Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

TITLE: Towards a Kingdom of Reproductive Life – the Core Sperm Proteome

Short title: The Core Sperm Proteome

AUTHORS: Taylor Pini[1*], Brett Nixon[2,3], Timothy L. Karr[4,5], Raffaele Teperino[6,7], Adrián Sanz-Moreno[8], Patricia da Silva-Buttkus[8], Frank Tüttelmann[9], Sabine Kliesch[10], Valérie Gailus-Durner[8], Helmut Fuchs[8], Susan Marschall[6], Martin Hrabě de Angelis[6,7,8,11], David A. Skerrett-Byrne[3,6,7,12*]

[1] School of Veterinary Science, The University of Queensland, Gatton, QLD, Australia

[2] Priority Research Centre for Reproductive Science, School of Environmental and Life Sciences, College of Engineering, Science and Environment, The University of Newcastle, Callaghan, NSW, Australia

[3] Hunter Medical Research Institute, Infertility and Reproduction Research Program, New Lambton Heights, NSW, Australia.

[4] Biosciences Mass Spectrometry Core Research Facility, Knowledge Enterprise, Arizona State University, USA

[5] ASU-Banner Neurodegenerative Disease Research Center, The Biodesign Institute, Arizona State University, USA

[6] Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

[7] German Center for Diabetes Research (DZD) Neuherberg, Germany

[8] Institute of Experimental Genetics, German Mouse Clinic, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[9] Institute of Reproductive Genetics, Centre of Medical Genetics, University and University Hospital of Münster, Germany

[10] Department of Clinical and Surgical Andrology, Centre of Reproductive Medicine and Andrology, University and University Hospital of Münster, Germany

[11] Chair of Experimental Genetics, TUM School of Life Sciences, Technische Universität München, Freising, Germany

[12] School of Biomedical Sciences and Pharmacy, College of Health, Medicine and Wellbeing, The University of Newcastle, Callaghan, NSW, Australia

*Corresponding authors: T.Pini@uq.edu.au (T.P.) & David.Skerrett-Byrne@helmholtz-munich.de (D.A.S.B)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

ABSTRACT

Reproductive biology is often considered in three siloed research areas; humans, agriculture and wildlife. Yet, each demand solutions for treatment of subfertility, fertility biomarkers, development of assisted reproductive technologies and effective contraception. To efficiently develop solutions applicable to all species, we must improve our understanding of the common biology underpinning reproductive processes. Accordingly, we integrate proteomic data from 29 publicly available datasets (>2 TB of data) to characterize mature sperm proteomes spanning 12 vertebrate species, identifying 13,853 proteins. Although human and mouse have relatively well-annotated sperm proteomes, many non-model species rely heavily on predicted or homology-inferred identifications. Despite variation in proteome size, composition and reproductive strategies, comparative analyses revealed that vertebrates share a fundamental molecular framework essential for sperm function. A core set of 45 species-level and 135 order-level conserved proteins mapped to critical processes, including energy generation, acrosome function, as well as novel signalling pathways (BAG2 and FAT10). Knockout mouse models further validate the significance of these conserved proteins, demonstrating that their disruption impairs sperm motility and fertilization capacity. Moreover, we discovered loss-of-function variants of two additional core sperm proteins in clinical samples, linking them to severe sperm defects. Intriguingly, *in-silico* analysis reveals function-driven, context-dependent diversity surpassing evolutionary patterns. Collectively, these results highlight the value of integrating publicly available datasets and underscore the need for improved genome/proteome annotation in non-model species in mammals. This work provides a foundation for developing cross-species strategies to enhance fertility treatments, assisted reproductive technologies, and conservation

efforts. All data is available via ShinySpermKingdom (https://reproproteomics.shinyapps.io/ShinySpermKingdom/).

KEYWORDS: Sperm, sperm proteome, fertility, data reanalysis, Pebp4, Echs1, Etfb, Ndufa10, Aldh7a1, proteomics, bioinformatics

IN BRIEF: Sperm function is essential for fertility across humans, agriculture, and wildlife, yet comparative studies remain limited. This study integrates multi-species proteomic data to identify a core sperm proteome, uncovering conserved molecular pathways and validating novel sperm proteins critical for motility and fertilization.

WORD COUNT: 7,358

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

INTRODUCTION

Reproductive biology is often considered in three siloed areas: humans, domesticated animals and wildlife. Despite their differences, there are several common needs across these species; efficient production, treatment of subfertility and infertility, development of assisted reproductive technologies (ARTs) and effective contraception (Comizzoli and Holt 2019, Duffy, et al. 2020). To effectively meet these needs, research should focus on developing solutions that are applicable across species where possible. However, to achieve this goal, we need a better understanding of the common reproductive biology across species.

Reproductive physiology is incredibly diverse across the animal kingdom, with many strategies unique to particular evolutionary lineages. In the context of mating, examples of this diversity include the location of semen deposition (Suarez and Pacey 2006), the ability to store spermatozoa in the female tract for extended periods (Holt and Lloyd 2010) and variable sperm morphologies (Fitzpatrick, et al. 2022). Even between species of the same class, some of these traits appear to differ significantly. There is potential to both take advantage of common pathways and potentially exploit strategies that are unique to some species.

This area of study has significant potential benefits on the male side of the reproductive equation. For example, high quality transcriptomic and proteomic studies on the formation of biological sperm storage may highlight pathways that could be targeted to prolong the *in vitro* shelf life of spermatozoa from a variety of species. As an example, koala sperm display an exceptional longevity of up to 42 days post-ejaculation during *ex vivo* storage (Johnston, et al. 2000, Johnston, et al. 2012, Skerrett-Byrne, et al. 2021a). Understanding the biochemical principles of this longevity would yield potential benefits as diverse as human ARTs and addressing the logistical issues of extending the shelf life of fresh semen in the beef and dairy

industries (Murphy, et al. 2017). Alternatively, similar studies of spermatozoa and seminal plasma from species with high sperm competition could identify proteins and non-coding RNA species that may be exploited to treat subfertility and improve ARTs. To make such advances, further basic discovery research is required.

With decreasing costs and increasing sensitivity, proteomic profiling of the male gamete has become widespread (Mohanty, et al. 2015). While sperm proteomes have been published for a variety of species, we are yet to capitalize on the suggestion of Oliva, Martínez-Heredia, and Estanyol (Oliva, et al. 2008), to identify conserved proteins from among the wealth of proteomic data that has been collected. As yet there has been limited exploration of cross species analyses; with a single study having compared the sperm proteomes of rodents and ungulates (Bayram, et al. 2016) and another comparing three closely related mouse species with differences in sperm competition (Vicens, et al. 2017). The results of these studies suggest that sperm proteins fall into two categories; (i) highly conserved "core" proteins and (ii) rapidly evolving proteins that are unique to species or taxonomic groups. Importantly, many of the core proteins had important biological roles (e.g. spermatogenesis, capacitation (Bayram, et al. 2016)), suggesting they may provide key avenues for developing cross species reproductive solutions.

A comprehensive cross species sperm proteomic analysis would provide the highest quality data from which to build further research. However, the physical collection of gametes from a large assortment of species involves many difficulties, not the least of which are the considerable expense and logistical management of samples. Thus, as a precursor to further experimental studies, we herein present an *in-silico* analysis of publicly available proteomic data from 12 species, representing the most comprehensive cross species sperm proteome published to date. Using this information, we establish an up-to-date core sperm proteome, highlighting candidate

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

pathways that are highly conserved across species. In addition, we compare proteomes across species based on biological contexts to highlight potentially important pathways for further study.

MATERIALS AND METHODS

Chemicals and reagents

Unless otherwise specified, all reagents were purchased from Merck.

Proteomic data sourcing

Publicly available proteomic data was sourced from the ProteomeXchange repository (www.proteomexchange.org) (Deutsch, et al. 2023), drawing data from a range of partner repositories including PRIDE (Perez-Riverol, et al. 2022), iProX (Chen, et al. 2022), and MassIVE (Choi, et al. 2020). The search term '*sperm*' was initially used to obtain all proteomic datasets containing this keyword. Next, datasets were refined by phylum, with only species in the Chordata phylum retained. In the remaining datasets, any without the RAW mass spectrometry output files (e.g. .WIFF, .RAW) available were excluded. From this filtered list, the remaining datasets were refined based on the following criteria; use of bottom-up proteomics employing data dependent acquisition on fresh, mature cauda epididymal or ejaculated spermatozoa from wildtype animals. Studies were not excluded based on sample enrichment techniques (e.g. density gradient centrifugation, isolation of plasma membrane proteins). Beyond these criteria, studies were excluded if the experimental treatment of each raw MS output file could not be determined or available RAW files deposited were not of sufficient detail or file type for reanalysis. In the first case, clarification was sought from the publishing authors to establish file identities prior to

exclusion. The application of these criteria resulted in 29 datasets (Table 1) and over 2TB of RAW data.

International Mouse Phenotyping Consortium mouse models, histology, and data collection

The International Mouse Phenotyping Consortium (IMPC) database (Dickinson, et al. 2016, Groza, et al. 2022) was mined for genetic knockout mice overlapping with those identified as the core sperm proteome (Fig. 3A), and with special access granted, we crossed referenced with the European Mouse Mutant Archive (EMMA) (Hagn, et al. 2007) to restrict to those gene KOs with available *in vitro* fertilization (IVF) and sperm data. The mouse models were generated using the IMPC targeting strategy with CRISPR/Cas technology at Helmholtz Munich (https://www.mousephenotype.org/understand/the-data/allele-design/). After genotyping, heterozygous × heterozygous matings were set up to generate sufficient mutant mice with littermate $^{+/+}$ controls for phenotyping analysis at the German Mouse Clinic as described, (Fuchs, et al. 2018) and in agreement with the standardized phenotyping pipeline of the IMPC including histopathological analysis (n=2) (https://www.mousephenotype.org/impress/PipelineInfo?id=14) for all lines except for *Aldh7a1* (n=5). We obtained data pertinent to *Aldh7a1* (Aldehyde dehydrogenase 7 family member A1), *Echs1* (Enoyl-CoA hydratase, short chain 1), *Etfb (*Electron transfer flavoprotein subunit beta), *Ndufa10* (NADH:ubiquinone oxidoreductase subunit A10), *Pebp4* (Phosphatidylethanolamine binding protein 4), with a wildtype reference control provided by EMMA. For further information on protocols used by EMMA for sperm collection, analysis and IVF, please see their publicly available resources and videos (https://www.infrafrontier.eu/emma/cryopreservation-protocols/). Briefly, histopathological analyses of formalin-fixed, paraffin-embedded and haematoxylin & eosin (H&E)-stained sections

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

(3 μm-thick) of testis, epididymis, prostate and seminal vesicles from control and mutant mice were performed blind by two pathologists. When applicable, the number of multinucleated giant cells (MGCs) in the seminiferous tubules was counted and expressed per unit area of testis.

Proteome Discoverer processing

Consistent with previous studies (Martin, et al. 2022, Murray, et al. 2021, Skerrett-Byrne, et al. 2022, Skerrett-Byrne, et al. 2021a, Skerrett-Byrne, et al. 2021b, Skerrett-Byrne, et al. 2021c, Smyth, et al. 2022, Staudt, et al. 2022, Trigg, et al. 2021), database searching of each study's RAW files was performed using Proteome Discoverer 2.5 (Thermo Fisher Scientific). SEQUEST HT was used to search against the appropriate UniProt database (Table S1, each downloaded 18th June 2022, including reviewed and unreviewed proteins). Highly stringent database searching criteria were utilized, including up to two missed cleavages, a precursor mass tolerance set to 10 ppm and fragment mass tolerance of 0.02 Da. Trypsin was designated as the digestion enzyme. Cysteine carbamidomethylation was set as a fixed modification while acetylation (K, N-terminus), phosphorylation (S,T,Y) and oxidation (M) were designated as dynamic modifications. Interrogation of the corresponding reversed database was also performed to evaluate the false discovery rate (FDR) of peptide identification using Percolator on the basis of *q*-values, which were estimated from the target-decoy search approach. To filter out target peptide spectrum matches over the decoy-peptide spectrum matches, a fixed FDR of 1% was set at the peptide level. The resultant protein list was exported from Proteome Discoverer 2.5 as an Excel file and further refined to include only those with a protein identification (FDR ≤ 0.01) with at least one or more unique peptides.

Phylogenetic trees and UniProt mapping

NCBI taxonomy numbers were submitted to phylot (v2) to generate a phylogenetic tree, visualized and exported from iTOL (Interactive Tree Of Life) (Letunic and Bork 2007, 2011). Utilising UniProt (https://www.uniprot.org/), each of the sperm proteomes were mapped to the UniProt Knowledge Base to ascertain the level of evidence of each protein (Skerrett-Byrne, et al. 2022, Skerrett-Byrne, et al. 2021a, Skerrett-Byrne, et al. 2021b, Skerrett-Byrne, et al. 2021c, Smyth, et al. 2022). UniProt protein evidence is a measure of the current, manually curated, type of evidence that supports the existence of that protein; experimental evidence 1) at protein level; 2) at transcript level; 3) protein inferred from homology; 4) protein predicted.

Humanization with OmicsBox

To conduct a comparative analysis between all species, a minimum cut-off of at least 500 proteins was applied before proceeding to humanization to maximize the comparisons possible. Data from each of the remaining eight species were uploaded to UniProt to generate a FASTA file for humanization using a custom workflow on the OmicsBox software (version 2.2.4, BioBam Bioinformatics, Valencia, Spain) (https://www.biobam.com/omicsbox). This workflow includes a cloud based DIAMOND BLAST protein search against the human proteome (Buchfink, et al. 2021, Götz, et al. 2008, Skerrett-Byrne, et al. 2021a, Zhang, et al. 2022), with the output restricted to an e-value cut-off of $4.07E^{-10}$ to ensure accurate homologues were obtained (97.5% average conversion).

Identifying conserved and species-specific proteins

Conserved proteins were identified at both species level (i.e., proteins present in all species used for further analysis) and order level (i.e., proteins present in at least one species of all taxonomic

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

orders used for further analysis). Humanized identifications (IDs) were employed for this analysis and lists were compared using jvenn (Bardou, et al. 2014) and DeepVenn (Hulsen 2022) to identify conserved proteins. The analysis at order level included the taxonomic orders Primates (*H. sapiens*), Rodentia (*M. musculus*), Artiodactyla (*B. taurus, O. aries, S. scrofa, T. truncates*), Lagomorpha (*O. cuniculus*), Diprotodontia (*P. cinereus*) and Crocodilia (*C. porosus*).

Comparing the sperm proteome based on biological contexts

Groups of species were compared based on several biological 'contexts', including location of testes (internal vs external), history of selective breeding (yes vs no) and sperm metabolism preference (glycolysis preference vs no preference). Humanized IDs were used for this analysis and species classified into each group are listed in Supplementary Table S17. To account for the stronger influence of human and mouse proteomes due to their extensive inventories, proteins were not included in the analysis if they were only identified in mouse or human spermatozoa. Lists were compared using jvenn (Bardou, et al. 2014) and DeepVenn (Hulsen 2022) to identify conserved and unique proteins.

Bioinformatic analyses of proteomic data

Bioinformatic analyses employed humanized IDs for analysis. High granularity pathway analysis was performed using the Ingenuity Pathway Analysis software package (IPA; Qiagen, Hilden, Germany) as previously described (Martin, et al. 2022, Murray, et al. 2021, Skerrett-Byrne, et al. 2022, Skerrett-Byrne, et al. 2021a, Skerrett-Byrne, et al. 2021b, Skerrett-Byrne, et al. 2021c, Smyth, et al. 2022, Staudt, et al. 2022, Trigg, et al. 2021, Zhang, et al. 2022). Each humanized proteomic list was analysed on the basis of predicted protein subcellular location and classification (other excluded), in addition to canonical pathways and disease and functions, using the IPA *p*-

value enrichment score (a strict cut-off of p-value ≤ 0.05) (Krämer, et al. 2013). The Database for Annotation, Visualization and Integrated Discovery (DAVID, www.david.ncifcrf.gov, v 2021, (Huang da, et al. 2009, Sherman, et al. 2022)) functional annotation clustering tool was used to identify enriched clusters based on gene ontology terms, protein-protein interactions, protein domains, pathways and literature. All searches were performed with default thresholds for similarity, classification and enrichment, using *Homo sapiens* as the background gene list. Clusters were classified as significantly enriched based on Benjamini adjusted p-values ≤ 0.05. Visual protein-protein interaction networks were generated using STRING (www.string-db.org, v 11.5). The humanized proteome was interrogated using UniProt to assess subcellular locations relevant to sperm cells, the following GO terms were used: acrosomal vesicle (GO:0001669), perinuclear theca (GO:0033011), nucleus (GO:0005634), mitochondrion (GO:0005739), axoneme (GO:0005930), cytoskeletal calyx (GO:0033150), sperm midpiece (GO:0097225), head-tail coupling apparatus (GO:0120212), principal piece (GO:0097228), end piece (GO:0097229), annulus (GO:0097227) and flagellum (GO:0036126).

