CSIRO PUBLISHING

VIEWPOINT
https://doi.org/10.1071/RD26034

Reproduction, Fertility and Development

Your field needs your data! Raising the standard of quality control and data accessibility in reproductive proteomics

David A. Skerrett-Byrne[A,B,C,D,E,F,*]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
David A. Skerrett-Byrne
Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
Email:
David.Skerrett-Byrne@helmholtz-munich.de

Handling Editor:
Graeme Martin

Received: 5 February 2026
Accepted: 18 March 2026
Published: 20 April 2026

Cite this: Skerrett-Byrne DA (2026) Your field needs your data! Raising the standard of quality control and data accessibility in reproductive proteomics. *Reproduction, Fertility and Development* **38**, RD26034. doi:10.1071/RD26034

ABSTRACT

The field needs your data. Despite rapid progress in reproductive proteomics, a major barrier to scientific advancement remains the limited availability and transparency of proteomic datasets. Although more than 2000 sperm proteomics studies are indexed on PubMed, fewer than 414 datasets have been deposited in ProteomeXchange, leaving the majority of published findings effectively inaccessible for reanalysis. This Viewpoint highlights the urgent need for improved data stewardship, standardised quality control and open access to raw mass spectrometry files across reproductive biology. In this article, I outline how transparent false discovery rate control, true biological replication and clearly defined quantitative thresholds are essential for generating robust and interpretable proteomic outputs. I further discuss how interactive data platforms, such as ShinyApps, can substantially improve the accessibility and usability of these complex reproductive proteomic datasets. Using recent examples, I demonstrate how public data reanalysis can uncover species-conserved pathways, improve proteome coverage, validate biological functions and enable new discoveries and insights far beyond the aims of the original studies. Finally, I present a practical roadmap for authors, reviewers and journals to ensure that reproductive proteomics embraces the FAIR data principles, and moves towards a culture where sharing raw data, comprehensive metadata and interactive applications becomes standard practice. To support implementation, a concise checklist is provided to summarise key criteria for data availability, quality control and metadata reporting. Improving data accessibility and quality will not only strengthen individual studies, but will accelerate discovery and create a more robust, connected and future-proof foundation for reproductive biology.

Keywords: data reuse, data stewardship, open science, publicly accessible data, quality control, reproductive proteomics, roadmap, Shiny app.

The data gap in reproductive proteomics

Over the past two decades, proteomics has become a central discovery platform in reproductive biology, with rapid expansion across species, tissues and biological contexts. Large-scale proteomic studies now underpin fundamental biology advancements in infertility diagnostics (Skerrett-Byrne *et al.* 2022; Kanaka *et al.* 2023; Kong *et al.* 2025), assisted reproductive technologies (ART) (Azpiazu *et al.* 2014; Potiris *et al.* 2024), contraceptive development (Zhang *et al.* 2024; Skerrett-Byrne *et al.* 2025*a*) and comparative reproductive biology (Bayram *et al.* 2016; Pini *et al.* 2025), spanning samples such as spermatozoa (Skerrett-Byrne *et al.* 2022), testes (Lawrence *et al.* 2025), oocytes (Pfeiffer *et al.* 2015), endometrium (Manousopoulou *et al.* 2019) and placenta (Bandres-Meriz *et al.* 2024). This growth reflects advancements in mass spectrometry (MS) technology and increased sensitivity (Liang *et al.* 2025), improved bioinformatical tools (Lou *et al.* 2023; Wang *et al.* 2025) and a broad recognition that post-transcriptional regulation is critical to reproductive function (Skerrett-Byrne *et al.* 2025*a*).

