3(1):95–98. doi:10.1016/j.cels.2016.07.002.
- Eix EF, Johnson CJ, Wartman KM, Kernien JF, Meudt JJ, Shanmuganayagam D, Gibson ALF, Nett JE. 2022. Ex vivo human and porcine skin effectively model *Candida auris* colonization,

- differentiating robust and poor fungal colonizers. *J Infect Dis*. 225(10):1791–1795. doi:[10.1093/infdis/jiac094](https://doi.org/10.1093/infdis/jiac094).
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6(5):e19379. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379).
- Flakus MJ, Wuschner AE, Wallat EM, Graham M, Shao W, Shanmuganayagam D, Christensen GE, Reinhardt JM, Bayouth JE. 2023a. Validation of CT-based ventilation and perfusion biomarkers with histopathology confirms radiation-induced pulmonary changes in a porcine model. *Sci Rep*. 13(1):9377. doi:[10.1038/s41598-023-36292-0](https://doi.org/10.1038/s41598-023-36292-0).
- Flakus MJ, Wuschner AE, Wallat EM, Shao W, Meudt J, Shanmuganayagam D, Christensen GE, Reinhardt JM, Bayouth JE. 2023b. Robust quantification of CT-ventilation biomarker techniques and repeatability in a porcine model. *Med Phys*. 50(10):6366–6378. doi:[10.1002/mp.16400](https://doi.org/10.1002/mp.16400).
- Flakus MJ, Wuschner AE, Wallat EM, Shao W, Shanmuganayagam D, Christensen GE, Reinhardt JM, Li K, Bayouth JE. 2023c. Quantifying robustness of CT-ventilation biomarkers to image noise. *Front Physiol*. 14:1040028. doi:[10.3389/fphys.2023.1040028](https://doi.org/10.3389/fphys.2023.1040028).
- Frangiogiannis NG. 2022. Why animal model studies are lost in translation. *J Cardiovasc Aging*. 2(2):22. doi:[10.20517/jca.2022.10](https://doi.org/10.20517/jca.2022.10).
- French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, Shepard KG, Price HC, King D, Witt KL, et al. 2015. Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. *Environ Health Perspect*. 123(3):237–245. doi:[10.1289/ehp.1408202](https://doi.org/10.1289/ehp.1408202).
- Gerdtz V, Wilson HL, Meurens F, van Drunen Littel-van den Hurk S, Wilson D, Walker S, Wheler C, Townsend H, Potter AA. 2015. Large animal models for vaccine development and testing. *ILAR J*. 56(1):53–62. doi:[10.1093/ilar/ilv009](https://doi.org/10.1093/ilar/ilv009).
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 9(2):e90346. doi:[10.1371/journal.pone.0090346](https://doi.org/10.1371/journal.pone.0090346).
- Gomez-Raya L, Rodriguez C, Barragan C, Silio L. 2015. Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genet Sel Evol*. 47(1):81. doi:[10.1186/s12711-015-0153-1](https://doi.org/10.1186/s12711-015-0153-1).
- Griffith BP, Goerlich CE, Singh AK, Rothblatt M, Lau CL, Shah A, Lorber M, Grazioli A, Saharia KK, Hong SN, et al. 2022. Genetically modified porcine-to-human cardiac xenotransplantation. *N Engl J Med*. 387(1):35–44. doi:[10.1056/NEJMoa2201422](https://doi.org/10.1056/NEJMoa2201422).
- Grimm D. 2023. U.S., European researchers face monkey shortage crisis. *Science*. 380(6645):567–568. doi:[10.1126/science.adi6512](https://doi.org/10.1126/science.adi6512).
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29(8):1072–1075. doi:[10.1093/bioinformatics/btt086](https://doi.org/10.1093/bioinformatics/btt086).
- Gutierrez K, Dicks N, Glanzner WG, Agellon LB, Bordignon V. 2015. Efficacy of the porcine species in biomedical research. *Front Genet*. 6:293. doi:[10.3389/fgene.2015.00293](https://doi.org/10.3389/fgene.2015.00293).
- Hahn ME. 2002. Aryl hydrocarbon receptors: diversity and evolution. *Chem Biol Interact*. 141(1-2):131–160. doi:[10.1016/s0009-2797\(02\)00070-4](https://doi.org/10.1016/s0009-2797(02)00070-4).
- Hahn ME, Karchner SI, Merson RR. 2017. Diversity as opportunity: insights from 600 million years of AHR evolution. *Curr Opin Toxicol*. 2:58–71. doi:[10.1016/j.cotox.2017.02.003](https://doi.org/10.1016/j.cotox.2017.02.003).