Clustering and network visualization

The refined pathways output from IPA were loaded into Perseus (version 1.6.10.43) (Tyanova, et al. 2016), to carry out unbiased hierarchical clustering across the species. Protein networks of the core sperm proteomes at the species (45 proteins) and order (135 proteins) taxonomic levels were investigated using STRING (version 12.0)(Szklarczyk, et al. 2021) and then visualized and modified using Cytoscape (version 3.8.2). (Shannon, et al. 2003) Basic data handling, if not otherwise stated, was conducted using Microsoft Excel 365 (Version 2211, Microsoft Corporation, Redmond, WA) and GraphPad Prism version 10.4.1 (GraphPad Software; San Diego, CA).

Shiny Application development

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

In accordance with Shiny blueprint outlined by ShinySperm(Skerrett-Byrne, et al. 2024), a Shiny Application was deployed to support the accessibility and interpretability of these datasets within, allowing for effective data-driven insights by the field. The full coding script supporting ShinySpermKingdom (https://reproproteomics.shinyapps.io/ShinySpermKingdom/), can be downloaded from GitHub – https://github.com/DavidSBEire/ShinySpermKingdom. In brief, the ShinySpermKingdom application was built using the shiny package (version 1.9.1) on RStudio (version 2024.04.1+748), with base *R* (version 4.3.3, 2024-02-29). Supporting the functionality and aesthetics of this application are several packages, including: DT, eulerr, ggplot2, openxlsx, plotly, readxl, reshape2, RColorBrewer, and shinydashboard.

Male Reproductive Genomics (MERGE) cohort

The MERGE cohort currently comprises exome/genome data of almost 3,000 men of whom most attended the Centre of Reproductive Medicine and Andrology (CeRA), Münster, for couple infertility. MERGE is continuously growing and for the current study, 2,882 datasets of men with quantitative and/or qualitative sperm defects were queried for the 135 genes encoding the identified conserved sperm proteins. Specifically, 2,327 men had very few or no sperm in the ejaculate (crypto- or azoospermia, HP:0030974/HP:0000027), 437 had various grades of reduced sperm counts (oligozoospermia, HP:0000798) often combined with reduced/impaired sperm motility and/or morphology (astheno-/teratozoospermia, HP:00122077/HP:0012864), and 118 had normal sperm counts but motility and/or morphology defects. The most recent description of MERGE including the details of sequencing are available in (Stallmeyer, et al. 2024). Only well-covered (>20x), rare (minor allele frequency [MAF] <0.01 in gnomAD 2.1.1), coding, homo- or hemizygous, loss-of-function variants (LoF: stop gained, frameshift, splice acceptor/donor) were prioritised.

RESULTS

Establishment of multispecies mature spermatozoa proteomes

A comprehensive search of the ProteomeXchange repository using the keyword '*sperm*' yielded a total of 146 datasets (Fig. 1A). After excluding species from outside the Chordata phylum, a total of 90 datasets remained. Of these, 48 studies were excluded based on our predefined inclusion criteria, such as only studies on functional mature sperm cells (See STAR methods). A further 13 studies were excluded based on insufficient information available for reanalysis. Datasets at each successive level of exclusion are listed in Table S1.

The final cohort comprised 29 datasets, representing 12 species: *Homo sapiens* (Castillo, et al. 2019, Pini, et al. 2020, Schiza, et al. 2018, Urizar-Arenaza, et al. 2019, Vandenbrouck, et al. 2016), *Mus musculus* (Bayram, et al. 2016, Castaneda, et al. 2017, Guyonnet, et al. 2014, Guyonnet, et al. 2012, Liu, et al. 2019, Skerrett-Byrne, et al. 2022, Xu, et al. 2020), *Sus scrofa* (Bayram, et al. 2016, Pérez-Patiño, et al. 2019, Xu, et al. 2021, Zhang, et al. 2022), *Bos taurus* (Bayram, et al. 2016, Byrne, et al. 2012, Kasvandik, et al. 2015, Ramesha, et al. 2020, Shen, et al. 2021), *Crocodylus porosus* (Nixon, et al. 2019), *Oryctolagus cuniculus* (Casares-Crespo, et al. 2019, Juárez, et al. 2020), *Tursiops truncatus* (Fuentes-Albero, et al. 2021), *Ovis aries* (Leahy, et al. 2020), *Phascolarctos cinereus* (Skerrett-Byrne, et al. 2021a), *Gallus gallus domesticus* (Labas, et al. 2015, Vitorino Carvalho, et al. 2021), *Rattus norvegicus* (Bayram, et al. 2016) and *Bubalus bubalis* (Batra, et al. 2021, Fu, et al. 2019) (Fig. 1A, Table S1). These datasets were reanalysed using a stringent and uniform pipeline implemented in Proteome Discover, and the resultant protein IDs are provided in Tables S2-13. To enhance accessibility, all data is also available on

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

ShinySpermKingdom, facilitating an interactive experience with these complex datasets (https://reproproteomics.shinyapps.io/ShinySpermKingdom/).

Unsurprisingly, human (9,296 proteins) and mouse (8,645 proteins) exhibited the most comprehensive sperm proteomes (Fig. 1B, Table 2), reflecting their status as well researched species. Returning nearly a third of the larger proteomes was the boar (3,298), closely followed by bull (3,177), crocodile (2,855), and rabbit (1,650). The proteome of each species was first assessed using UniProt to determine their current curated level of protein evidence (Fig. 1C; Tables S2-13). Predictably, the sperm proteomes of well-characterized species like human (91.4%), mouse (94.4%) and rat (83.1%) were all well annotated at protein and transcript level. Interestingly, buffalo sperm harboured 77.3% of its evidence at the transcript level. Within the sperm proteomes of the remaining 8 species, in most cases >90% of protein identifications were only predicted or inferred from homology, indicating that experimental evidence for the existence of most proteins remains poor in non-traditional model species (Fig. 1C). Due to the low number of protein identifications in the chicken (223 IDs), Norwegian rat (89 IDs) and buffalo (84 IDs), these species were excluded from further downstream analyses (Fig. 1B). These exclusions ensured a focus on datasets with sufficient coverage and quality for robust comparative analysis.

Humanized sperm proteomes redefine evolutionary links

To advance the understanding of the remaining 9 species, each sperm proteome was converted to their respective human homologues to allow utilization of human focused bioinformatic tools, facilitating a standardized cross species analysis. Humanization was achieved using the OmicsBox software as previously described.(Skerrett-Byrne, et al. 2021a) Conversion rates to humanized IDs were exceptionally high, ranging from 95.9% – 99.1%, producing a total of 13,853 proteins (Fig. 2A, Table 2, Table S14). These newly generated sperm proteomes were subject to analyses using

Ingenuity Pathway Analysis (IPA) to provide an overall classification of protein types present. Notably, despite variations in proteome size, proportional compositional analysis of each species revealed broad consistency (Fig. 2B). Enzymes were the dominated category, accounting for ~72.5% of all sperm proteins, followed by transporters (~15.5%), transcription (~5.3%) and translation (~2.7%) regulators, receptors (~1.6%), ion channels (~1.6%), cytokines (~0.6%) and growth factors (~0.5%).Whilst there were no overt differences in the proportional composition of protein classification types, the two largest sperm proteomes (i.e., human and mouse), featured proportionally more transcription regulators than that of all other species assessed (i.e., ~9.3% vs ~4.1%). This enrichment may indicate potential differences in sperm-specific regulatory mechanisms between traditional model organisms and less-studied species, or reflects poorer annotation in these latter species. Further interrogation with UniProt subcellular localization and the Human Protein Atlas(Uhlén, et al. 2015) sperm subcellular resource allowed mapping to key sperm locations (Fig. S1): acrosome (146 proteins), perinuclear theca (17 proteins), nucleus (3,879 proteins), calyx (15 proteins), equatorial segment (13 proteins), connecting piece (45 proteins), mid-piece (118 proteins), mitochondria (394 proteins), annulus (16 proteins), flagellum (204 proteins), flagellar centriole (21 proteins), axoneme (176 proteins), principal piece (91 proteins), and end piece (31 proteins) (Table S14).

Seeking to delve further into the functional relationships among these sperm proteomes, the canonical pathways node of IPA was utilized leading to the identification of 482 unique pathways significantly enriched in at least one species *(p*-value ≤ 0.05). Unbiased hierarchical clustering based on the conservation of these pathways revealed mouse spermatozoa to be the most functionally related to that of their human counterparts (Fig. 2C). This finding contrasts that of the genomic lineage tracing, which indicated that among the species assessed, rabbits were the closest

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

evolutionary relative to humans. Notably, this hierarchical clustering approach achieved the division of the assessed species into two broad groupings; 1) crocodile, dolphin, bull and boar; 2) sheep, koala, rabbit, mouse and human. This reorganization suggests that functional relationships based on sperm proteomes may not strictly align with evolutionary distances derived from genomic data. Amongst the most significantly enriched pathways across all species were those related to energy metabolism ("Mitochondrial Dysfunction", "Oxidative Phosphorylation", "Glycolysis", "Gluconeogenesis", and "Fatty Acid β-oxidation") and capacitation ("Sirtuin Signalling Pathway", "Protein Ubiquitination Pathway").

Characterization of the core sperm proteome

To ascertain the proteins most fundamental to the functional competency of spermatozoa across the 9 assessed species, a comparison of all protein identifications was conducted. This strategy uncovered a modest 45 and 135 conserved proteins at the taxonomic level of species and orders, respectively, which we hereafter refer to as the core sperm proteome (Fig. 3A, Table S15). The 9 species were collapsed into six taxonomic orders, namely Primate (human; 9,186 proteins), Rodentia (mouse; 6,707 proteins), Artiodactyl (boar, bull, dolphin and sheep; 3,795 proteins), Lagomorpha (rabbit; 1,548 proteins), Diprotodontia (koala; 556 proteins), and Crocodilia (crocodile; 2,103 proteins). Of those proteins that were identified in more than one order, the largest overlaps were between Primate and Rodentia (830), Primate, Rodentia and Artiodactyla (489), and Rodentia and Artiodactyla (457) (Fig. 3A).

Further interrogation of these conserved proteins with the UniProt Alignment tool (Clustal Omega(Sievers, et al. 2011)) revealed many proteins with known roles in sperm motility, mitochondria function, capacitation, and acrosome reaction have high levels of sequence similarity (Fig. 3B–D); heat shock protein family A member 2 (HSPA2; 98.9%), protein kinase cAMP-

dependent type I regulatory subunit alpha (PRKAR1A; 97.4%), and voltage dependent anion channel 3 (VDAC3; 96.8%). However, a notable divergence was observed for human VDAC3, which differs by ~20% compared to the other 8 species. Conversely, proteins critical for sperm motility, zona pellucida binding, and penetration exhibited greater variability in sequence conservation (Fig. 3E–G); acrosin (ACR, 65.1%), calcium binding tyrosine phosphorylation regulated (CABYR; 64.8%), and the zona pellucida binding protein (ZPBP; 82.3%).

Both species and order lists were subjected to analysis with IPA, focusing on molecular functions and pathways. Strong consistency between both groups was observed, with the significant enrichment (p-value ≤ 0.05) of key reproductive processes, including synthesis of ATP, movement of cilia, acrosome reaction, binding of sperm and zona pellucida, and fertilization (Fig. 3H, Table S16). Pathway analysis of the core sperm proteomes displayed significant enrichment of pathways involved in proteostasis, metabolism, and oxidative stress (Fig. 3I, Table S16). The most significant enriched pathways were Bcl2-associated athanogene 2 (BAG2) and Ubiquitin-like protein FAT10 signalling. Complementary STRING analysis identified distinct protein interaction networks (Fig. S2), with clusters of proteins being readily detected associated with chaperone functions, the proteasome, ribosome function, metabolism, sperm morphogenesis and zona pellucida binding. Further DAVID analysis of the core sperm proteome revealed significant enrichment of similar annotation clusters, including chaperone functions, the proteasome, ribosome function, glycolysis, the TCA cycle and flagella. Additional enriched annotation clusters included secretory granules, mitochondrial function and ATP binding. These findings collectively underscore the critical roles of these conserved proteins in ensuring sperm functionality across diverse species.

Knockout mouse models confirm conserved proteins affect sperm fertilization competency

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

To investigate the functional relevance of these 135 conserved sperm proteins, we leveraged resources from The International Mouse Phenotyping Consortium (IMPC) database(Dickinson, et al. 2016, Groza, et al. 2022) and the European Mouse Mutant Archive (EMMA)(Hagn, et al. 2007) to obtain knockout (KO) models of these protein-coding genes of interest. This effort yielded five candidates: *Aldh7a1* (Aldehyde dehydrogenase 7 family member A1), *Echs1* (Enoyl-CoA hydratase, short chain 1), *Etfb* (Electron transfer flavoprotein subunit beta), *Ndufa10* (NADH:ubiquinone oxidoreductase subunit A10), and *Pebp4* (Phosphatidylethanolamine binding protein 4). We first established the evolutionary conservation of these proteins across species (Fig. 4A), yielding relatively high conservation between species: Aldh7a1 (median 86.8%), Echs1 (83.4%), Etfb (89.3%), Ndufa10 (77.5%), and Pebp4 (52.3%). Notably, species-specific variations were observed for Aldh7a1 (rabbit) and Pebp4 (crocodile).

Through EMMA, we obtained unpublished *in-vitro* fertilization (IVF) data pertaining to heterozygous and homozygous KO mice (Fig. 4). Notably, *Echs1*, *Etfb* and *Ndufa10* are homozygous lethal, and as such, data presented for these genes are from heterozygous males. Total sperm motility analysis of male KO mice showed marked reductions for *Aldh7a1* (27.7% decrease) and *Etfb* (21.7%), compared to wildtype (Fig. 4B). Forward progressive motility was further impaired, with proportional reductions of 36.5% (*Etfb*), 32.7% (*Aldh7a1*), and 25% (*Ndufa10*) (Fig. 4B). Milder decreases were observed for *Pebp4* (13.5% loss) and *Echs1* (9.6% loss). Next, using sperm from these KO males, we performed IVF to evaluate their fertilization potential. The first point of examination was the cleavage rate to the 2-cell stage (Fig. 4 C), which mirrored motility trends with marked reductions: *Aldh7a1* (29%), *Etfb* (31.2%), and *Ndufa10* (32.3%). Rates for *Pebp4* and *Echs1* remained within expected values. Blastocyst formation rates revealed severe effects, particularly for KOs of *Ndufa10* and *Etfb* with success rates of 10% and 29% respectively

(Fig. 4C). Almost halving of success was observed for *Pebp4* (53%), whilst *Aldh7a1* (15%) and *Echs1* (25%) were the least impacted. Pregnancy rates were low for all KOs compared to the expected rate of WTs, with *Etfb* being the most greatly affected (33%), followed by *Echs1* (56%) and remaining KOs ranging from 61% to 68%. Where pregnancy was achieved, litter sizes were within range of expected of age matched wildtype controls (Fig. S3A) with no significant shifts in foetal sex distribution (Fig. S3B).

In complement to the IVF studies, the histology of the male reproductive tract was examined to provide a detailed understanding of its structure and function at microscopic level, and explore any potential contributing effects played by these pivotal tissues. For each KO mouse line, hematoxylin and eosin (H&E) staining was carried out on sections of the testis and epididymis (spermatogenesis (Hermo, et al. 2010a, b) & sperm maturation (Nixon, et al. 2020)), as well as the prostate gland and seminal vesicles (major contributors to the seminal plasma (Robert and Gagnon 1994, Schjenken, et al. 2018)) at 16 weeks of age (Fig. 4D, S4). The *Pebp4* KO displayed the most pronounced abnormalities, including significant presence of multinucleated giant cells (MGC), (symplasts), associated with mild multifocal seminiferous tubular degeneration (Creasy, et al. 2012), and presence of sloughed germ cells in the epididymal caudal tubules, a good indicator of spermatogenic disruption in the testis (De Grava Kempinas and Klinefelter 2014) (Fig. 4D, S4). A similar phenotype of higher number of MGCs was found in *Ndufa10* KOs (Fig.s 4D, S4), but did not achieve statistical significance (p-value = 0.054), and was accompanied by mild focal vacuolation. Across the five KOs, there were no histopathological changes observed for the prostate or seminal vesicles (Fig. S4), with the exception of *Ndufa10* with hyperplasia in the anterior prostate (or coagulating gland) noted (Fig. S4).