Despite this expansion, the reproductive proteomics literature remains strikingly data-poor. As an example, while 2069 sperm proteomics studies are indexed in PubMed, fewer than 414 corresponding datasets have been deposited in ProteomeXchange partner repositories such as PRIDE (PRoteomics IDEntifications) (Fig. 1*a*). Consequently, more

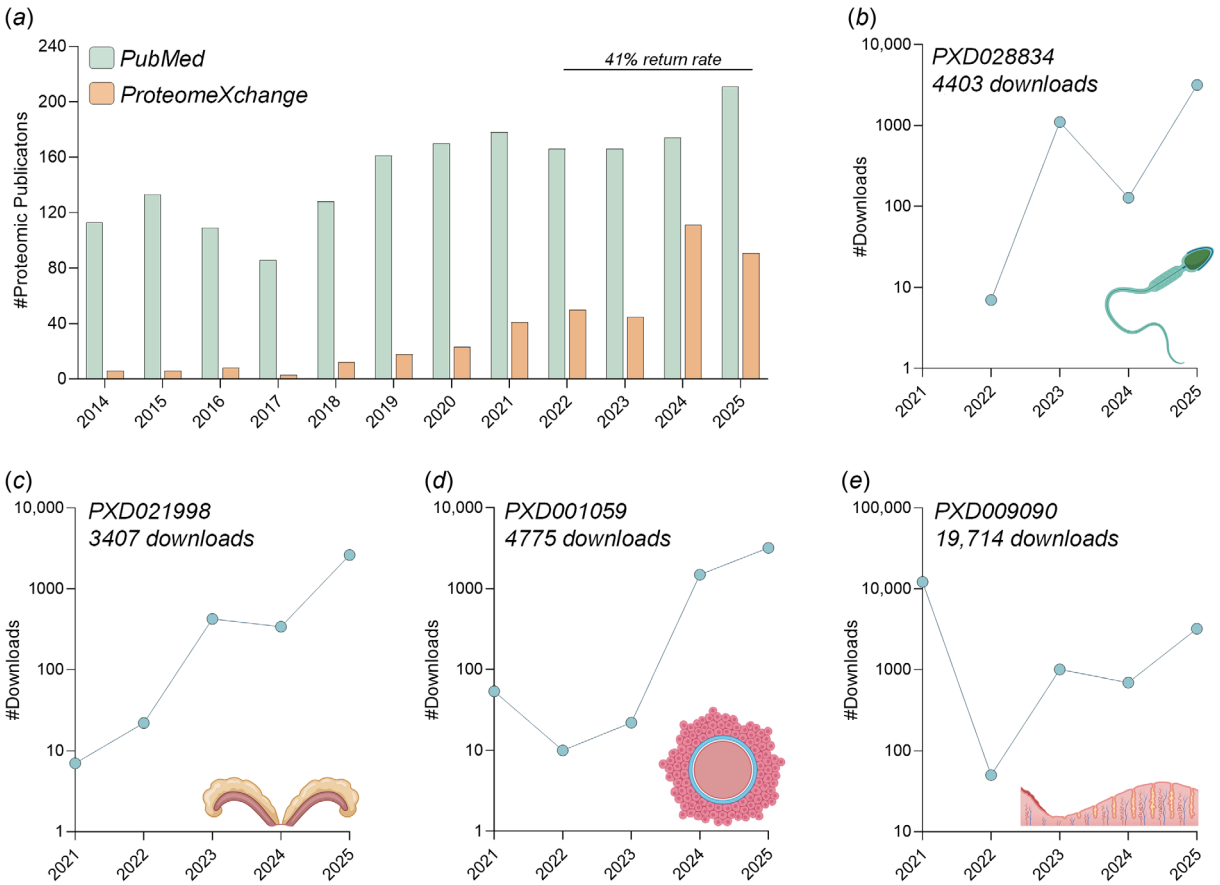

Fig. 1. (*a*) Timeline of reproductive proteomics publications (2014–2025, accessed 19th January 2026), green indicating searches on PubMed and orange for searches on PRIDE. Annual PRIDE downloads (2021–2025) of raw MS data from reproductive proteomic studies of (*b*) spermatozoa (Skerrett-Byrne *et al.* 2022), (*c*) seminal vesicles (Skerrett-Byrne *et al.* 2021), (*d*) oocytes (Pfeiffer *et al.* 2015), and the (*e*) endometrium (Manousopoulou *et al.* 2019).