- Hanna AS, Hellenbrand DJ, Schomberg DT, Salamat SM, Loh M, Wheeler L, Hanna B, Ozaydin B, Meudt J, Shanmuganayagam D. 2022. Brachial plexus anatomy in the miniature swine as compared to human. *J Anat*. 240(1):172–181. doi:[10.1111/joa.13525](https://doi.org/10.1111/joa.13525).
- Hasler-Rapacz J, Ellegren H, Fridolfsson A-K, Kirkpatrick B, Kirk S, Andersson L, Rapacz J. 1998. Identification of a mutation in the low density lipoprotein receptor gene associated with recessive familial hypercholesterolemia in swine. *Am J Med Genet*. 76(5):379–386. doi:[10.1002/ajmg.1330760521](https://doi.org/10.1002/ajmg.1330760521).
- Hess MK, Rowe SJ, Van Stijn TC, Henry HM, Hickey SM, Brauning R, McCulloch AF, Hess AS, Kirk MR, Kumar S, et al. 2020. A restriction enzyme reduced representation sequencing approach for low-cost, high-throughput metagenome profiling. *PLoS One*. 15(4):e0219882. doi:[10.1371/journal.pone.0219882](https://doi.org/10.1371/journal.pone.0219882).
- Horton MV, Johnson CJ, Kernien JF, Patel TD, Lam BC, Cheong JZA, Meudt JJ, Shanmuganayagam D, Kalan LR, Nett JE. 2020. *Candida auris* forms high-burden biofilms in skin niche conditions and on porcine skin. *mSphere*. 5(1):e00910-19. doi:[10.1128/mSphere.00910-19](https://doi.org/10.1128/mSphere.00910-19).
- Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 33(6):1635–1638. doi:[10.1093/molbev/msw046](https://doi.org/10.1093/molbev/msw046).
- Jiang H, Lei R, Ding SW, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*. 15(1):182. doi:[10.1186/1471-2105-15-182](https://doi.org/10.1186/1471-2105-15-182).
- Johnson CJ, Cheong JZA, Shanmuganayagam D, Kalan LR, Nett JE. 2022. Modeling *Candida auris* colonization on porcine skin ex vivo. *Methods Mol Biol*. 2517:251–258. doi:[10.1007/978-1-0716-2417-3_20](https://doi.org/10.1007/978-1-0716-2417-3_20).
- Johnson CJ, Eix EF, Lam BC, Wartman KM, Meudt JJ, Shanmuganayagam D, Nett JE. 2021. Augmenting the activity of chlorhexidine for decolonization of *Candida auris* from porcine skin. *J Fungi (Basel)*. 7(10):804. doi:[10.3390/jof7100804](https://doi.org/10.3390/jof7100804).
- Jouhilahti EM, Peltonen S, Heape AM, Peltonen J. 2011. The pathoetiology of neurofibromatosis 1. *Am J Pathol*. 178(5):1932–1939. doi:[10.1016/j.ajpath.2010.12.056](https://doi.org/10.1016/j.ajpath.2010.12.056).
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5):540–546. doi:[10.1038/s41587-019-0072-8](https://doi.org/10.1038/s41587-019-0072-8).
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 9(4):357–359. doi:[10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).
- Lelovas PP, Kostomitsopoulos NG, Xanthos TT. 2014. A comparative anatomic and physiologic overview of the porcine heart. *J Am Assoc Lab Anim Sci*. 53(5):432–438.
- Letunic I, Bork P. 2024. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 52(W1):W78–W82. doi:[10.1093/nar/gkae268](https://doi.org/10.1093/nar/gkae268).
- Li K, Xu P, Wang J, Yi X, Jiao Y. 2023. Identification of errors in draft genome assemblies at single-nucleotide resolution for quality assessment and improvement. *Nat Commun*. 14(1):6556. doi:[10.1038/s41467-023-42336-w](https://doi.org/10.1038/s41467-023-42336-w).
- Lin Z, Guo F, Gregorich ZR, Sun R, Zhang H, Hu Y, Shanmuganayagam D, Ge Y. 2018. Comprehensive characterization of swine cardiac troponin T proteoforms by top-down mass spectrometry. *J Am Soc Mass Spectrom*. 29(6):1284–1294. doi:[10.1007/s13361-018-1925-y](https://doi.org/10.1007/s13361-018-1925-y).
- Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ, Lowe R, Yan Q, Zhang J, Vu H, et al. 2023. Universal DNA methylation age across mammalian tissues. *Nat Aging*. 3(9):1144–1166. doi:[10.1038/s43587-023-00462-6](https://doi.org/10.1038/s43587-023-00462-6).
- Lu H, Giordano F, Ning Z. 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*. 14(5):265–279. doi:[10.1016/j.gpb.2016.05.004](https://doi.org/10.1016/j.gpb.2016.05.004).
- Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. 2021. Importance of the pig as a human biomedical model. *Sci Transl Med*. 13(621):eabd5758. doi:[10.1126/scitranslmed.abd5758](https://doi.org/10.1126/scitranslmed.abd5758).

- Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. *Curr Protoc.* 1(12):e323. doi:[10.1002/cpz1.323](https://doi.org/10.1002/cpz1.323).
- Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 14(1):e1005944. doi:[10.1371/journal.pcbi.1005944](https://doi.org/10.1371/journal.pcbi.1005944).
- McLean DT, Meudt JJ, Lopez Rivera LD, Schomberg DT, Pavelec DM, Duellman TT, Buehler DG, Schwartz PB, Graham M, Lee LM, et al. 2023. Single-cell RNA sequencing of neurofibromas reveals a tumor microenvironment favorable for neural regeneration and immune suppression in a neurofibromatosis type 1 porcine model. *Front Oncol.* 13:1253659. doi:[10.3389/fonc.2023.1253659](https://doi.org/10.3389/fonc.2023.1253659).
- Meurens F, Summerfield A, Nauwynck H, Saif L, Gerds V. 2012. The pig: a model for human infectious diseases. *Trends Microbiol.* 20(1):50–57. doi:[10.1016/j.tim.2011.11.002](https://doi.org/10.1016/j.tim.2011.11.002).
- Meyerholz DK, Burrough ER, Kirchhof N, Anderson DJ, Helke KL. 2024. Swine models in translational research and medicine. *Vet Pathol.* 61(4):512–523. doi:[10.1177/03009858231222235](https://doi.org/10.1177/03009858231222235).
- Miranpuri GS, Schomberg DT, Stan P, Chopra A, Buttar S, Wood A, Radzin A, Meudt JJ, Resnick DK, Shanmuganayagam D. 2018. Comparative morphometry of the Wisconsin Miniature Swine(TM) thoracic spine for modeling human spine in translational spinal cord injury research. *Ann Neurosci.* 25(4):210–218. doi:[10.1159/000488022](https://doi.org/10.1159/000488022).
- Moazami N, Stern JM, Khalil K, Kim JI, Narula N, Mangiola M, Weldon EP, Kagermazova L, James L, Lawson N, et al. 2023. Pig-to-human heart xenotransplantation in two recently deceased human recipients. *Nat Med.* 29(8):1989–1997. doi:[10.1038/s41591-023-02471-9](https://doi.org/10.1038/s41591-023-02471-9).
- Namous H, Krueger C, Cheng Y, Melo PHC, Peppas A, Kaluza GL, Stoffregen WC, Reed J, Khatib H, Granada JF. 2023a. Longitudinal dynamics of circulating mirnas in a swine model of familial hypercholesterolemia during early atherosclerosis. *Sci Rep.* 13(1):19355. doi:[10.1038/s41598-023-46762-0](https://doi.org/10.1038/s41598-023-46762-0).
- Namous H, Strillacci MG, Braz CU, Shanmuganayagam D, Krueger C, Peppas A, Stoffregen WC, Reed J, Granada JF, Khatib H. 2023b. ITGB2 is a central hub-gene associated with inflammation and early fibro-atheroma development in a swine model of atherosclerosis. *Atheroscler Plus.* 54:30–41. doi:[10.1016/j.athplu.2023.11.001](https://doi.org/10.1016/j.athplu.2023.11.001).
- National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Health and Medicine Division; Institute for Laboratory Animal Research; Board on Health Sciences Policy; Committee on the State of the Science and Future Needs for Nonhuman Primate Model Systems. 2023. The national academies collection: reports funded by national institutes of health. In: Yost OC, Downey A, Ramos KS, editors. *Nonhuman Primate Models in Biomedical Research: State of the Science and Future Needs*. National Academies Press (US). p. 1–268. Copyright 2023 by the National Academy of Sciences. All rights reserved.
- Ortiz EM. 2019. Vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo. doi:[10.5281/zenodo.2540861](https://doi.org/10.5281/zenodo.2540861).