Conserved proteins loss of function variants linked to human sperm defects

 Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Seeking to further support the clinical relevance of these core sperm proteins, we interrogated our access to nearly 2,900 exomes and genomes of men with quantitative and/or qualitative sperm defects from the MERGE cohort (Stallmeyer, et al. 2024). This analysis returned two men with homo- or hemizygous loss of function (LoF) variants: subject M2218 is homozygous for a stop-gain variant in parkin coregulated: *PACRG*; NM_152410.2:c.369T>A p.(Tyr123Ter). He repeatedly had normal sperm counts but 99-100% sperm head defects and significantly impaired motility. Subject M3692 is homozygous for a frameshift variant in dynein axonemal light intermediate chain 1: *DNALI1*; NM_003462.5:c.490dup p.(Tyr164LeufsTer20). He repeatedly had normal sperm counts but almost all sperm were immotile.

Characterization of sperm proteins solely identified in different species

The total number of protein IDs for each species was strongly correlated to the number of "unique" (detectable) proteins in that species' proteome (r_s = 0.88, p = 0.002). While a considerable proportion of proteins in the human (66.3%) and mouse (30.1%) sperm proteomes were only identified in these species, all other species (with much more poorly characterized proteomes) contained <6% proteins not identified in any other species (Table 2, Table S17).

DAVID analysis of unique-to-species proteins highlighted significantly enriched annotation clusters in some species (Table S17). In mice, a range of enriched clusters were observed, including those involved in transcription (RNA splicing (enrichment score (ES) 18.8), small non-coding RNA processing (ES 5.2), RNA helicases (ES 4.2)) and translation (ribosome/mitochondrial translation (ES 8.8). In both cattle and the koala, there was enrichment for secreted proteins (ES 4.3, 3.5 respectively), however the proteins within these clusters were species-specific. Secreted proteins only identified in the cattle sperm proteome included beta defensins (DEFB108B, DEFB116), cytokines (IL12B, IL34, CCL2, CTF1), proteins with

antimicrobial activity (VIP, ADM) and RNase 1. In contrast, the enriched cluster of secreted proteins only identified in the koala sperm proteome largely contained pro-hormones (AVP, RLN2, PTH, INSL5, POMC). Proteins uniquely identified in dolphin sperm showed weak but significant enrichment of functions associated with cell and gonadal differentiation (ES 2.6), including AMH, TSPY2 and TSPY8. There were no significantly enriched annotation clusters in unique-to-species proteins in the boar, rabbit, sheep or crocodile sperm proteomes.

As described above, depending on the species, 0.9 – 4.1% of protein identifications were not successfully converted to humanized IDs for further analysis. While many of these were poorly characterized proteins that may indeed have human equivalents, it is likely that some of these represent additional unique-to-species proteins. Examples of such proteins that are thought to be specific to one or several closely related species (albeit most with homologues in more diverse species) include mouse seminal vesicle proteins (SVS3A, SVS3B, SVS4, SVS5, SVS6)(Karn, et al. 2008), boar carbohydrate-binding protein AQN-1(Kraus, et al. 2005), bovine spermadhesin Z13 (Haase, et al. 2005), and rabbit semen coagulum protein (SVP200) (Lundwall, et al. 2020).

Reproductive strategies reflected in sperm proteomes

In seeking to investigate the influence of evolutionary and imposed reproductive strategies on sperm protein composition, we further interrogated each species proteome by focusing on sperm metabolism preference (glycolysis preference vs no preference), location of the testes (internal vs external), and history of selective breeding (yes vs no). Species were stratified into their appropriate groups and Venn Diagram analyses were conducted to determine the proteins unique to each biological context. Firstly, focusing on sperm metabolism, a comparison of species with a preference for glycolytic sperm metabolism against those with no preference between glycolysis and oxidative phosphorylation, resulted in a 49.7% shared overlap of proteins (1,857) (Fig. 5A).

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

The larger component of this comparison was those species with no sperm metabolism preference (cattle, rabbit, sheep), with 1,385 unique proteins. Only species that preferentially use glycolysis showed enrichment for a variety of metabolically relevant pathways, including degradation of amino acids (valine, isoleucine, tryptophan, leucine, cysteine, phenylalanine, alanine, glutamine) and N-acetylglucosamine (Table S18).

A subsequent comparison of sperm proteomes based on species with internal vs external testes revealed 41.5% of proteins (1,825) were shared, with only 13.1% (574) proteins unique to species with internal testes (crocodile, dolphin) (Fig. 5B). Species with external testes showed significant enrichment for signalling involving a range of interleukins (e.g. IL-3, -13, -17), reactive oxygen species production and NRF2-mediated oxidative stress response pathways, as well as clathrin mediated endocytosis signalling, adherens junction signalling, HIF1α signalling, acute phase response signalling, and integrin signalling. By comparison, species with internal testes showed significant enrichment for pathways involved in glycine and cholesterol synthesis (Table S18).

Lastly, species were compared based on whether they have a history of selective breeding; a strategy that revealed 42.9% (1,853) proteins were shared regardless of breeding practices (Fig. 5C). However, species under selective breeding pressures (cattle, boar, rabbit, sheep) yielded the greater number of unique proteins amounting to a total of 1,853 proteins (42.1%). Conversely, sperm from those species without selective pressures were associated with the unique expression of 658 proteins (8.2%). Selectively bred species showed higher enrichment for endothelin-1, EIF2 and CDK5 signalling. In comparison, species with no history of selective breeding showed higher enrichment for adherens junction signalling (Table S18). Unique and overlapping protein IDs for each of the 3 biological contexts comparisons are supplied in Table S18.

DISCUSSION

From a total of 29 datasets across 12 vertebrate species (>2TB of RAW data), we have generated the most comprehensive cross species analysis of the mature sperm proteome to date, identifying a grand total of 13,853 unique proteins residing in the sperm of those species studied herein. Within this dataset, we discerned a core sperm proteome of 45 proteins conserved at the species level and 135 proteins conserved at the order level, underscoring a fundamental molecular framework essential for generating fertilization-competent spermatozoa. Enrichment analyses linked these core proteins to critical pathways in proteostasis and sperm metabolism, while knockout mouse models of selected conserved proteins confirmed their influential roles in governing sperm motility, fertilization success, and overall reproductive health. Together, our findings demonstrate that despite the remarkable diversity in reproductive strategies observed across vertebrates, including differences in sperm metabolic preferences, testicular location, and histories of selective breeding, a fundamental set of molecular components remain steadfastly conserved. This universal baseline provides a platform upon which species-specific adaptations are layered, shaping the unique sperm proteomes that ultimately define each organism's reproductive biology. Complementing the full breadth of analyses discussed here, we have deployed a Shiny application, ShinySpermKingdom (https://reproproteomics.shinyapps.io/ShinySpermKingdom/), to support the accessibility and interpretability of these datasets beyond static documents, allowing for effective data-driven insights by researchers in the field (Skerrett-Byrne, et al. 2024).

Proteomics technologies in reproductive biology have exploded more than 20-fold increase in 20 years, paralleling advancements in mass spectrometry (MS) and bioinformatic pipelines. (Skerrett-Byrne, et al. 2024) As a field, proteomics is reaching an intriguing point, as these RAW datasets can be viewed as '*living datasets*', from which we can continuously yield new insights as

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

bioinformatic tools, database annotations, and computational capacities advance.(Dai, et al. 2024, Drew, et al. 2017) Whilst there are several variables affecting their '*mine-ability*', early analyses often only scratch the surface of what these complex spectra contain; proteins may remain unidentified simply because the necessary reference sequences or annotation frameworks are inadequately annotated at the time.(Willems, et al. 2020) Through these advancements, the continuous reanalysis of MS datasets is not merely a luxury, but a crucial endeavour to fully leverage these rich repositories of biological information, ensuring that the scientific community continues to extract meaningful knowledge from the collective body of proteomic data. In complement to this, large-scale proteomic atlases, such as the human and mouse draft proteomes,(Giansanti, et al. 2022, Wang, et al. 2019) provide a systematic framework for characterising could be adapted to non-model species, ultimately capturing the full diversity of sperm proteomes across taxa.

Of the >2,000 sperm proteome studies found on PubMed,(Skerrett-Byrne, et al. 2024) a mere 218 studies have data deposited in a publicly available repository, a poor return of ~10%. Until now, to our knowledge there have yet to be any studies reanalysing the RAW MS data of previous studies focused on mature sperm cells, here, we have demonstrated the capacity to assemble proteomic profiles for 12 taxa, achieving an unprecedented depth of coverage and unveiling numerous unique and conserved proteins. Comparing the proteomic depths of the studies included to generate these new sperm proteomes, we demonstrate improved depths in rabbit (1,360 (Juárez, et al. 2020), to1,605; 18% increase) and crocodile (1,119 (Nixon, et al. 2019), to 2,855; 155% increase). Despite these strides, the proteomic annotation of non-model species still lags behind that of humans and established model organisms, highlighted best by the discrepancy between human and koala sperm proteome, with the latter being only 6.3% the size of the former.

Consequently, the wealth of newly identified proteins reported here, while indicative of greater analytical depth, is still constrained by the uneven quality of reference data, highlighting the need for greater annotation placed upon non-model species in NCBI and UniProt databases, and the pursuit of new analytical tools.(Heck and Neely 2020, Van den Broeck, et al. 2023)

Interestingly, even beyond direct protein-to-protein comparisons, our functional analyses highlight complex patterns of convergence and divergence that transcend traditional phylogenetic boundaries. For instance, despite their substantial evolutionary distance, koalas and sheep emerge as functionally aligned in terms of pathways underpinning their sperm proteomes, underscoring the powerful influence of reproductive strategies and physiological demands on proteome composition. Conversely, we observed species-specific pathways enrichment, with crocodiles displaying the great enrichment of fatty acid β-oxidation, an important source of long-term sustained energy as sperm ascends several meters of the female reproductive tract (Gist, et al. 2008, Nixon, et al. 2019). These cross-species comparisons highlight both the remarkable flexibility and deep conservation of sperm biology.

The identification of a core sperm proteome comprising 45 proteins at the species level and 135 at the order level underscores the presence of a conserved molecular foundation essential for sperm function across evolutionary lineages. Notably, given the size of the smallest proteome, the koala (556 proteins), these conserved proteins represent a 24.3% level of conservation at the order level, highlighting the functional importance of these proteins despite proteome size variation. Focusing on the sequence conservation of these core sperm proteins, we observed a median of 94.4% conservation at the amino acid level. Proteins such as heat shock protein A2 (HSPA2) and voltage dependent anion channel 3 (VDAC3) maintain exceptionally high sequence conservation, reflecting their critical and broadly indispensable roles in processes like protein folding, sperm

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

maturation, and capacitation.(Arcelay, et al. 2008, Li, et al. 2023, Nixon, et al. 2017, Redgrove, et al. 2012, Sampson, et al. 2001) By contrast, proteins critical for zona pellucida binding and penetration, acrosin (ACR)(Dudkiewicz 1983, Hua, et al. 2023, Liu and Gordon Baker 1993), zona pellucida-binding protein (ZPBP),(Dun, et al. 2010, Lin, et al. 2007) and calcium-binding tyrosine-phosphorylation regulated protein (CABYR),(Naaby-Hansen, et al. 2002, Skerrett-Byrne, et al. 2022) exhibit greater sequence variability, perhaps signifying adaptations to specific reproductive environments or selective pressures. Notably, expected proteins such as histones and protamines, were not highly conserved across all species, with about 16 histone proteins retained across at least 4/9 species, with testis-specific Histone H2B type 1-A (H2BC1) (Montellier, et al. 2013) detected in 6/9 species. Once more, pointing to the need for greater annotation of non-model species.

This is supported by previous work which sought to compare the sperm proteome within three closely related *Mus* species that experience different level of sexual selection(Vicens, et al. 2017). Whilst not a defined difference in protein sequence, this work highlighted significant interspecific protein abundance divergence of proteins which govern sperm–egg interactions, including ACR, ZPBPs, and conversely no significant shifts in HSPA2 or protein kinase cAMP-dependent type I regulatory subunit alpha (PRKAR1A). Seeking to understand how these core sperm proteins may interact, *in-silico* enrichment analyses grouped them into known networks integral to sperm motility (flagellar motility), energy source of the sperm cell (oxidative phosphorylation [OXPHOS]), and fertilization, encompassing zona pellucida binding and acrosome reaction. A key study from 2016, included in our work, carried out proteomics on sperm collected from 19 placental mammalian species, identified a core sperm proteome of 623 proteins (Bayram, et al. 2016). Importantly, this study leverages two advantages: 1) comparatively narrower range of species, ungulates and rodents; 2) all proteomics sample preparation, MS analysis and

data processing are uniform. However, there is considerable overlap with our study in regards to the significantly enriched pathways and functions which connect the core sperm proteome, namely metabolic processes such as OXPHOS, glycolysis and the tricarboxylic acid cycle (TCA), acrosome assembly, zone pellucida binding and proteasome function (Bayram, et al. 2016). Notably, the emergence of the highly enriched proteostasis-associated pathways, involving proteins like BAG2 and FAT10, provides intriguing hints of previously underappreciated quality control processes that may safeguard sperm function by ensuring proper protein folding, timely degradation, regulation of capacitation and acrosomal reaction, thereby facilitating successful fertilization.(Cafe, et al. 2021, Smyth, et al. 2024) For instance, the enrichment of BAG2 in the conserved sperm proteome, known to modulate HSPA2 in spermatogenesis,(Yin, et al. 2020) raises the potential of a post-testicular role in maintaining HSPA2 functionality in facilitating sperm-egg adhesion and binding. (Nixon, et al. 2015, Smyth, et al. 2024)

Seeking to build upon this important database and demonstrate how these analyses can help build new knowledge on sperm biology, we were granted special access to the European Mouse Mutant Archive (EMMA)(Hagn, et al. 2007) to obtain sperm and *in vitro* fertilization (IVF) data, from knockout (KO) mouse models targeting selected proteins from the core proteome. We elected to focus on proteins which had not previously been implicated in sperm morphology or functional maturation; metabolic enzyme Aldh7a1 (Aldehyde dehydrogenase 7 family member A1)(Korasick and Tanner 2021); mitochondrial enzymes Echs1 (Enoyl-CoA hydratase, short chain 1)(Burgin and McKenzie 2020) and Etfb (Electron transfer flavoprotein subunit beta)(Henriques, et al. 2021); mitochondrial complex I subunit Ndufa10 (NADH:ubiquinone oxidoreductase subunit A10) (Formosa, et al. 2018); and a member of the phosphatidylethanolamine-binding proteins, Pebp4, although detected in bull sperm and semen previously,(An, et al. 2012, Somashekar, et al. 2017)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

no studies have demonstrated its role in sperm function. Here for the first time, KOs of these five core genes have been shown to directly influence sperm quality, motility, and fertilization potential. Intriguingly, the largest decreases observed in total sperm motility and progressive motility were associated with the four proteins located in the mitochondrial matrix, each of which play key roles in mitochondrial function. Given the pivotal role mitochondria play in the energy required to drive sperm motility(Piomboni, et al. 2012), it lends to reason these proteins may have regulatory roles in ensuring efficient OXPHOS/ATP production in sperm mitochondria. Further to the KO mouse work, we sought to investigate the clinical relevance of these proteins in the context of human infertility. Utilising the MERGE cohort, we queried our 135 conserved sperm proteins against the almost 2,900 males with quantitative and/or qualitative sperm defects, identifying two men with homozygous loss-of-function variants in *PACRG* and *DNALI1*. Notably, a recent mouse KO study demonstrated that PACRG interacts within a manchette-associated complex, and is essential for proper sperm assembly (Yap, et al. 2023). Moreover, KO mice exhibited a significant reduction in sperm count, with the remaining sperm characterized by abnormally shaped heads and bent tails. DNALI1 is a component of the inner dynein arms and in men affected by biallelic pathogenic variants, immotile sperm exhibiting an asymmetric fibrous sheath of the flagella have been described (Sha, et al. 2022). Together, these experimental models lend strong empirical support to the notion that the core sperm proteome is not merely a historical residue of evolutionary conservation, but rather a functional blueprint vital for the generation and maintenance of fertilization-competent sperm across species.