than 80% of published studies cannot currently be reanalysed, compared across laboratories, or integrated into larger cross-species or meta-analytic frameworks. While this Viewpoint draws heavily on sperm proteomics as a case study, the same challenges apply across the female and male reproductive tracts, where raw data availability and metadata completeness are similarly inconsistent. ProteomeXchange and its partner repositories such as PRIDE, were established specifically to address these challenges, providing a stable, FAIR (Findable, Accessible, Interoperable, Reusable) compliant infrastructure for long-term storage, discovery and reuse of MS data (Wilkinson *et al.* 2016; Perez-Riverol *et al.* 2024). Yet uptake within reproductive biology lags behind what is seen in other biomedical domains, limiting reproducibility and long-term scientific value. The consequences are by no means trivial, with missing data preventing meaningful comparison between studies, apparent species- or condition-specific proteins cannot be reliably validated, and valuable datasets are frequently lost when laboratories close, personnel move on, or local storage fails.

Encouragingly, data deposition practices are improving (Fig. 1*a*). The proportion of reproductive proteomics studies depositing RAW data has increased steadily over time, with 2024 showing the highest annual deposition rate to date (63.8%), and an average 41% return rate between 2022 and 2025 (Fig. 1*a*). However, without stronger community norms and journal-level requirements, this progress remains incremental. Closing the data gap will be essential if reproductive proteomics is to mature from a collection of isolated protein lists into a cumulative, integrative and future-proof discipline. With this in mind, this Viewpoint covers the following key areas: the importance of open access to raw data; quality control in proteomics; the role of interactive data platforms; and lastly, a roadmap for high-quality reproductive proteomics.

Quality control in proteomics: a minimum standard

Proteomics offers unparalleled power to interrogate reproductive biology, yet this power carries a responsibility to apply rigorous and transparent quality control (QC). Without

clearly defined analytical standards, large protein lists can unintentionally create a false sense of biological insight and promote over-interpretation. As reproductive proteomics continues to expand across reproductive tissues, cells and species, inconsistent QC practices increasingly undermine comparability and reproducibility.

At a minimum, confidence in protein identification depends on stringent database search parameters and false discovery rate (FDR) control. Given the high level of accuracy and sensitivity in contemporary MS technology, database searches should employ appropriately stringent mass tolerances that reflect instrument performance and acquisition mode. Typically, these tolerances should be approximately 10 ppm (parts per million) and 0.05 Da (Daltons) for MS1 precursors and MS2 fragments, respectively. Likewise, protein, peptide and peptide-spectra match (PSM) identifications should be filtered at least ≤ 1% FDR, applied consistently and reported transparently. Relaxing on the parameters assigned to these thresholds may inflate protein numbers, but they do so at the cost of confidence, particularly in studies making claims about rare reproductive tissues/cells, condition-specific or species-specific proteins.

The quality of protein annotation further shapes biological interpretation, especially in non-model organisms (Pini *et al.* 2025). Reliance on outdated or poorly curated proteomes can exaggerate apparent novelty and inflate claims of 'unique' proteins. When analysed under consistent QC frameworks and updated annotated databases (please see NCBI and UniProt), many such identifications are revealed to be conserved homologues or artefacts of limited depth or annotation gaps. Cross-species studies therefore require orthologue mapping strategies to avoid false absences and misleading biological conclusions (Pini *et al.* 2025).

Beyond identification confidence, biological replication remains a frequently overlooked component of QC. Some reproductive proteomic studies continue to rely on single biological samples that are repeatedly measured and treated as independent replicates. Within each experimental group (e.g. control and treated), biological replicates must represent independent biological units, typically different individuals rather than repeated measurements of the same individual under identical conditions. Sampling the same individual is of course appropriate when examining temporal or conditional effects on that one individual. While practical constraints are acknowledged, such as rare species (Nixon *et al.* 2019), insufficient biological replicates fundamentally limits statistical power, obscures biological variability and restricts interpretability.