- Pabst R. 2020. The pig as a model for immunology research. *Cell Tissue Res.* 380(2):287–304. doi:[10.1007/s00441-020-03206-9](https://doi.org/10.1007/s00441-020-03206-9).
- Perleberg C, Kind A, Schnieke A. 2018. Genetically engineered pigs as models for human disease. *Dis Model Mech.* 11(1):dmm030783. doi:[10.1242/dmm.030783](https://doi.org/10.1242/dmm.030783).
- Peterson L, Yacoub MH, Ayares D, Yamada K, Eisenson D, Griffith BP, Mohiuddin MM, Eyestone W, Venter JC, Smolenski RT, et al. 2024. Physiological basis for xenotransplantation from genetically modified pigs to humans. *Physiol Rev.* 104(3):1409–1459. doi:[10.1152/physrev.00041.2023](https://doi.org/10.1152/physrev.00041.2023).
- Pirasteh A, Periyasamy S, Meudt JJ, Liu Y, Lee LM, Schachtschneider KM, Schook LB, Gaba RC, Mao L, Said A, et al. 2022. Staging liver fibrosis by fibroblast activation protein inhibitor PET in a human-sized swine model. *J Nucl Med.* 63(12):1956–1961. doi:[10.2967/jnumed.121.263736](https://doi.org/10.2967/jnumed.121.263736).
- Poland A, Palen D, Glover E. 1994. Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol Pharmacol.* 46(5):915–921. doi:[10.1016/S0026-895X\(25\)09833-5](https://doi.org/10.1016/S0026-895X(25)09833-5).
- Porrett PM, Orandi BJ, Kumar V, Houpp J, Anderson D, Cozette Killian A, Hauptfeld-Dolejssek V, Martin DE, Macedon S, Budd N, et al. 2022. First clinical-grade porcine kidney xenotransplant using a human decedent model. *Am J Transplant.* 22(4):1037–1053. doi:[10.1111/ajt.16930](https://doi.org/10.1111/ajt.16930).
- Rapacz J, Hasler-Rapacz JO, Hu ZL, Rapacz JM, Vögeli P, Hojný J, Janik A. 1994. Identification of new apolipoprotein B epitopes and haplotypes and their distribution in swine populations. *Anim Genet.* 25 Suppl 1(S1):51–57. doi:[10.1111/j.1365-2052.1994.tb00403.x](https://doi.org/10.1111/j.1365-2052.1994.tb00403.x).
- Raval NR, Nasser A, Madsen CA, Beschornier N, Beaman EE, Juhl M, Lehel S, Palner M, Svarer C, Plavén-Sigraý P, et al. 2022. An in vivo pig model for testing novel positron emission tomography radioligands targeting cerebral protein aggregates. *Front Neurosci.* 16:847074. doi:[10.3389/fnins.2022.847074](https://doi.org/10.3389/fnins.2022.847074).
- Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics.* 19(1):460. doi:[10.1186/s12859-018-2485-7](https://doi.org/10.1186/s12859-018-2485-7).
- Rubinstein CD, McLean DT, Lehman BP, Meudt JJ, Schomberg DT, Krentz KJ, Reichert JL, Meyer MB, Adams M, Konsitzke CM, et al. 2021. Assessment of mosaicism and detection of cryptic alleles in CRISPR/cas9-engineered neurofibromatosis type 1 and TP53 mutant porcine models reveals overlooked challenges in precision modeling of human diseases. *Front Genet.* 12:721045. doi:[10.3389/fgene.2021.721045](https://doi.org/10.3389/fgene.2021.721045).
- Schachtschneider KM, Schook LB, Meudt JJ, Shanmuganayagam D, Zoller JA, Haghani A, Li CZ, Zhang J, Yang A, Raj K, et al. 2021. Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *Geroscience.* 43(5):2467–2483. doi:[10.1007/s11357-021-00439-6](https://doi.org/10.1007/s11357-021-00439-6).
- Scheben A, Batley J, Edwards D. 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J.* 15(2):149–161. doi:[10.1111/pbi.12645](https://doi.org/10.1111/pbi.12645).
- Schomberg DT, Tellez A, Meudt JJ, Brady DA, Dillon KN, Arowolo FK, Wicks J, Rousselle SD, Shanmuganayagam D. 2016. Miniature swine for preclinical modeling of complexities of human disease for translational scientific discovery and accelerated development of therapies and medical devices. *Toxicol Pathol.* 44(3):299–314. doi:[10.1177/0192623315618292](https://doi.org/10.1177/0192623315618292).
- Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics.* 37(12):1639–1643. doi:[10.1093/bioinformatics/btaa1016](https://doi.org/10.1093/bioinformatics/btaa1016).
- Song J, Clagett-Dame M, Peterson RE, Hahn ME, Westler WM, Sicinski RR, DeLuca HF. 2002. A ligand for the aryl hydrocarbon receptor isolated from lung. *Proc Natl Acad Sci U S A.* 99(23):14694–14699. doi:[10.1073/pnas.232562899](https://doi.org/10.1073/pnas.232562899).
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9):1312–1313. doi:[10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Suenderhauf C, Parrott N. 2013. A physiologically based pharmacokinetic model of the minipig: data compilation and model implementation. *Pharm Res.* 30(1):1–15. doi:[10.1007/s11095-012-0911-5](https://doi.org/10.1007/s11095-012-0911-5).
- Tang GQ, Xue J, Lian MJ, Yang RF, Liu TF, Zeng ZY, Jiang AA, Jiang YZ, Zhu L, Bai L, et al. 2013. Inbreeding and genetic diversity in three

- imported swine breeds in China using pedigree data. *Asian-Australas J Anim Sci.* 26(6):755–765. doi:[10.5713/ajas.2012.12645](https://doi.org/10.5713/ajas.2012.12645).
- Tuttle AH, Philip VM, Chesler EJ, Mogil JS. 2018. Comparing phenotypic variation between inbred and outbred mice. *Nat Methods.* 15(12):994–996. doi:[10.1038/s41592-018-0224-7](https://doi.org/10.1038/s41592-018-0224-7).
- Wallat EM, Wuschner AE, Flakus MJ, Christensen GE, Reinhardt JM, Shanmuganayagam D, Bayouth JE. 2021. Radiation-induced airway changes and downstream ventilation decline in a swine model. *Biomed Phys Eng Express.* 7(6):10.1088/2057-1976/ac3197. doi:[10.1088/2057-1976/ac3197](https://doi.org/10.1088/2057-1976/ac3197).
- Wang N, Yuan Y, Wang H, Yu D, Liu Y, Zhang A, Gowda M, Nair SK, Hao Z, Lu Y, et al. 2020. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Sci Rep.* 10(1):16308. doi:[10.1038/s41598-020-73321-8](https://doi.org/10.1038/s41598-020-73321-8).
- Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 39(11):1348–1365. doi:[10.1038/s41587-021-01108-x](https://doi.org/10.1038/s41587-021-01108-x).
- Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, et al. 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience.* 9(6):giaa051. doi:[10.1093/gigascience/giaa051](https://doi.org/10.1093/gigascience/giaa051).
- Wolfensohn S. 2010. Old world monkeys. In: *The UFAW handbook on the care and management of laboratory and other research animals.* John Wiley & Sons Ltd. p. 592–617.
- Wuschner AE, Flakus MJ, Wallat EM, Reinhardt JM, Shanmuganayagam D, Christensen GE, Bayouth JE. 2022a. Measuring indirect radiation-induced perfusion change in fed vasculature using dynamic contrast CT. *J Pers Med.* 12(8):1254. doi:[10.3390/jpm12081254](https://doi.org/10.3390/jpm12081254).
- Wuschner AE, Flakus MJ, Wallat EM, Reinhardt JM, Shanmuganayagam D, Christensen GE, Gerard SE, Bayouth JE. 2022b. CT-derived vessel segmentation for analysis of post-radiation therapy changes in vasculature and perfusion. *Front Physiol.* 13:1008526. doi:[10.3389/fphys.2022.1008526](https://doi.org/10.3389/fphys.2022.1008526).
- Wuschner AE, Wallat EM, Flakus MJ, Shanmuganayagam D, Meudt J, Christensen GE, Reinhardt JM, Miller JR, Lawless MJ, Baschnagel AM, et al. 2021. Radiation-induced hounsfield unit change correlates with dynamic CT perfusion better than 4DCT-based ventilation measures in a novel-swine model. *Sci Rep.* 11(1):13156. doi:[10.1038/s41598-021-92609-x](https://doi.org/10.1038/s41598-021-92609-x).
- Yang L, Yin H, Bai L, Yao W, Tao T, Zhao Q, Gao Y, Teng J, Xu Z, Lin Q, et al. 2024. Mapping and functional characterization of structural variation in 1060 pig genomes. *Genome Biol.* 25(1):116. doi:[10.1186/s13059-024-03253-3](https://doi.org/10.1186/s13059-024-03253-3).

Editor: E. Clark