Beyond their universally conserved core, sperm proteomes also reflect diverse biological contexts that shape reproductive strategies and outcomes. For instance, contrasting testicular environments, as represented by species with either external or internal testes, impart distinct

proteomic signatures on mature sperm. Species with external testes exhibit enrichments in signalling pathways associated with interleukin-mediated communication, reactive oxygen species management, and NRF2-mediated oxidative stress responses, as well as pathways related to cell-cell interaction and endocytic processes. These findings suggest that externalized gonads, potentially evolving under selective pressures linked to temperature regulation or hypoxic conditions, have prompted the refinement of sperm's molecular toolkit to bolster resilience and maintain fertilisation capacity. Conversely, species with internal testes show a bias toward pathways involving glycine and cholesterol synthesis, hinting at metabolic adjustments that support sperm function in a more thermally stable yet potentially resource-limited environment.

Selective historical breeding regimes and unique metabolic preferences leave distinguishable marks on the sperm proteome. For instance microRNA biogenesis and endothelin-1 signalling are enriched in domesticated species, reflecting human-driven selection for traits impacting sperm maturation, and early embryonic development through epigenetic and signalling mechanisms.(Conine and Rando 2022, Sharma, et al. 2018, Spadafora 2023, Tomar, et al. 2024, Trigg, et al. 2021) In contrast, non-selected species' enrichment for adherens junction signalling points to a more basal state of sperm cell-cell interaction and membrane integrity. (Lui, et al. 2003, Wen, et al. 2016) Additionally, the preferential use of glycolysis as an energy source(du Plessis, et al. 2015) in certain taxa is mirrored by an increased representation of amino acid degradation and N-acetylglucosamine metabolism pathways, underscoring how ecological pressures shape sperm energetics. In essence, the sperm proteome reflects not only a shared, evolutionarily ancient blueprint but also the distinct biological contexts that steer reproductive strategies and outcomes in varied environmental and evolutionary landscapes.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Despite the breadth and depth of the current dataset, several limitations constrain the interpretation and generalisability of our findings. Foremost, the use of publicly available proteomic data, generated over a span of years and employing diverse sample preparation protocols, mass spectrometry platforms, and analytical pipelines, precludes direct, quantitative '*apples to apples*' comparisons between species. Although advancements discussed, many datasets remain hampered by incomplete metadata, and limited raw data availability hinder reproducibility and the clarity of species-to-species comparisons. Additionally, our reliance on human orthologues for pathway analyses, inevitably restricts our ability to fully explore lineage-specific adaptations or signatures of selective pressure acting on particular sperm proteins. Similarly, we have focused exclusively on chordate species, thus excluding the vast diversity of reproductive strategies found in invertebrates and other non-chordate taxa. Moreover, the marked disparity in proteome sizes across species highlights that coverage bias remains a significant concern, but regardless our current dataset offers a glimpse of the full sperm proteome complexity. Recognizing these acquisition biases paves the way for new funding proposals aimed at refining reference annotations and expanding high-quality MS coverage across diverse species. By systematically addressing these limitations, the field can move closer to an accurate and comprehensive understanding of sperm proteomes at a global scale. Until then, our findings, while important and innovative, should be viewed as a framework upon which to build as the global sperm proteome landscape becomes more thoroughly mapped and understood.

Future investigations will benefit from a more controlled and integrated experimental framework, enabling true comparisons across species. Standardising sample collection protocols, employing uniform mass spectrometry methodologies, and harmonising data processing pipelines will help overcome current disparities in data quality and annotation. As proteome databases are

further refined, and as more non-model species achieve higher-quality genome assemblies, future analyses will be better positioned to distinguish genuine species-specific proteins from those simply missing in incomplete datasets. Additionally, exploring the role of post-translational modifications (PTMs), such as phosphorylation, acetylation, and glycosylation, across multiple taxa will be critical for understanding how subtle regulatory mechanisms modulate sperm functionality. Indeed, since sperm protein composition is largely defined during epididymal maturation,(Skerrett-Byrne, et al. 2022) incorporating temporal and spatial sampling strategies alongside PTM profiling may reveal how species adapt their sperm at the molecular level to an array of environmental pressures.

Ultimately, comprehensive multispecies studies employing rigorous and consistent workflows have the potential to create living, expandable proteomic databases that will evolve as analytical capabilities and reference annotations improve. By repeatedly revisiting and reanalysing mass spectrometry datasets, the field can unlock layers of complexity that have, until now, remained concealed. Such iterative proteomic approaches, informed by emerging knowledge of sperm physiology and integrated with genomic, transcriptomic, and epigenomic data, will open new avenues for understanding the evolutionary and functional contexts of sperm proteomes. Here, we demonstrate the power of systematically revisiting and reanalysing proteomic data, identifying a set of 135 conserved proteins critical to sperm function, including several that have never been previously implicated in sperm biology. This study underscores the importance of continuous proteomic refinement, allowing for the identification of previously overlooked but evolutionally and functionally essential proteins. By embracing this approach, future research to not only refine our grasp of the universal underpinnings of sperm biology but also reveal how species-specific adaptations arise and shape reproductive success across diverse taxa.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

DECLARATION OF INTERESTS

The authors declare no competing interests.

FUNDING

This research was supported by a National Health and Medical Research Council of Australia (NHMRC) Emerging Leadership Fellowship (APP2034392) and a College of Engineering, Science & Environment (University of Newcastle) Accelerator Fellowship, both awarded to D.A.S.B.. Additionally, F.T. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Clinical Research Unit 'Male Germ Cells' (CRU326, project number 329621271). F.T. and S.K. were supported by the German Federal Ministry for Education and Research (BMBF) as part of the Junior Scientist Research Centre 'ReproTrack.MS' (grant 01GR2303).

AUTHOR CONTRIBUTIONS

Conceptualisation, T.P., B.N., and D.A.S.B.; Methodology, T.P. and D.A.S.B.; Software, D.A.S.B.; Investigation, T.P., B.N., T.K., R.T., A.S.M., P.S.B., F.T., T.S., S.K., V.G.D., H.F., S.M., M.H.A., and D.A.S.B.; Formal Analysis D.A.S.B.; Validation R.T., A.S.M., P.S.B., V.G.D., H.F., S.M., M.H.A., and D.A.S.B; Visualisation, T.P., and D.A.S.B; Writing – Original Draft, T.P., and D.A.S.B; Writing – Review & Editing, B.N., T.K., R.T., A.S.M., P.S.B., F.T., T.S., S.K., V.G.D., H.F., S.M., M.H.A.; Funding Acquisition, B.N. and D.A.S.B; Resources, B.N. R.T., F.T., T.S., S.K., V.G.D., H.F., S.M., M.H.A., and D.A.S.B.; Supervision, B.N. and D.A.S.B.

ACKNOWLEDGEMENTS

We thank the Academic and Research Computing Support team, The University of Newcastle who provided High Performance Computing Infrastructure to support the bioinformatics analyses. We also thank David MacKenzie for their graphic design input and support. We thank, S. Dunst and B. Rey, for technical support with the sperm and IVF culture experiments. We thank the technicians and animal caretakers of the German Mouse Clinic. We thank Dr Wei Zhou for his insightful questions at The Society for Reproductive Biology Annual Scientific Meeting.

 Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

REFERENCES

An, L-P, T Maeda, T Sakaue, K Takeuchi, T Yamane, P-G Du, I Ohkubo, and H Ogita 2012 Purification, molecular cloning and functional characterization of swine phosphatidylethanolamine-binding protein 4 from seminal plasma. *Biochemical and Biophysical Research Communications* 423 690-696.

Arcelay, E, AM Salicioni, E Wertheimer, and PE Visconti 2008 Identification of proteins undergoing tyrosine phosphorylation during mouse sperm capacitation. *International Journal of Developmental Biology* 52.

Bardou, P, J Mariette, F Escudié, C Djemiel, and C Klopp 2014 jvenn: an interactive Venn diagram viewer. *BMC bioinformatics* 15 1-7.

Batra, V, V Bhushan, SA Ali, P Sarwalia, A Pal, S Karanwal, S Solanki, A Kumaresan, R Kumar, and TK Datta 2021 Buffalo sperm surface proteome profiling reveals an intricate relationship between innate immunity and reproduction. *BMC Genomics* 22 480.

Bayram, HL, AJ Claydon, PJ Brownridge, JL Hurst, A Mileham, P Stockley, RJ Beynon, and DE Hammond 2016 Cross-species proteomics in analysis of mammalian sperm proteins. *J Proteomics* 135 38-50.

Buchfink, B, K Reuter, and H-G Drost 2021 Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature methods* 18 366-368.

Burgin, HJ, and M McKenzie 2020 Understanding the role of OXPHOS dysfunction in the pathogenesis of ECHS1 deficiency. *FEBS Letters* 594 590-610.

Byrne, K, T Leahy, R McCulloch, ML Colgrave, and MK Holland 2012 Comprehensive mapping of the bull sperm surface proteome. *Proteomics* 12 3559-3579.

Cafe, SL, B Nixon, H Ecroyd, JH Martin, DA Skerrett-Byrne, and EG Bromfield 2021 Proteostasis in the Male and Female Germline: A New Outlook on the Maintenance of Reproductive Health. *Front Cell Dev Biol* 9 660626.

Casares-Crespo, L, P Fernández-Serrano, and MP Viudes-de-Castro 2019 Proteomic characterization of rabbit (Oryctolagus cuniculus) sperm from two different genotypes. *Theriogenology* 128 140-148.

Castaneda, JM, R Hua, H Miyata, A Oji, Y Guo, Y Cheng, T Zhou, X Guo, Y Cui, B Shen, Z Wang, Z Hu, Z Zhou, J Sha, R Prunskaite-Hyyrylainen, Z Yu, R Ramirez-Solis, M Ikawa, MM Matzuk, and M Liu 2017 TCTE1 is a conserved component of the dynein regulatory complex and is required for motility and metabolism in mouse spermatozoa. *Proc Natl Acad Sci U S A* 114 E5370-e5378.

Castillo, J, OA Bogle, M Jodar, F Torabi, D Delgado-Dueñas, JM Estanyol, JL Ballescà, D Miller, and R Oliva 2019 Proteomic Changes in Human Sperm During Sequential in vitro Capacitation and Acrosome Reaction. *Front Cell Dev Biol* 7 295.

Chen, T, J Ma, Y Liu, Z Chen, N Xiao, Y Lu, Y Fu, C Yang, M Li, S Wu, X Wang, D Li, F He, H Hermjakob, and Y Zhu 2022 iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res* 50 D1522-d1527.

Choi, M, J Carver, C Chiva, M Tzouros, T Huang, T-H Tsai, B Pullman, OM Bernhardt, R Hüttenhain, and GC Teo 2020 MassIVE. quant: a community resource of quantitative mass spectrometry–based proteomics datasets. *Nature methods* 17 981-984.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Comizzoli, P, and WV Holt 2019 Breakthroughs and new horizons in reproductive biology of rare and endangered animal species. *Biol Reprod* 101 514-525.

Conine, CC, and OJ Rando 2022 Soma-to-germline RNA communication. *Nature Reviews Genetics* 23 73-88.

Creasy, D, A Bube, Ed Rijk, H Kandori, M Kuwahara, R Masson, T Nolte, R Reams, K Regan, S Rehm, P Rogerson, and K Whitney 2012 Proliferative and Nonproliferative Lesions of the Rat and Mouse Male Reproductive System. *Toxicologic Pathology* 40 40S-121S.

Dai, C, J Pfeuffer, H Wang, P Zheng, L Käll, T Sachsenberg, V Demichev, M Bai, O Kohlbacher, and Y Perez-Riverol 2024 quantms: a cloud-based pipeline for quantitative proteomics enables the reanalysis of public proteomics data. *Nat Methods* 21 1603-1607.

De Grava Kempinas, W, and GR Klinefelter 2014 Interpreting histopathology in the epididymis. *Spermatogenesis* 4 e979114.

Deutsch, EW, N Bandeira, Y Perez-Riverol, V Sharma, JJ Carver, L Mendoza, DJ Kundu, S Wang, C Bandla, S Kamatchinathan, S Hewapathirana, BS Pullman, J Wertz, Z Sun, S Kawano, S Okuda, Y Watanabe, B MacLean, MJ MacCoss, Y Zhu, Y Ishihama, and JA Vizcaíno 2023 The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res* 51 D1539-d1548.

Dickinson, ME, AM Flenniken, X Ji, L Teboul, MD Wong, JK White, TF Meehan, WJ Weninger, H Westerberg, H Adissu, CN Baker, L Bower, JM Brown, LB Caddle, F Chiani, D Clary, J Cleak, MJ Daly, JM Denegre, B Doe, ME Dolan, SM Edie, H Fuchs, V Gailus-Durner, A Galli, A Gambadoro, J Gallegos, S Guo, NR Horner, C-W Hsu, SJ Johnson, S Kalaga, LC Keith, L Lanoue, TN Lawson, M Lek, M Mark, S Marschall, J Mason, ML McElwee, S Newbigging, LMJ Nutter, KA Peterson, R Ramirez-Solis, DJ Rowland, E Ryder, KE Samocha, JR Seavitt, M Selloum, Z

Szoke-Kovacs, M Tamura, AG Trainor, I Tudose, S Wakana, J Warren, O Wendling, DB West, L Wong, A Yoshiki, M McKay, B Urban, C Lund, E Froeter, T LaCasse, A Mehalow, E Gordon, LR Donahue, R Taft, P Kutney, S Dion, L Goodwin, S Kales, R Urban, K Palmer, F Pertuy, D Bitz, B Weber, P Goetz-Reiner, H Jacobs, E Le Marchand, A El Amri, L El Fertak, H Ennah, D Ali-Hadji, A Ayadi, M Wattenhofer-Donze, S Jacquot, P André, M-C Birling, G Pavlovic, T Sorg, I Morse, F Benso, ME Stewart, C Copley, J Harrison, S Joynson, R Guo, D Qu, S Spring, et al. 2016 High-throughput discovery of novel developmental phenotypes. *Nature* 537 508-514.

Drew, K, C Lee, RL Huizar, F Tu, B Borgeson, CD McWhite, Y Ma, JB Wallingford, and EM Marcotte 2017 Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol Syst Biol* 13 932.

du Plessis, SS, A Agarwal, G Mohanty, and M van der Linde 2015 Oxidative phosphorylation versus glycolysis: what fuel do spermatozoa use? *Asian J Androl* 17 230-235.

Dudkiewicz, AB 1983 Inhibition of fertilization in the rabbit by anti-acrosin antibodies. *Gamete research* 8 183-197.

Duffy, JMN, GD Adamson, E Benson, S Bhattacharya, S Bhattacharya, M Bofill, K Brian, B Collura, C Curtis, JLH Evers, RG Farquharson, A Fincham, S Franik, LC Giudice, E Glanville, M Hickey, AW Horne, ML Hull, NP Johnson, V Jordan, Y Khalaf, JML Knijnenburg, RS Legro, S Lensen, J MacKenzie, D Mavrelos, BW Mol, DE Morbeck, H Nagels, EHY Ng, C Niederberger, AS Otter, L Puscasiu, S Rautakallio-Hokkanen, L Sadler, I Sarris, M Showell, J Stewart, A Strandell, C Strawbridge, A Vail, M van Wely, M Vercoe, NL Vuong, AY Wang, R Wang, J Wilkinson, K Wong, TY Wong, CM Farquhar, H AlAhwany, O Balaban, F Barton, Y Beebeejaun, J Boivin, JJA Bosteels, C Calhaz-Jorge, A D'Angelo, FD L, JDJ C, E du Mez, AF R, MO Gerval,

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

JG L, EM Greenblatt, G Hartshorne, C Helliwell, LJ Hughes, J Jo, J Jovanović, L Kiesel, C Kietpeerakool, E Kostova, T Kucuk, R Kumar, RL Lawrence, N Lee, KE Lindemann, OM Loto, PJ Lutjen, M MacKinven, M Mascarenhas, H McLaughlin, SM Mourad, LK Nguyen, RJ Norman, M Olic, KL Overfield, M Parker-Harris, S Repping, R Rizzo, P Salacone, CH Saunders, R Sengupta, IA Sfontouris, NR Silverman, HL Torrance, EP Uphoff, SA Wakeman, T Wischmann, et al. 2020 Top 10 priorities for future infertility research: an international consensus development study† ‡. *Hum Reprod* 35 2715-2724.