Finally, clear and transparently reported quantitative criteria are equally important in defining your true list of proteins. I contend that each identified protein should contain a quantitative value in at least 80% of your biological replicates (e.g. four out of five biological replicates) to be included. Next, differential abundance claims should be supported by defined cut-offs applied across this substantial proportion of biological replicates. At minimum a threshold for significant dysregulation should be a fold change of ± 1.5 (or $-0.585 \leq \text{Log}_2\ \text{FC} \geq 0.585$), supported by testing such as *P*-value or ANOVA where appropriate, with a minimum of ≤ 0.05 (or $-\text{Log}_{10}$ *P*-value ≥ 1.3). However, fold change thresholds should be interpreted in the context of absolute protein abundance, as modest fold changes in highly abundant proteins may represent substantial biological shifts, whereas large fold changes in low-abundance proteins may reflect limited quantitative impact (Kammers *et al.* 2015). Pascovici *et al.* (2016) further highlight that conventional multiple-testing corrections can be overly stringent in proteomics and may mask genuine biologically meaningful differences, particularly in low-power experimental designs.

Together, these considerations define the foundational analytical standard required for reproducible and interpretable reproductive proteomics. Adoption of transparent, consistent QC frameworks will not restrict discovery, instead, it will ensure that reported findings are robust, comparable and biologically meaningful across the increasingly diverse landscape of reproductive proteomics.

The role of interactive data platforms

Even when proteomic datasets are publicly available, their practical utility is often limited by how they are presented. Static supplementary tables, frequently large, fragmented and poorly annotated, remain the dominant mode of data dissemination. While sufficient for archival purposes, these formats create high barriers to exploration, comparison and reuse, particularly for researchers without specialised bioinformatics expertise. As reproductive proteomics datasets grow in size and complexity, new approaches are required to make shared data genuinely accessible and intuitively usable.

Interactive data platforms, such as Shiny web-based applications or similar frameworks, provide a powerful solution to this challenge. Recent examples from reproductive biology illustrate what becomes possible when open data are paired with interactive visualisation. ShinySperm (Skerrett-Byrne *et al.* 2024), for instance, demonstrate how publicly available reproductive proteomics datasets can be transformed into intuitive, searchable resources that support hypothesis generation. Rather than treating datasets as fixed end products, these platforms allow researchers to interrogate whole protein lists by functional category (e.g. kinases) or subcellular localisation (e.g. acrosome), with live visualisation of filtered proteomic data (e.g. volcano plots and heatmaps). This enables users to tailor exploration directly to their specific biological questions. Extending this concept, ShinyEpididymis (Trigg *et al.* 2026), integrates several independent datasets spanning epididymosomes, epididymal spermatozoa and tissue, enabling cross-study comparisons that would be more impractical using traditional supplementary materials alone. The inclusion of features such as 'snapshot' allows researchers to query large

lists of proteins/genes at once, and assess their reported detection across all available datasets. Importantly, these principles are not limited to proteomics. Similar interactive frameworks have now been applied to sperm epigenome datasets, demonstrating the broader applicability of open, platform-based data exploration across reproductive omics (Skerrett-Byrne *et al.* 2025*b*; Laurent *et al.* 2026). Together, these tools enhance transparency, reproducibility and longevity by allowing the community to explore exactly the datasets underpinning published conclusions.

Beyond technical accessibility, interactive platforms democratise data use across the reproductive biology community. Clinicians, veterinarians, wildlife biologists and fundamental biology researchers can all interrogate the same datasets without requiring bespoke pipelines or advanced coding skills. This shared interface fosters cross-disciplinary engagement and maximises the return on investment from often scarce or ethically sensitive biological samples. As reproductive proteomics continues to scale, journals and funders should consider encouraging the integration of interactive data resources alongside manuscripts. When coupled with open access to raw data and transparent QC, such platforms transform proteomic datasets from static supplements into living community resources, capable of evolving, being reanalysed and generating insights well beyond the scope of the original study.

The importance of open access to raw data

Open access to raw MS data represents the single most effective intervention for improving rigour, transparency and long-term value in reproductive proteomics. While processed protein lists and summary tables capture the conclusions of a study, only raw data preserve the full informational content required for independent validation, reanalysis and future reuse. Without access to raw MS files, reproducibility is inherently limited. Public repositories such as ProteomeXchange and its partner databases, including PRIDE, provide a mature, FAIR-compliant infrastructure for the deposition of raw MS files alongside processed outputs and metadata. When fully utilised, these resources enable datasets to be reanalysed using updated software, improved databases and more stringent quality control frameworks, often yielding increased proteome coverage, improved quantification and reduced false discovery rates compared with the original analyses.

The impact of raw data accessibility is already evident in large-scale reanalysis efforts such as ShinySpermKingdom (Pini *et al.* 2025). This resource was only possible because the underlying datasets included accessible raw files with sufficient metadata to support uniform reprocessing. Reanalysis of 29 publicly available sperm proteomic datasets, amounting to > 2 TB of raw MS data, using contemporary stringent pipelines substantially improved proteome coverage across several species (> 13,800 proteins), identified a core set of 135 conserved proteins underpinning essential sperm functions, and both conserved and species-specific molecular pathways. Importantly, this work extended beyond this reanalysis by utilisation of knockout mouse models, confirming roles for these conserved proteins in motility and fertilisation capacity, as well as clinical evidence underscoring these proteins' importance. These advances extend well beyond the original aims of the contributing studies, illustrating the compounding scientific value of open data. Furthermore, utilising PRIDE's project usage metrics, several reproductive proteomic datasets show substantial growth in data downloads. These include datasets from spermatozoa (Skerrett-Byrne *et al.* 2022) (Fig. 1*b*), seminal vesicles (Skerrett-Byrne *et al.* 2021) (Fig. 1*c*), oocytes (Pfeiffer *et al.* 2015) (Fig. 1*d*) and endometrium (Manousopoulou *et al.* 2019) (Fig. 1*e*), each recording a > 64-fold increase in downloads from 2022 to 2025, demonstrating accelerating uptake of reproductive proteomics data across and beyond the field.

Conversely, the absence of raw data imposes hard limits on what the field can achieve. Studies that provide only filtered protein tables cannot be meaningfully reanalysed, compared across platforms or integrated into meta-analyses. Incomplete metadata, such as missing information on species, tissue origin, enrichment strategies, acquisition mode, instrument model or fragmentation settings, further erodes data utility, even when raw files are available. The Sample and Data Relationship Format (SDRF), introduced by the proteomics community, provides a practical solution by standardising experimental metadata and explicitly linking samples, variables and raw data files in a reusable format (Dai *et al.* 2021).

An important consideration for human studies is the growing recognition of Indigenous data sovereignty, where biological samples and associated data are regarded as belonging to Indigenous peoples and their communities rather than individual researchers or institutions (Garrison *et al.* 2019). In such cases, culturally appropriate governance frameworks developed through meaningful engagement with Indigenous communities and Indigenous scientists, alongside community consent processes, may limit unrestricted public deposition of datasets in international repositories, highlighting the need to balance open science with respect for Indigenous data governance principles (reviewed in detail by Hudson *et al.* 2020) and the CARE principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) by Carroll *et al.* (2023).

Mandatory deposition of raw MS files, accompanied by processed outputs and comprehensive metadata (SDRF), would immediately elevate standards across reproductive proteomics. Such datasets represent 'living' resources from which we can continuously yield new insights (Dai *et al.* 2024; Pini *et al.* 2025), enabling robust cross-study comparisons spanning species, tissues and developmental contexts. For rare human samples, endangered species or longitudinal clinical cohorts, failure to share raw data represents an irreversible loss of biological information. Embracing routine raw data

deposition is therefore not merely a technical consideration, but a necessary cultural shift toward cumulative, sustainable and integrative reproductive science.