Dun, MD, LA Mitchell, RJ Aitken, and B Nixon 2010 Sperm–zona pellucida interaction: molecular mechanisms and the potential for contraceptive intervention. *Fertility Control* 139-178.

Fitzpatrick, JL, AF Kahrl, and RR Snook 2022 SpermTree, a species-level database of sperm morphology spanning the animal tree of life. *Sci Data* 9 30.

Formosa, LE, MG Dibley, DA Stroud, and MT Ryan 2018 Building a complex complex: Assembly of mitochondrial respiratory chain complex I. *Seminars in Cell & Developmental Biology* 76 154-162.

Fu, Q, L Pan, D Huang, Z Wang, Z Hou, and M Zhang 2019 Proteomic profiles of buffalo spermatozoa and seminal plasma. *Theriogenology* 134 74-82.

Fuchs, H, JA Aguilar-Pimentel, OV Amarie, L Becker, J Calzada-Wack, YL Cho, L Garrett, SM Hölter, M Irmler, M Kistler, M Kraiger, P Mayer-Kuckuk, K Moreth, B Rathkolb, J Rozman, P da Silva Buttkus, I Treise, A Zimprich, K Gampe, C Hutterer, C Stöger, S Leuchtenberger, H Maier, M Miller, A Scheideler, M Wu, J Beckers, R Bekeredjian, M Brielmeier, DH Busch, M Klingenspor, T Klopstock, M Ollert, C Schmidt-Weber, T Stöger, E Wolf, W Wurst, A Yildirim, A Zimmer, V Gailus-Durner, and M Hrabě de Angelis 2018 Understanding gene functions and

disease mechanisms: Phenotyping pipelines in the German Mouse Clinic. *Behav Brain Res* 352 187-196.

Fuentes-Albero, MC, L González-Brusi, P Cots, C Luongo, S Abril-Sánchez, JL Ros-Santaella, E Pintus, S Ruiz-Díaz, C Barros-García, MJ Sánchez-Calabuig, D García-Párraga, M Avilés, MJ Izquierdo Rico, and FA García-Vázquez 2021 Protein Identification of Spermatozoa and Seminal Plasma in Bottlenose Dolphin (Tursiops truncatus). *Front Cell Dev Biol* 9 673961.

Giansanti, P, P Samaras, Y Bian, C Meng, A Coluccio, M Frejno, H Jakubowsky, S Dobiasch, RR Hazarika, J Rechenberger, J Calzada-Wack, J Krumm, S Mueller, CY Lee, N Wimberger, L Lautenbacher, Z Hassan, YC Chang, C Falcomatà, FP Bayer, S Bärthel, T Schmidt, R Rad, SE Combs, M The, F Johannes, D Saur, MH de Angelis, M Wilhelm, G Schneider, and B Kuster 2022 Mass spectrometry-based draft of the mouse proteome. *Nat Methods* 19 803-811.

Gist, DH, A Bagwill, V Lance, DM Sever, and RM Elsey 2008 Sperm storage in the oviduct of the American alligator. *J Exp Zool A Ecol Genet Physiol* 309 581-587.

Götz, S, JM García-Gómez, J Terol, TD Williams, SH Nagaraj, MJ Nueda, M Robles, M Talón, J Dopazo, and A Conesa 2008 High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36 3420-3435.

Groza, T, FL Gomez, HH Mashhadi, V Muñoz-Fuentes, O Gunes, R Wilson, P Cacheiro, A Frost, P Keskivali-Bond, B Vardal, A McCoy, TK Cheng, L Santos, S Wells, D Smedley, A-M Mallon, and H Parkinson 2022 The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* 51 D1038-D1045.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Guyonnet, B, N Egge, and GA Cornwall 2014 Functional amyloids in the mouse sperm acrosome. *Mol Cell Biol* 34 2624-2634.

Guyonnet, B, M Zabet-Moghaddam, S SanFrancisco, and GA Cornwall 2012 Isolation and proteomic characterization of the mouse sperm acrosomal matrix. *Mol Cell Proteomics* 11 758-774.

Haase, B, C Schlötterer, ME Hundrieser, H Kuiper, O Distl, E Töpfer-Petersen, and T Leeb 2005 Evolution of the spermadhesin gene family. *Gene* 352 20-29.

Hagn, M, S Marschall, and M Hrabè de Angelis 2007 EMMA—The European mouse mutant archive. *Briefings in Functional Genomics* 6 186-192.

Heck, M, and BA Neely 2020 Proteomics in Non-model Organisms: A New Analytical Frontier. *J Proteome Res* 19 3595-3606.

Henriques, BJ, R Katrine Jentoft Olsen, CM Gomes, and P Bross 2021 Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease. *Gene* 776 145407.

Hermo, L, RM Pelletier, DG Cyr, and CE Smith 2010a Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 1: background to spermatogenesis, spermatogonia, and spermatocytes. *Microscopy research and technique* 73 241-278.

Hermo, L, RM Pelletier, DG Cyr, and CE Smith 2010b Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 2: changes in spermatid organelles associated with development of spermatozoa. *Microscopy research and technique* 73 279-319.

Holt, WV, and RE Lloyd 2010 Sperm storage in the vertebrate female reproductive tract: how does it work so well? *Theriogenology* 73 713-722.

Hua, R, R Xue, Y Liu, Y Li, X Sha, K Li, Y Gao, Q Shen, M Lv, Y Xu, Z Zhang, X He, Y Cao, and H Wu 2023 ACROSIN deficiency causes total fertilization failure in humans by preventing the sperm from penetrating the zona pellucida. *Human Reproduction* 38 1213-1223.

Huang da, W, BT Sherman, and RA Lempicki 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4 44-57.

Hulsen, T 2022 DeepVenn--a web application for the creation of area-proportional Venn diagrams using the deep learning framework Tensorflow. js. *arXiv preprint arXiv:2210.04597*.

Johnston, SD, MR McGowan, NJ Phillips, and P O'Callaghan 2000 Optimal physicochemical conditions for the manipulation and short-term preservation of koala (Phascolarctos cinereus) spermatozoa. *J Reprod Fertil* 118 273-281.

Johnston, SD, YP Zee, C López-Fernández, and J Gosálvez 2012 The effect of chilled storage and cryopreservation on the sperm DNA fragmentation dynamics of a captive population of koalas. *J Androl* 33 1007-1015.

Juárez, JD, F Marco-Jiménez, A Talavan, X García-Domínguez, MP Viudes-De-Castro, R Lavara, and JS Vicente 2020 Evaluation by re-derivation of a paternal line after 18 generations on seminal traits, proteome and fertility. *Livestock Science* 232 103894.

Karn, RC, NL Clark, ED Nguyen, and WJ Swanson 2008 Adaptive evolution in rodent seminal vesicle secretion proteins. *Mol Biol Evol* 25 2301-2310.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Kasvandik, S, G Sillaste, A Velthut-Meikas, AV Mikelsaar, T Hallap, P Padrik, T Tenson, Ü Jaakma, S Kõks, and A Salumets 2015 Bovine sperm plasma membrane proteomics through biotinylation and subcellular enrichment. *Proteomics* 15 1906-1920.

Korasick, DA, and JJ Tanner 2021 Impact of missense mutations in the ALDH7A1 gene on enzyme structure and catalytic function. *Biochimie* 183 49-54.

Krämer, A, J Green, J Pollard, Jr, and S Tugendreich 2013 Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30 523-530.

Kraus, M, M Tichá, B Zelezná, J Peknicová, and V Jonáková 2005 Characterization of human seminal plasma proteins homologous to boar AQN spermadhesins. *J Reprod Immunol* 65 33-46.

Labas, V, I Grasseau, K Cahier, A Gargaros, G Harichaux, AP Teixeira-Gomes, S Alves, M Bourin, N Gérard, and E Blesbois 2015 Qualitative and quantitative peptidomic and proteomic approaches to phenotyping chicken semen. *J Proteomics* 112 313-335.

Leahy, T, JP Rickard, T Pini, BM Gadella, and SP de Graaf 2020 Quantitative Proteomic Analysis of Seminal Plasma, Sperm Membrane Proteins, and Seminal Extracellular Vesicles Suggests Vesicular Mechanisms Aid in the Removal and Addition of Proteins to the Ram Sperm Membrane. *Proteomics* 20 e1900289.

Letunic, I, and P Bork 2007 Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23 127-128.

Letunic, I, and P Bork 2011 Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 39 W475-478.

Li, C, R Yu, H Liu, J Qiao, F Zhang, S Mu, M Guo, H Zhang, Y Li, and X Kang 2023 Sperm acrosomal released proteome reveals MDH and VDAC3 from mitochondria are involved in acrosome formation during spermatogenesis in Eriocheir sinensis. *Gene* 887 147784.

Lin, YN, A Roy, W Yan, KH Burns, and MM Matzuk 2007 Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Mol Cell Biol* 27 6794-6805.

Liu, DY, and H Gordon Baker 1993 Inhibition of acrosin activity with a trypsin inhibitor blocks human sperm penetration of the zona pellucida. *Biology of Reproduction* 48 340-348.

Liu, F, X Liu, X Liu, T Li, P Zhu, Z Liu, H Xue, W Wang, X Yang, J Liu, and W Han 2019 Integrated Analyses of Phenotype and Quantitative Proteome of CMTM4 Deficient Mice Reveal Its Association with Male Fertility. *Mol Cell Proteomics* 18 1070-1084.

Lui, WY, WM Lee, and CY Cheng 2003 Sertoli-germ cell adherens junction dynamics in the testis are regulated by RhoB GTPase via the ROCK/LIMK signaling pathway. *Biol Reprod* 68 2189-2206.

Lundwall, Å, M Persson, K Hansson, and M Jonsson 2020 Identification of the major rabbit and guinea pig semen coagulum proteins and description of the diversity of the REST gene locus in the mammalian clade Glires. *PLoS One* 15 e0240607.

Martin, JH, R Mohammed, SJ Delforce, DA Skerrett-Byrne, CC de Meaultsart, JG Almazi, AN Stephens, NM Verrills, E Dimitriadis, Y Wang, ER Lumbers, and KG Pringle 2022 Role of the prorenin receptor in endometrial cancer cell growth. *Oncotarget* 13 587-599.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Mohanty, G, N Swain, and L Samanta 2015 Sperm Proteome: What Is on the Horizon? *Reprod Sci* 22 638-653.

Montellier, E, F Boussouar, S Rousseaux, K Zhang, T Buchou, F Fenaille, H Shiota, A Debernardi, P Héry, S Curtet, M Jamshidikia, S Barral, H Holota, A Bergon, F Lopez, P Guardiola, K Pernet, J Imbert, C Petosa, M Tan, Y Zhao, M Gérard, and S Khochbin 2013 Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev* 27 1680-1692.

Murphy, EM, C Murphy, C O'Meara, G Dunne, B Eivers, P Lonergan, and S Fair 2017 A comparison of semen diluents on the in vitro and in vivo fertility of liquid bull semen. *J Dairy Sci* 100 1541-1554.

Murray, HC, AK Enjeti, RGS Kahl, HM Flanagan, J Sillar, DA Skerrett-Byrne, JG Al Mazi, GG Au, CE de Bock, K Evans, ND Smith, A Anderson, B Nixon, RB Lock, MR Larsen, NM Verrills, and MD Dun 2021 Quantitative phosphoproteomics uncovers synergy between DNA-PK and FLT3 inhibitors in acute myeloid leukaemia. *Leukemia* 35 1782-1787.

Naaby-Hansen, S, A Mandal, MJ Wolkowicz, B Sen, VA Westbrook, J Shetty, SA Coonrod, KL Klotz, Y-H Kim, LA Bush, CJ Flickinger, and JC Herr 2002 CABYR, a Novel Calcium-Binding Tyrosine Phosphorylation-Regulated Fibrous Sheath Protein Involved in Capacitation. *Developmental Biology* 242 236-254.

Nixon, B, EG Bromfield, J Cui, and GN De Iuliis 2017 Heat Shock Protein A2 (HSPA2): Regulatory Roles in Germ Cell Development and Sperm Function. *Adv Anat Embryol Cell Biol* 222 67-93.

Nixon, B, EG Bromfield, MD Dun, KA Redgrove, EA McLaughlin, and RJ Aitken 2015 The role of the molecular chaperone heat shock protein A2 (HSPA2) in regulating human sperm-egg recognition. *Asian J Androl* 17 568-573.

Nixon, B, SL Cafe, AL Eamens, GN De Iuliis, EG Bromfield, JH Martin, DA Skerrett-Byrne, and MD Dun 2020 Molecular insights into the divergence and diversity of post-testicular maturation strategies. *Molecular and Cellular Endocrinology* 517 110955.

Nixon, B, SD Johnston, DA Skerrett-Byrne, AL Anderson, SJ Stanger, EG Bromfield, JH Martin, PM Hansbro, and MD Dun 2019 Modification of Crocodile Spermatozoa Refutes the Tenet That Post-testicular Sperm Maturation Is Restricted To Mammals. *Mol Cell Proteomics* 18 S58-s76.

Oliva, R, J Martínez-Heredia, and JM Estanyol 2008 Proteomics in the study of the sperm cell composition, differentiation and function. *Syst Biol Reprod Med* 54 23-36.

Pérez-Patiño, C, I Parrilla, J Li, I Barranco, EA Martínez, H Rodriguez-Martínez, and J Roca 2019 The Proteome of Pig Spermatozoa Is Remodeled During Ejaculation. *Mol Cell Proteomics* 18 41-50.

Perez-Riverol, Y, J Bai, C Bandla, D García-Seisdedos, S Hewapathirana, S Kamatchinathan, DJ Kundu, A Prakash, A Frericks-Zipper, M Eisenacher, M Walzer, S Wang, A Brazma, and JA Vizcaíno 2022 The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50 D543-d552.

Pini, T, J Parks, J Russ, M Dzieciatkowska, KC Hansen, WB Schoolcraft, and M Katz-Jaffe 2020 Obesity significantly alters the human sperm proteome, with potential implications for fertility. *J Assist Reprod Genet* 37 777-787.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Piomboni, P, R Focarelli, A Stendardi, A Ferramosca, and V Zara 2012 The role of mitochondria in energy production for human sperm motility. *International Journal of Andrology* 35 109-124.

Ramesha, KP, P Mol, U Kannegundla, LN Thota, L Gopalakrishnan, E Rana, N Azharuddin, KK Mangalaparthi, M Kumar, G Dey, A Patil, K Saravanan, SK Behera, S Jeyakumar, A Kumaresan, MA Kataktalware, and TSK Prasad 2020 Deep Proteome Profiling of Semen of Indian Indigenous Malnad Gidda (Bos indicus) Cattle. *J Proteome Res* 19 3364-3376.

Redgrove, KA, B Nixon, MA Baker, L Hetherington, G Baker, DY Liu, and RJ Aitken 2012 The molecular chaperone HSPA2 plays a key role in regulating the expression of sperm surface receptors that mediate sperm-egg recognition. *PLoS One* 7 e50851.

Robert, M, and C Gagnon 1994 Sperm motility inhibitor from human seminal plasma: presence of a precursor molecule in seminal vesicle fluid and its molecular processing after ejaculation. *Int J Androl* 17 232-240.

Sampson, MJ, WK Decker, AL Beaudet, W Ruitenbeek, D Armstrong, MJ Hicks, and WJ Craigen 2001 Immotile sperm and infertility in mice lacking mitochondrial voltage-dependent anion channel type 3. *Journal of Biological Chemistry* 276 39206-39212.

Schiza, C, D Korbakis, E Panteleli, K Jarvi, AP Drabovich, and EP Diamandis 2018 Discovery of a Human Testis-specific Protein Complex TEX101-DPEP3 and Selection of Its Disrupting Antibodies. *Mol Cell Proteomics* 17 2480-2495.

Schjenken, JE, DJ Sharkey, and SA Robertson 2018 Seminal vesicle—secretion.

Sha, Y, W Liu, H Nie, L Han, C Ma, X Zhang, Z Xiao, W Qin, X Jiang, and X Wei 2022 Homozygous mutation in DNALI1 leads to asthenoteratozoospermia by affecting the inner dynein arms. *Front Endocrinol (Lausanne)* 13 1058651.

Shannon, P, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, and T Ideker 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13 2498-2504.

Sharma, U, F Sun, CC Conine, B Reichholf, S Kukreja, VA Herzog, SL Ameres, and OJ Rando 2018 Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Dev Cell* 46 481-494.e486.