A roadmap for high-quality reproductive proteomics

Improving data quality and accessibility in reproductive proteomics does not require new technologies or extensive infrastructure. Instead, it requires clear expectations, shared standards and consistent enforcement across authors, reviewers and journals. The following practical recommendations outline achievable steps that would substantially strengthen rigour, reproducibility and long-term data value across the field.

For authors, raw MS files, processed outputs should be deposited in a ProteomeXchange partner repository prior to manuscript submission, accompanied by comprehensive metadata describing species, tissue, experimental design, acquisition parameters and analysis pipelines. Studies should employ true biological replication and apply stringent and transparent QC thresholds. Where feasible, authors should consider providing interactive data resources to facilitate exploration and reuse for the community (Skerrett-Byrne *et al.* 2024, 2025*b*; Pini *et al.* 2025).

For journals, mandatory deposition of raw data with private reviewer login details during peer review should become standard practice. Requiring structured QC and metadata checklists, covering FDR control, replication, quantitative thresholds and metadata completeness, would support reviewers and promote consistency across publications. Journals are also well placed to encourage interactive data visualisation applications as complementary resources, particularly for large or integrative proteomic studies.

For reviewers, scrutiny should extend beyond headline protein numbers to include evaluation of QC practices, replication, metadata completeness and raw data availability. Where analytical parameters appear weak or inconsistent, reviewers should feel empowered to request clarification or reanalysis. However, meaningful assessment is only possible if data is accessible at the time of review, reinforcing the importance of journal-level deposition requirements.

Together, these actions provide a realistic roadmap toward a more transparent, reproducible and integrative reproductive proteomics ecosystem. Establishing shared expectations around data quality and accessibility will not constrain discovery, but rather, it will ensure that individual studies contribute meaningfully to a cumulative and durable scientific foundation.

Conclusion

This Viewpoint provides a clear, constructive, and forward-looking case for improving data quality and accessibility in reproductive proteomics. It advocates for shared standards across species, systems and journals, with the goal of accelerating discovery and strengthening confidence in published proteomic findings. Reproductive proteomics has benefited from extraordinary technological advances and substantial investment of time, expertise and biological material. When raw data, rigorous QC and transparent metadata are embraced as standard practice, each dataset becomes far more than a single publication – it becomes a reusable community resource that can be reanalysed, integrated and extended as new questions and tools emerge. The future of reproductive proteomics will be shaped not only by the proteins we report, but by the data we choose to share.

The field does not just need new findings. The field needs your data.

Supplementary material

Supplementary material can be accessed from the article page online.

References

Azpiazu R, Amaral A, Castillo J, Estanyol JM, Guimerà M, Ballescà JL, Balasch J, Oliva R (2014) High-throughput sperm differential proteomics suggests that epigenetic alterations contribute to failed assisted reproduction. *Human Reproduction (Oxford, England)* **29**(6), 1225–1237. doi:10.1093/humrep/deu073

Bandres-Meriz J, Sanz-Cuadrado Marta I, Hurtado de Mendoza I, Majali-Martinez A, Honeder Sophie E, Cindrova-Davies T, Birner-Gruenberger R, Dalgaard Louise T, Desoye G (2024) MCM proteins are up-regulated in placentas of women with reduced insulin sensitivity. *Bioscience Reports* **44**(10). doi:10.1042/bsr20240430

Bayram HL, Claydon AJ, Brownridge PJ, Hurst JL, Mileham A, Stockley P, Beynon RJ, Hammond DE (2016) Cross-species proteomics in analysis of mammalian sperm proteins. *Journal of Proteomics* **135**, 38–50. doi:10.1016/j.jprot.2015.12.027

Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodriguez-Lonebear D, Rowe R (2023) The CARE principles for indigenous data governance. *Open Scholarship Press Curated Volumes: Policy* **19**(31). doi:10.5334/dsj-2020-043