Shen, D, C Zhou, M Cao, W Cai, H Yin, L Jiang, and S Zhang 2021 Differential Membrane Protein Profile in Bovine X- and Y-Sperm. *J Proteome Res* 20 3031-3042.

Sherman, BT, M Hao, J Qiu, X Jiao, MW Baseler, HC Lane, T Imamichi, and W Chang 2022 DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* 50 W216-221.

Sievers, F, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, and J Söding 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 7 539.

Skerrett-Byrne, DA, AL Anderson, EG Bromfield, IR Bernstein, JE Mulhall, JE Schjenken, MD Dun, SJ Humphrey, and B Nixon 2022 Global profiling of the proteomic changes associated with the post-testicular maturation of mouse spermatozoa. *Cell Rep* 41 111655.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Skerrett-Byrne, DA, AL Anderson, L Hulse, C Wass, MD Dun, EG Bromfield, GN De Iuliis, M Pyne, V Nicolson, SD Johnston, and B Nixon 2021a Proteomic analysis of koala (phascolarctos cinereus) spermatozoa and prostatic bodies. *Proteomics* 21 e2100067.

Skerrett-Byrne, DA, EG Bromfield, HC Murray, MFB Jamaluddin, AG Jarnicki, M Fricker, AT Essilfie, B Jones, TJ Haw, D Hampsey, AL Anderson, B Nixon, RJ Scott, PAB Wark, MD Dun, and PM Hansbro 2021b Time-resolved proteomic profiling of cigarette smoke-induced experimental chronic obstructive pulmonary disease. *Respirology* 26 960-973.

Skerrett-Byrne, DA, R Teperino, and B Nixon 2024 ShinySperm: navigating the sperm proteome landscape. *Reprod Fertil Dev* 36.

Skerrett-Byrne, DA, NA Trigg, EG Bromfield, MD Dun, IR Bernstein, AL Anderson, SJ Stanger, LA MacDougall, T Lord, RJ Aitken, SD Roman, SA Robertson, B Nixon, and JE Schjenken 2021c Proteomic Dissection of the Impact of Environmental Exposures on Mouse Seminal Vesicle Function. *Mol Cell Proteomics* 20 100107.

Smyth, SP, B Nixon, AL Anderson, HC Murray, JH Martin, LA MacDougall, SA Robertson, DA Skerrett-Byrne, and JE Schjenken 2022 Elucidation of the protein composition of mouse seminal vesicle fluid. *Proteomics* 22 e2100227.

Smyth, SP, B Nixon, DA Skerrett-Byrne, ND Burke, and EG Bromfield 2024 Building an Understanding of Proteostasis in Reproductive Cells: The Impact of Reactive Carbonyl Species on Protein Fate. *Antioxid Redox Signal* 41 296-321.

Somashekar, L, S Selvaraju, S Parthipan, SK Patil, BK Binsila, MM Venkataswamy, S Karthik Bhat, and JP Ravindra 2017 Comparative sperm protein profiling in bulls differing in fertility and

identification of phosphatidylethanolamine-binding protein 4, a potential fertility marker. *Andrology* 5 1032-1051.

Spadafora, C 2023 The epigenetic basis of evolution. *Prog Biophys Mol Biol* 178 57-69.

Stallmeyer, B, C Bühlmann, R Stakaitis, AK Dicke, F Ghieh, L Meier, A Zoch, D MacKenzie MacLeod, J Steingröver, Ö Okutman, D Fietz, A Pilatz, A Riera-Escamilla, MJ Xavier, C Ruckert, S Di Persio, N Neuhaus, AS Gurbuz, A Şalvarci, N Le May, K McEleny, C Friedrich, G van der Heijden, MJ Wyrwoll, S Kliesch, JA Veltman, C Krausz, S Viville, DF Conrad, D O'Carroll, and F Tüttelmann 2024 Inherited defects of piRNA biogenesis cause transposon de-repression, impaired spermatogenesis, and human male infertility. *Nat Commun* 15 6637.

Staudt, DE, HC Murray, DA Skerrett-Byrne, ND Smith, MFB Jamaluddin, RGS Kahl, RJ Duchatel, ZP Germon, T McLachlan, ER Jackson, IJ Findlay, PS Kearney, A Mannan, HP McEwen, AM Douglas, B Nixon, NM Verrills, and MD Dun 2022 Phospho-heavy-labeled-spiketide FAIMS stepped-CV DDA (pHASED) provides real-time phosphoproteomics data to aid in cancer drug selection. *Clin Proteomics* 19 48.

Suarez, SS, and AA Pacey 2006 Sperm transport in the female reproductive tract. *Hum Reprod Update* 12 23-37.

Szklarczyk, D, AL Gable, KC Nastou, D Lyon, R Kirsch, S Pyysalo, NT Doncheva, M Legeay, T Fang, P Bork, LJ Jensen, and C von Mering 2021 The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49 D605-d612.

Tomar, A, M Gomez-Velazquez, R Gerlini, G Comas-Armangué, L Makharadze, T Kolbe, A Boersma, M Dahlhoff, JP Burgstaller, M Lassi, J Darr, J Toppari, H Virtanen, A Kühnapfel, M

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Scholz, K Landgraf, W Kiess, M Vogel, V Gailus-Durner, H Fuchs, S Marschall, M Hrabě de Angelis, N Kotaja, A Körner, and R Teperino 2024 Epigenetic inheritance of diet-induced and sperm-borne mitochondrial RNAs. *Nature* 630 720-727.

Trigg, NA, DA Skerrett-Byrne, MJ Xavier, W Zhou, AL Anderson, SJ Stanger, AL Katen, GN De Iuliis, MD Dun, SD Roman, AL Eamens, and B Nixon 2021 Acrylamide modulates the mouse epididymal proteome to drive alterations in the sperm small non-coding RNA profile and dysregulate embryo development. *Cell Rep* 37 109787.

Tyanova, S, T Temu, P Sinitcyn, A Carlson, MY Hein, T Geiger, M Mann, and J Cox 2016 The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* 13 731-740.

Uhlén, M, L Fagerberg, BM Hallström, C Lindskog, P Oksvold, A Mardinoglu, Å Sivertsson, C Kampf, E Sjöstedt, A Asplund, I Olsson, K Edlund, E Lundberg, S Navani, CA Szigyarto, J Odeberg, D Djureinovic, JO Takanen, S Hober, T Alm, PH Edqvist, H Berling, H Tegel, J Mulder, J Rockberg, P Nilsson, JM Schwenk, M Hamsten, K von Feilitzen, M Forsberg, L Persson, F Johansson, M Zwahlen, G von Heijne, J Nielsen, and F Pontén 2015 Proteomics. Tissue-based map of the human proteome. *Science* 347 1260419.

Urizar-Arenaza, I, N Osinalde, V Akimov, M Puglia, L Candenas, FM Pinto, I Muñoa-Hoyos, M Gianzo, R Matorras, J Irazusta, B Blagoev, N Subiran, and I Kratchmarova 2019 Phosphoproteomic and Functional Analyses Reveal Sperm-specific Protein Changes Downstream of Kappa Opioid Receptor in Human Spermatozoa. *Mol Cell Proteomics* 18 S118-s131.

Van den Broeck, L, DK Bhosale, K Song, CF Fonseca de Lima, M Ashley, T Zhu, S Zhu, B Van De Cotte, P Neyt, AC Ortiz, TR Sikes, J Aper, P Lootens, AM Locke, I De Smet, and R Sozzani

2023 Functional annotation of proteins for signaling network inference in non-model species. *Nature Communications* 14 4654.

Vandenbrouck, Y, L Lane, C Carapito, P Duek, K Rondel, C Bruley, C Macron, A Gonzalez de Peredo, Y Couté, K Chaoui, E Com, A Gateau, AM Hesse, M Marcellin, L Méar, E Mouton-Barbosa, T Robin, O Burlet-Schiltz, S Cianferani, M Ferro, T Fréour, C Lindskog, J Garin, and C Pineau 2016 Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J Proteome Res* 15 3998-4019.

Vicens, A, K Borziak, TL Karr, ERS Roldan, and S Dorus 2017 Comparative Sperm Proteomics in Mouse Species with Divergent Mating Systems. *Mol Biol Evol* 34 1403-1416.

Vitorino Carvalho, A, L Soler, A Thélie, I Grasseau, L Cordeiro, D Tomas, AP Teixeira-Gomes, V Labas, and E Blesblois 2021 Proteomic Changes Associated With Sperm Fertilizing Ability in Meat-Type Roosters. *Front Cell Dev Biol* 9 655866.

Wang, D, B Eraslan, T Wieland, B Hallström, T Hopf, DP Zolg, J Zecha, A Asplund, LH Li, C Meng, M Frejno, T Schmidt, K Schnatbaum, M Wilhelm, F Ponten, M Uhlen, J Gagneur, H Hahne, and B Kuster 2019 A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* 15 e8503.

Wen, Q, EI Tang, X Xiao, Y Gao, DS Chu, DD Mruk, B Silvestrini, and CY Cheng 2016 Transport of germ cells across the seminiferous epithelium during spermatogenesis-the involvement of both actin- and microtubule-based cytoskeletons. *Tissue Barriers* 4 e1265042.

Willems, P, I Fijalkowski, and P Van Damme 2020 Lost and found: re-searching and re-scoring proteomics data aids genome annotation and improves proteome coverage. *Msystems* 5 10.1128/msystems. 00833-00820.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Xu, K, L Yang, L Zhang, and H Qi 2020 Lack of AKAP3 disrupts integrity of the subcellular structure and proteome of mouse sperm and causes male sterility. *Development* 147.

Xu, Y, Q Han, C Ma, Y Wang, P Zhang, C Li, X Cheng, and H Xu 2021 Comparative Proteomics and Phosphoproteomics Analysis Reveal the Possible Breed Difference in Yorkshire and Duroc Boar Spermatozoa. *Front Cell Dev Biol* 9 652809.

Yap, YT, W Li, Q Huang, Q Zhou, D Zhang, Y Sheng, L Mladenovic-Lucas, S-P Yee, KE Orwig, JG Granneman, DC Williams, Jr., RA Hess, A Toure, and Z Zhang 2023 DNALI1 interacts with the MEIG1/PACRG complex within the manchette and is required for proper sperm flagellum assembly in mice. *eLife* 12 e79620.

Yin, Y, S Cao, H Fu, X Fan, J Xiong, Q Huang, Y Liu, K Xie, TG Meng, Y Liu, D Tang, T Yang, B Dong, S Qi, L Nie, H Zhang, H Hu, W Xu, F Li, L Dai, QY Sun, and Z Li 2020 A noncanonical role of NOD-like receptor NLRP14 in PGCLC differentiation and spermatogenesis. *Proc Natl Acad Sci U S A* 117 22237-22248.

Zhang, M, RZ Chiozzi, DA Skerrett-Byrne, T Veenendaal, J Klumperman, AJR Heck, B Nixon, JB Helms, BM Gadella, and EG Bromfield 2022 High Resolution Proteomic Analysis of Subcellular Fractionated Boar Spermatozoa Provides Comprehensive Insights Into Perinuclear Theca-Residing Proteins. *Front Cell Dev Biol* 10 836208.

FIGURE LEGENDS

Figure 1: Characterization of multispecies sperm proteomes. (A) Across 12 different species, over 2 TB of RAW spectral data was sourced from public repositories and processed using Proteome Discoverer 2.5, utilising highly stringent criteria. (B) The number of proteins identified in each of the 12 species and (C) the proportion their respective level of evidence for protein evidence as curated by to the UniProt Knowledge Base; 1) at protein level; 2) at transcript level; 3) protein inferred from homology; 4) protein predicted.

Figure 2: Humanization of sperm proteomes. (A) Summary of the number of proteins which were successfully converted to human homologues (duplicates removed). (B) The percentage of the original species proteome retained. (C) Heatmap depicts the top 25 unique pathways significantly enriched in at least one species (p-value ≤ 0.05), white denoting absence of detection. To the left, in gold is a phylogenetic tree depicts the evolutionary distances and relationships between the 9 species (generated with phylot and iTOL). Mirrored to the right in blue is the unbiased hierarchical clustering based upon the full remit of 482 pathways (sperm function).

Figure 3: Core sperm proteome. (A) Venn Diagram depicts the overlap of sperm proteomes at the level of taxonomic orders; Primates (H. sapiens), Rodentia (M. musculus), Artiodactyla (B. taurus, O. aries, S. scrofa, T. truncates), Lagomorpha (O. cuniculus), Diprotodontia (P. cinereus) and Crocodilia (C. porosus). UniProt Alignment tool (Clustal Omega) was utilized to interrogate the sequence alignment between each species for (B) Heat shock protein family A member 2 (HSPA2), (C) Protein kinase cAMP-dependent type I regulatory subunit alpha (PRKAR1A), (D) Voltage dependent anion channel 3 (VDAC3), (E) Acrosin (ACR), (F) Calcium binding tyrosine phosphorylation regulated (CABYR), and (G) Zona pellucida binding protein (ZPBP). The median alignment (% Similarity) is denoted in gold next to each protein symbol. The conserved sperm proteomes at the level of species (45 proteins) and order (135 proteins) were subjected to

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

analysis using Ingenuity Pathway Analysis (IPA). Heatmaps depict the comparative analysis of species and orders, with a refined focus on reproductive related (H) molecular functions and (I) pathways. Pathways and function known to important to motility, energy and egg interactions are highlight by blue, green and gold boxes respectively.

Figure 4: Core sperm protein mice knockouts affect sperm motility and fertilization capacity. (A) UniProt Alignment tool (Clustal Omega) was used to aligned protein sequence for Phosphatidylethanolamine binding protein 4 (Pebp4), Enoyl-CoA hydratase, short chain 1 (Echs1), Electron transfer flavoprotein subunit beta (ETFB), NADH:ubiquinone oxidoreductase subunit A10 (NDUFA10) and Aldehyde dehydrogenase 7 family member A1 (ALDH7A1). Gene knockout (KO) models were generated and sperm from heterozygous males were used for *in-vitro* fertilization (IVF). From these IVF experiments, the heatmaps depict the percentage of (B) motile sperm and those with progressive motility for each KO compared to wildtype (WT). Fertilization capacity was tracked and heatmaps depict the (C) 2-cell stage cleavage rate (%CR), blastocyte rate (%BR) and pregnancy rate (%PR). (D) Representative H&E images of the testis and epididymis from WT, *Pebp4*$^{-/-}$, and *Ndufa10*$^{+/-}$ 16-week-old mice, where multinucleated giant cells (MGCs) are indicated by arrowheads. Scale bars = 250 μm and 50 μm for insert image. Quantification of number of MGCs / mm^2 is represented by a bar chart as Mean ± SEM, with individual datapoints plotted; *** $p < 0.0001$

Figure 5: Reproductive strategies analyses. Sperm proteomes were stratified into three analyses, investigating the influence of evolutionary and imposed reproductive strategies on sperm protein composition, focusing on (A) sperm metabolism preference (glycolysis preference vs no preference), (B) location of the testes (internal vs external), and (C) history of selective breeding (yes vs no). Each analyses included Venn Diagrams to determine unique proteins to each biological context, which were further subject to analysis using Ingenuity Pathway Analysis (IPA). Heatmaps depict the comparative analysis of the resultant molecular functions.

SUPPLEMENTAL FIGURES

Figure S1: Sperm protein localization. Interrogation with UniProt maps proteins to their known sperm localizations.

Figure S2: Protein-to-protein interaction networks of core sperm proteomes. STRING network analyses of the core sperm proteome at the (A) species (45 proteins) and (B) order (135 proteins) taxonomic levels, drawn and coloured with Cytoscape.

Figure S3: Knockout mouse model pregnancy outcomes. Following successful births, the (A) litter size and (B) foetal sex distribution was recorded.

Figure S4: Histopathology of five KO models across the male reproductive tract. Representative images of H&E-stained sections from testis, cauda epididymis, anterior prostate and seminal vesicles of controls and knockout mice for the following genes: Aldehyde dehydrogenase 7 family member A1 (*Aldh7a1*), Phosphatidylethanolamine binding protein 4 (*Pebp4*), Enoyl-CoA hydratase, short chain 1 (*Echs1*), Electron transfer flavoprotein subunit beta (*Etfb*), and NADH:ubiquinone oxidoreductase subunit A10 (*Ndufa10*). Mice were 16 weeks old in all lines except for *Aldh7a1* (19 weeks old). Multinucleated giant cells (MGCs) are indicated by arrowheads. Scale bars are 250 μm and 50 μm.

 Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

SUPPLEMENTAL TABLES

Table S1

Complete publicly available proteome datasets investigated in this study. Also includes information on the FASTA files used across species.

Tables S2-13

The refined proteomic list for all 12 species, including UniProt accession, protein name, gene, level of annotated evidence, reviewed status and function from UniProt.

Table S14

All humanized proteomic datasets.

Table S15

The 45 proteins conserved at the species level and 135 proteins conserved at the order level.

Table S16

Core sperm proteome analysis outputs from Ingenuity Pathway Analysis (IPA)

Table S17

All proteins uniquely detected within each species and their respective outputs from Database for Annotation, Visualization and Integrated Discovery (DAVID).

Table S18

The analyses pertaining to the biological context comparisons, with unique proteins identifications listed, alongside the functional outputs from Ingenuity Pathway Analysis (IPA).

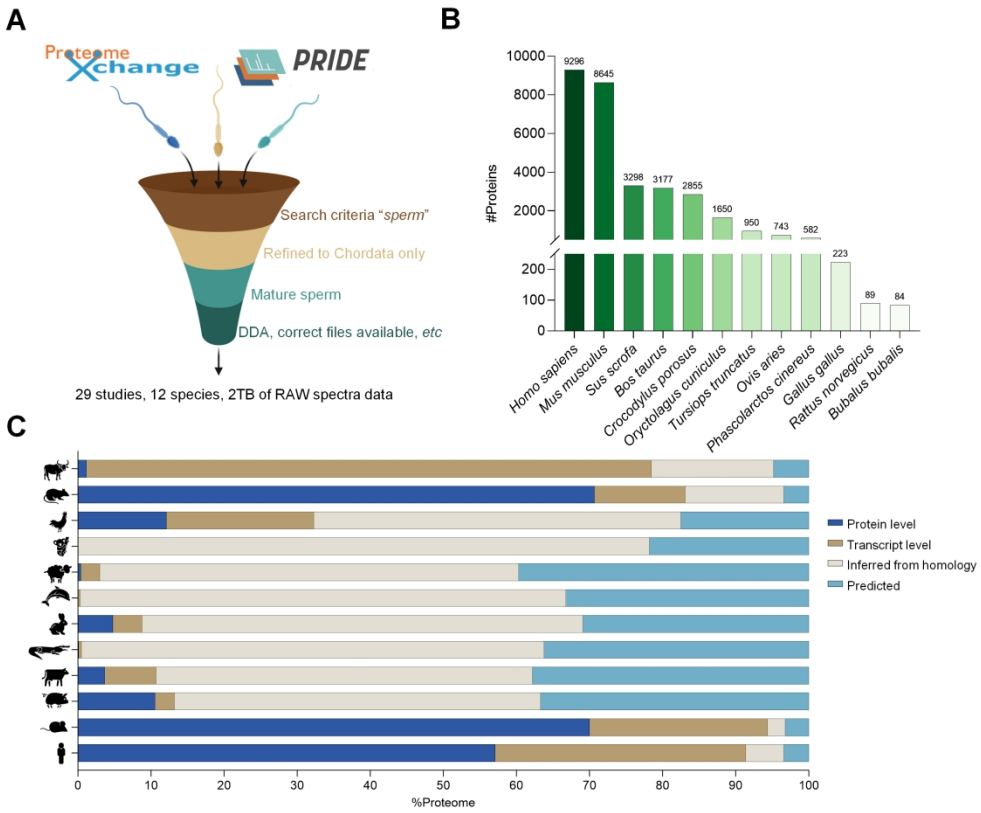

Figure 1: Characterization of multispecies sperm proteomes. (A) Across 12 different species, over 2 TB of RAW spectral data was sourced from public repositories and processed using Proteome Discoverer 2.5, utilising highly stringent criteria. (B) The number of proteins identified in each of the 12 species and (C) the proportion their respective level of evidence for protein evidence as curated by to the UniProt Knowledge Base; 1) at protein level; 2) at transcript level; 3) protein inferred from homology; 4) protein predicted.

180x149mm (400 x 400 DPI)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

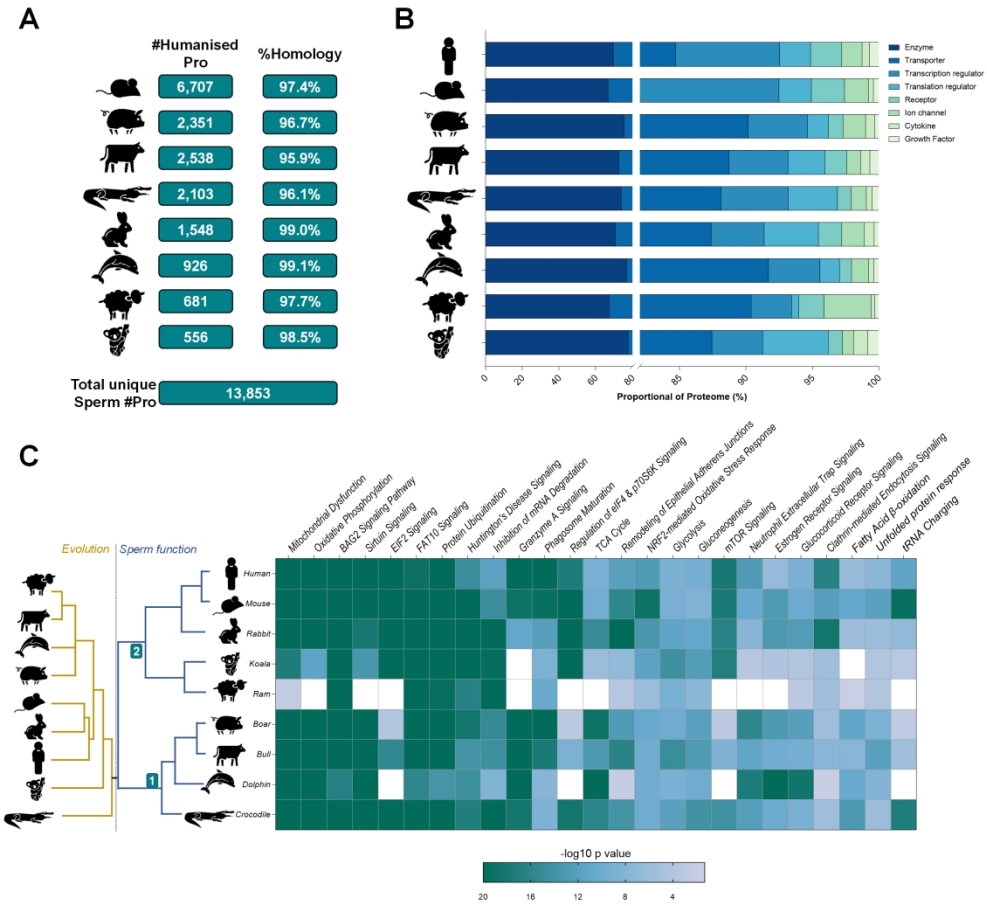

Figure 2: Humanization of sperm proteomes. (A) Summary of the number of proteins which were successfully converted to human homologues (duplicates removed). (B) The percentage of the original species proteome retained. (C) Heatmap depicts the top 25 unique pathways significantly enriched in at least one species (p-value ≤ 0.05), white denoting absence of detection. To the left, in gold is a phylogenetic tree depicts the evolutionary distances and relationships between the 9 species (generated with phylot and iTOL). Mirrored to the right in blue is the unbiased hierarchical clustering based upon the full remit of 482 pathways (sperm function).

162x149mm (400 x 400 DPI)

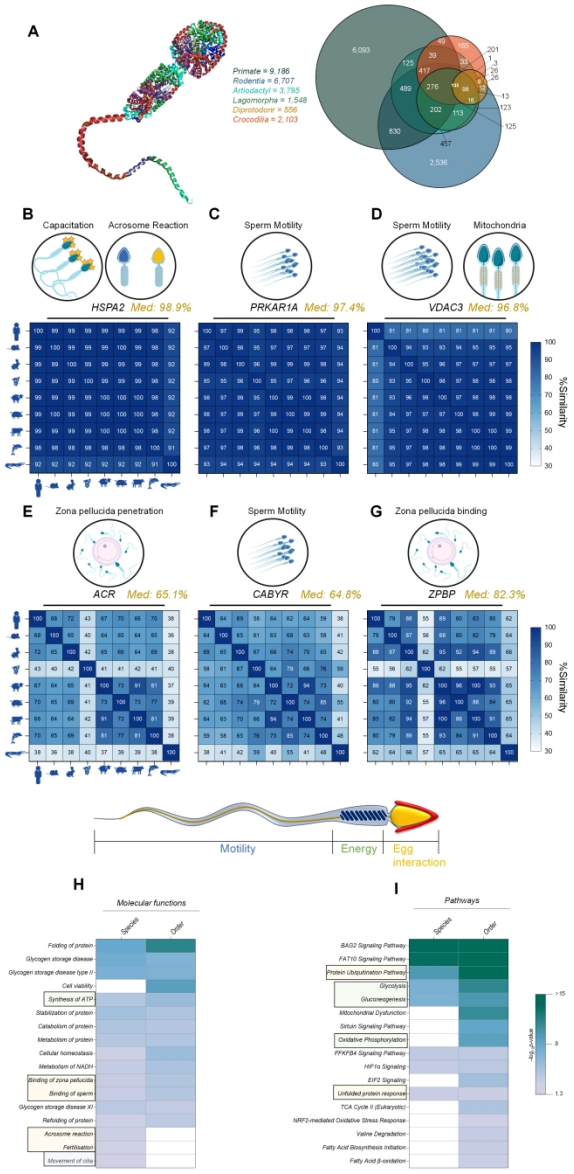

Figure 3: Core sperm proteome. (A) Venn Diagram depicts the overlap of sperm proteomes at the level of taxonomic orders; Primates (H. sapiens), Rodentia (M. musculus), Artiodactyla (B. taurus, O. aries, S. scrofa, T. truncates), Lagomorpha (O. cuniculus), Diprotodontia (P. cinereus) and Crocodilia (C. porosus). UniProt Alignment tool (Clustal Omega) was utilized to interrogate the sequence alignment between each species for (B) Heat shock protein family A member 2 (HSPA2), (C) Protein kinase cAMP-dependent type I regulatory subunit alpha (PRKAR1A), (D) Voltage dependent anion channel 3 (VDAC3), (E) Acrosin (ACR), (F) Calcium binding tyrosine phosphorylation regulated (CABYR), and (G) Zona pellucida binding protein (ZPBP). The median alignment (% Similarity) is denoted in gold next to each protein symbol. The conserved sperm proteomes at the level of species (45 proteins) and order (135 proteins) were subjected to analysis using Ingenuity Pathway Analysis (IPA). Heatmaps depict the comparative analysis of species and orders, with a refined focus on reproductive related (H) molecular functions and (I) pathways. Pathways and function known to important to motility, energy and egg interactions are highlight by blue, green and gold boxes respectively.

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

71x149mm (800 x 800 DPI)

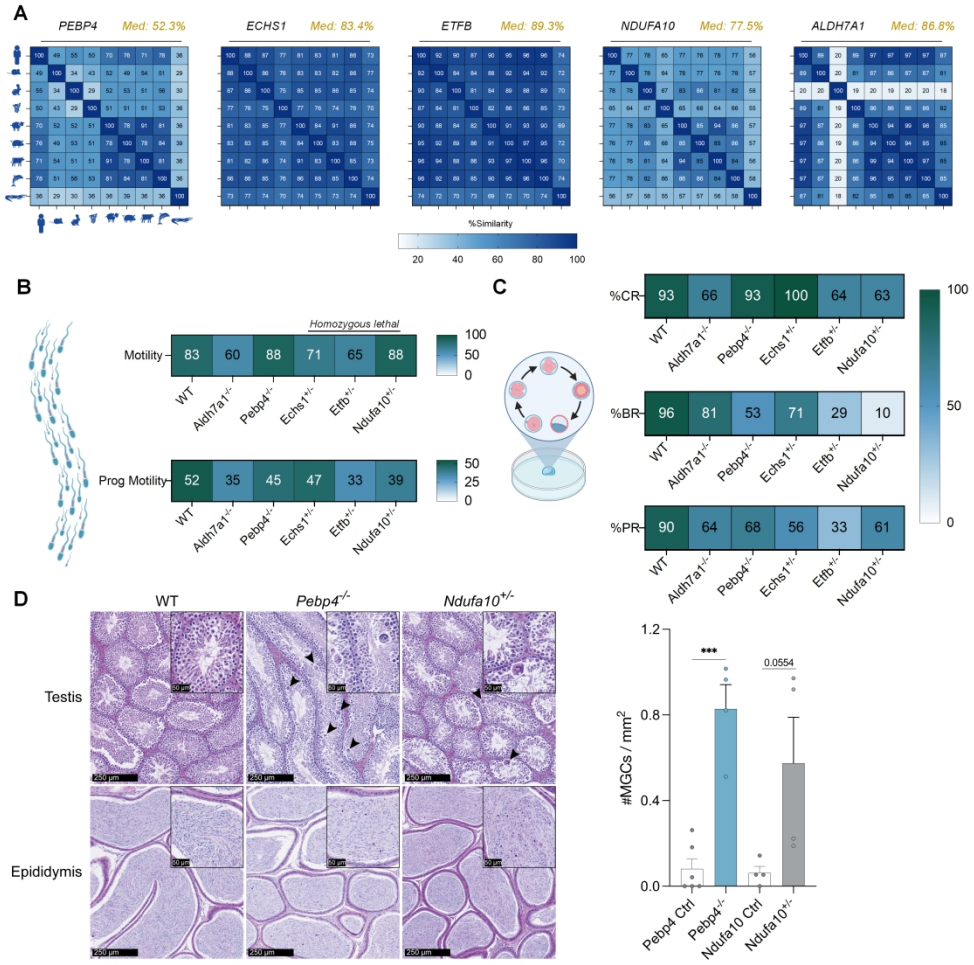

Figure 4: Core sperm protein mice knockouts affect sperm motility and fertilization capacity. (A) UniProt Alignment tool (Clustal Omega) was used to aligned protein sequence for Phosphatidylethanolamine binding protein 4 (Pebp4), Enoyl-CoA hydratase, short chain 1 (Echs1), Electron transfer flavoprotein subunit beta (ETFB), NADH:ubiquinone oxidoreductase subunit A10 (NDUFA10) and Aldehyde dehydrogenase 7 family member A1 (ALDH7A1). Gene knockout (KO) models were generated and sperm from heterozygous males were used for in-vitro fertilization (IVF). From these IVF experiments, the heatmaps depict the percentage of (B) motile sperm and those with progressive motility for each KO compared to wildtype (WT). Fertilization capacity was tracked and heatmaps depict the (C) cleavage rate (%CR), blastocyte rate (%BR) and pregnancy rate (%PR). (D) Representative H&E images of the testis and epididymis from WT, Pebp4-/-, and Ndufa10+/- 16-week-old mice, where multinucleated giant cells (MGCs) are indicated by arrowheads. Scale bars = 250 µm and 50 µm for insert image. Quantification of number of MGCs / mm2 is represented by a bar chart as Mean ± SEM, with individual datapoints plotted; *** $p < 0.0001$

149x149mm (600 x 600 DPI)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

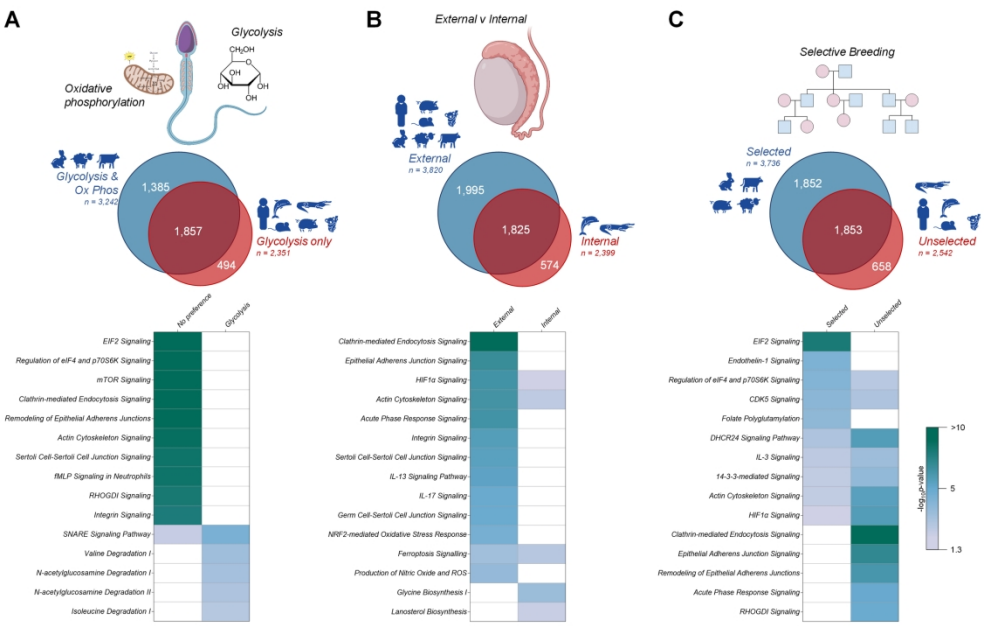

Figure 5: Reproductive strategies analyses. Sperm proteomes were stratified into three analyses, investigating the influence of evolutionary and imposed reproductive strategies on sperm protein composition, focusing on (A) sperm metabolism preference (glycolysis preference vs no preference), (B) location of the testes (internal vs external), and (C) history of selective breeding (yes vs no). Each analyses included Venn Diagrams to determine unique proteins to each biological context, which were further subject to analysis using Ingenuity Pathway Analysis (IPA). Heatmaps depict the comparative analysis of the resultant molecular functions.

209x134mm (300 x 300 DPI)

Table 1. Final proteomic datasets included in this study

Study	Dataset identifier	PMID	Species
Byrne et al. 2012	PXD000007	23081703	*Bos indicus*
Ramesha et al. 2020	PXD014172	32508098	*Bos indicus*
Kasvandik et al. 2015	PXD001096	25603787	*Bos taurus*
Shen et al. 2021	PXD019435	34009990	*Bos taurus*
Fu et al. 2019	PXD003859	31146187	*Bubalus bubalis*
Batra et al. 2021	PXD022114	34174811	*Bubalus bubalis*
Nixon et al. 2019	MSV000082258	30072580	*Crocodylus porosus*
Labas et al. 2015	PXD001254	25086240	*Gallus gallus*
Vitorino Carvalho et al. 2021	PXD022322	33898456	*Gallus gallus*
Vandenbrouck et al. 2016	PXD003947	27444420	*Homo sapiens*
Schiza et al. 2018	PXD007515	30097533	*Homo sapiens*
Urizar-Arenaza et al. 2019	PXD011290	30622161	*Homo sapiens*
Pini et al. 2020	PXD014849	32026202	*Homo sapiens*
Castillo et al. 2019	PXD014871	31824947	*Homo sapiens*
Guyonnet et al. 2012	PXD000575	22707618	*Mus musculus*
Guyonnet et al. 2014	PXD000592	24797071	*Mus musculus*
Castaneda et al. 2017	PXD005343	28630322	*Mus musculus*
Liu et al. 2019	PXD013092	30867229	*Mus musculus*
Xu et al. 2020	PXD016928	31969357	*Mus musculus*
Skerrett-Byrne et al. 2022	PXD028834	36384108	*Mus musculus*
Casares-Crespo et al. 2019	PXD007989	30753958	*Oryctolagus cuniculus*
Juárez et al. 2020	PXD015510	N/A	*Oryctolagus cuniculus*
Leahy et al. 2020	PXD017537	32383290	*Ovis aries*
Skerrett-Byrne et al. 2021a	PXD024250	34411425	*Phascolarctos cinereus*
Xu et al. 2021	PXD025607	34336820	*Sus scrofa*
Pérez-Patiño et al. 2019	PXD010062	30257877	*Sus scrofa*
Zhang et al. 2022	PXD030020	35252197	*Sus scrofa*
Fuentes-Albero et al. 2021	PXD024588	34336830	*Tursiops truncatus*
Bayram et al. 2016	PXD003164	26768581	*Mus musculus, Bos taurus, Sus scrofa, Rattus norvegicus*

 Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

Table 2. Summary of protein identifications by species

Species	**Studies, *n***	**Original protein IDs, *n***	**Humanized IDs, *n***	**Conversion rate (%)**	**Unique to species, *n* (%)**
Human (*Homo sapiens*)	5	9296			6093 (66.3)
House mouse (*Mus musculus*)	7	8645	8424	97.4	2536 (30.1)
Boar (*Sus scrofa*)	4	3298	3190	96.7	111 (3.5)
Cattle (*Bos taurus/indicus*)	5	3177	3048	95.9	149 (4.9)
Saltwater crocodile (*Crocodylus porosus*)	1	2855	2743	96.1	165 (6.0)
European rabbit (*Oryctolagus cuniculus*)	2	1650	1633	99.0	59 (3.6)
Common bottlenose dolphin (*Tursiops truncates*)	1	950	941	99.1	15 (1.6)
Sheep (*Ovis aries*)	1	743	726	97.7	40 (5.5)
Koala (*Phascolarctos cinereus*)	1	582	573	98.5	26 (4.5)
Chicken (*Gallus gallus*)	2	223			
Brown rat (*Rattus norvengicus*)	1	89			
Domestic water buffalo (*Bubalus bubalis*)	2	84			

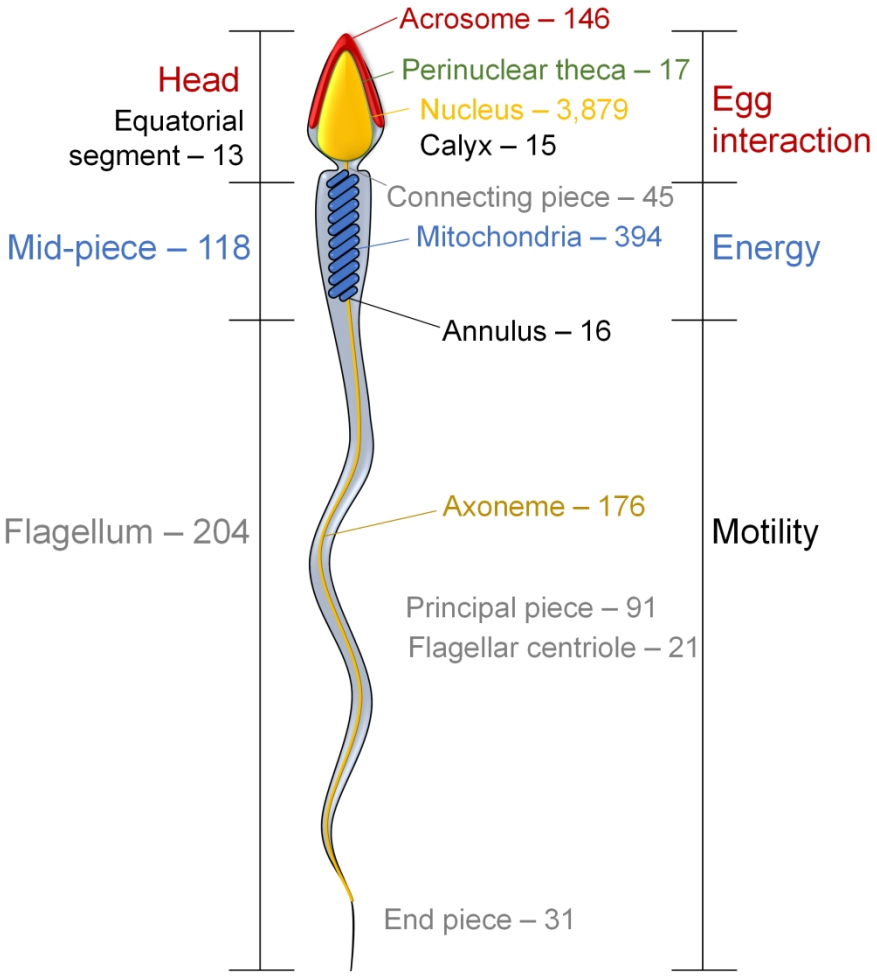

Figure S1: Sperm protein localization. Interrogation with UniProt maps proteins to their known sperm localizations.

152x149mm (400 x 400 DPI)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

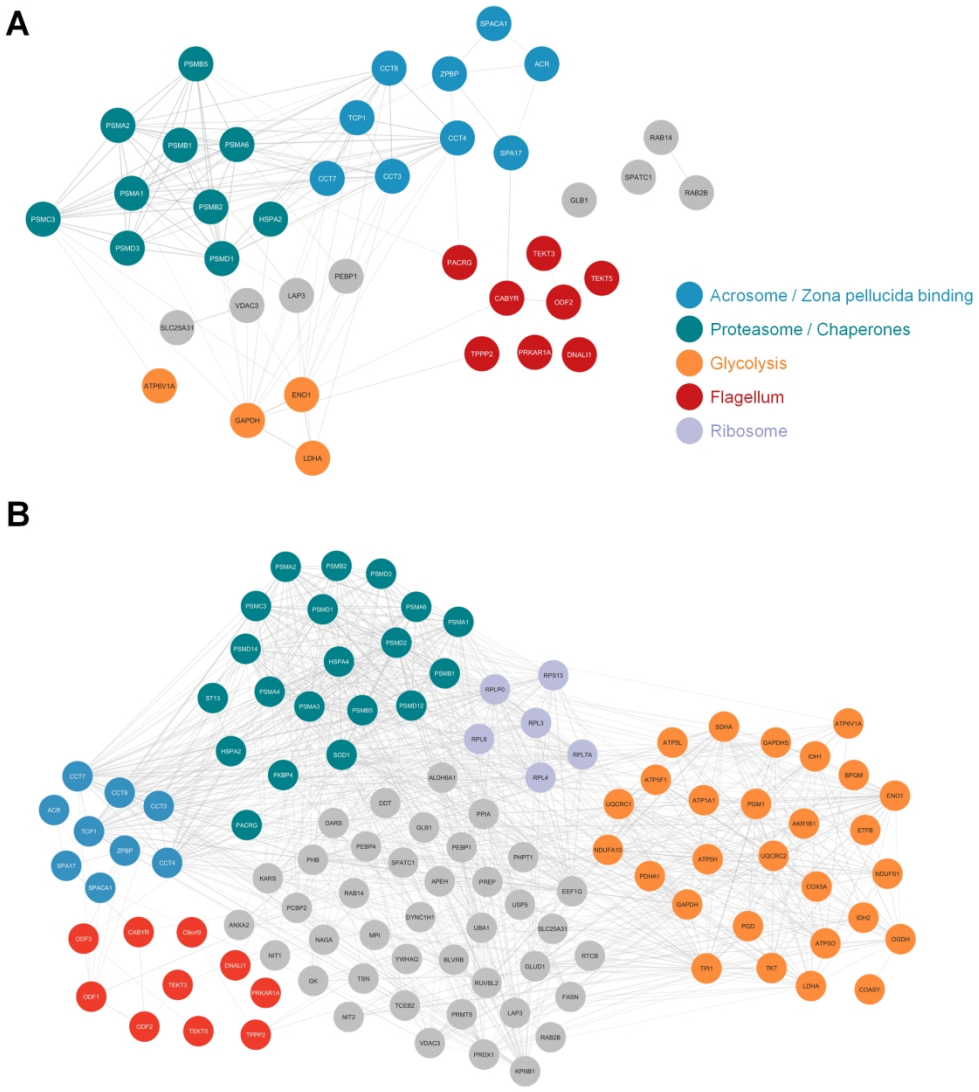

Figure S2: Protein-to-protein interaction networks of core sperm proteomes. STRING network analyses of the core sperm proteome at the (A) species (45 proteins) and (B) order (135 proteins) taxonomic levels, drawn and coloured with Cytoscape.

136x150mm (500 x 500 DPI)

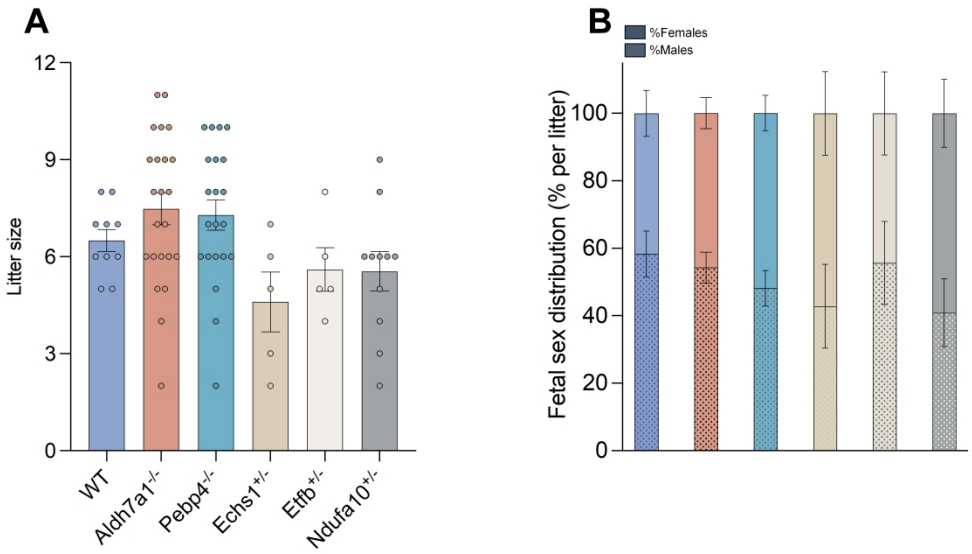

Figure S3: Knockout mouse model pregnancy outcomes. Following successful births, the (A) litter size and (B) foetal sex distribution was recorded.

238x150mm (500 x 500 DPI)

Accepted Manuscript published as REP-25-0105.R1. Accepted for publication: 28-Apr-2025

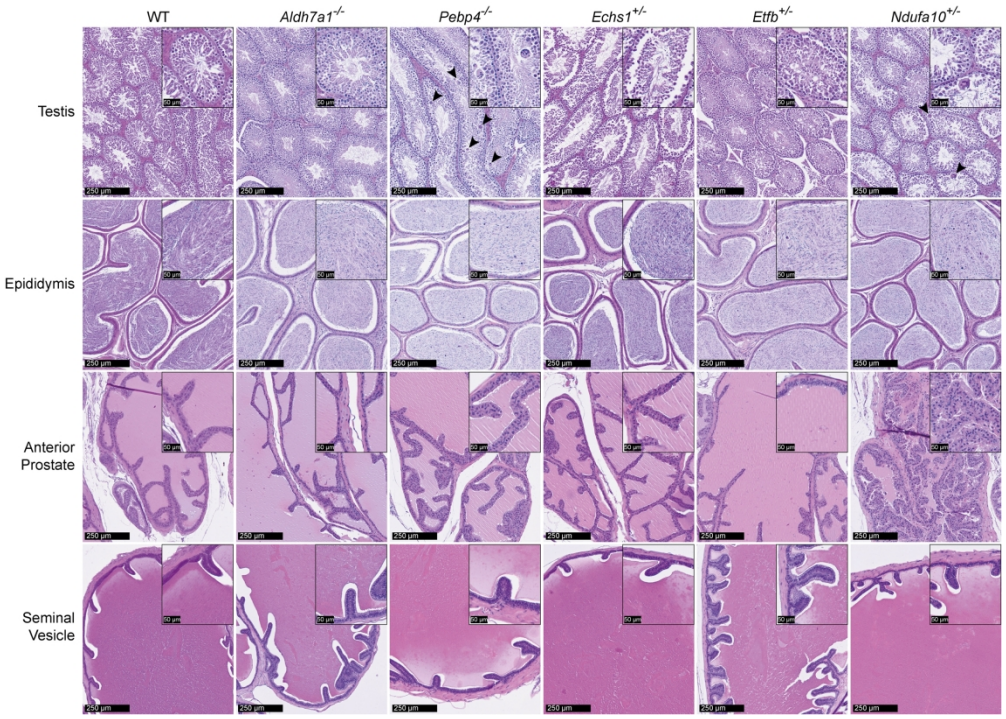

Figure S4: Histopathology of five KO models across the male reproductive tract. Representative images of H&E-stained sections from testis, cauda epididymis, anterior prostate and seminal vesicles of controls and knockout mice for the following genes: Aldehyde dehydrogenase 7 family member A1 (Aldh7a1), Phosphatidylethanolamine binding protein 4 (Pebp4), Enoyl-CoA hydratase, short chain 1 (Echs1), Electron transfer flavoprotein subunit beta (Etfb), and NADH:ubiquinone oxidoreductase subunit A10 (Ndufa10). Mice were 16 weeks old in all lines except for Aldh7a1 (19 weeks old). Multinucleated giant cells (MGCs) are indicated by arrowheads. Scale bars are 250 μm and 50 μm.

210x150mm (350 x 350 DPI)