Dai C, Füllgrabe A, Pfeuffer J, Solovyeva EM, Deng J, Moreno P, Kamatchinathan S, Kundu DJ, George N, Fexova S, Grüning B, Föll MC, Griss J, Vaudel M, Audain E, Locard-Paulet M, Turewicz M, Eisenacher M, Uszkoreit J, Van Den Bossche T, Schwämmle V, Webel H, Schulze S, Bouyssié D, Jayaram S, Duggineni VK, Samaras P, Wilhelm M, Choi M, Wang M, Kohlbacher O, Brazma A, Papatheodorou I, Bandeira N, Deutsch EW, Vizcaíno JA, Bai M, Sachsenberg T, Levitsky LI, Perez-Riverol Y (2021) A proteomics sample metadata representation for multiomics integration and big data analysis. *Nature Communications* **12**(1), 5854. doi:10.1038/s41467-021-26111-3

Dai C, Pfeuffer J, Wang H, Zheng P, Käll L, Sachsenberg T, Demichev V, Bai M, Kohlbacher O, Perez-Riverol Y (2024) quantms: a cloud-based pipeline for quantitative proteomics enables the reanalysis of public proteomics data. *Nature Methods* **21**(9), 1603–1607. doi:10.1038/s41592-024-02343-1

Garrison NA, Hudson M, Ballantyne LL, Garba I, Martinez A, Taualii M, Arbour L, Caron NR, Rainie SC (2019) Genomic research through an indigenous lens: understanding the expectations. *Annual Review of Genomics and Human Genetics* **20**, 495–517. doi:10.1146/annurev-genom-083118-015434

Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, Anderson J, Wilcox P, Arbour L, Brown A (2020) Rights, interests and expectations: Indigenous perspectives on unrestricted access to

genomic data. *Nature Reviews Genetics* **21**(6), 377–384. doi:10.1038/s41576-020-0228-x

Kammers K, Cole RN, Tiengwe C, Ruczinski I (2015) Detecting significant changes in protein abundance. *EuPA Open Proteomics* **7**, 11–19. doi:10.1016/j.euprot.2015.02.002

Kanaka V, Drakakis P, Loutradis D, Tsangaris GT (2023) Proteomics in the study of female fertility: an update. *Expert Review Proteomics* **20**(12), 319–330. doi:10.1080/14789450.2023.2275683

Kong R, Yan K, Liu D, Wu W (2025) A comprehensive mass spectrometry-based proteomic dataset of human spermatozoa. *Scientific Data* **12**(1), 1509. doi:10.1038/s41597-025-05824-w

Laurent K, Teperino R, Hrabě de Angelis M, Skerrett-Byrne DA, Beckers J (2026) Epigenetic legacy: the role of sperm miRNAs in the paternal inheritance of diabetes and obesity development. *Diabetes/Metabolism Research and Reviews* **630**(8017), 720–727. doi:10.1038/s41586-024-07472-3

Lawrence BM, O'Donnell L, Gannon AL, Skerrett-Byrne DA, Parameswaran S, Abbott I, Smith S, Handelsman DJ, Rebourcet D, Smith LB (2025) Functional analysis of HSD17B3-deficient male mice reveals roles for HSD17B7 and HSD17B12 in testosterone biosynthesis. *Endocrinology* **166**(6), bqaf078. doi:10.1210/endocr/bqaf078

Liang J, Tian J, Zhang H, Li H, Chen L (2025) Proteomics: an in-depth review on recent technical advances and their applications in biomedicine. *Medicinal Research Reviews* **45**(4), 1021–1044. doi:10.1002/med.22098

Lou R, Cao Y, Li S, Lang X, Li Y, Zhang Y, Shui W (2023) Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. *Nature Communications* **14**(1), 94. doi:10.1038/s41467-022-35740-1

Manousopoulou A, Hamdan M, Fotopoulos M, Garay-Baquero DJ, Teng J, Garbis SD, Cheong Y (2019) Integrated eutopic endometrium and non-depleted serum quantitative proteomic analysis identifies candidate serological markers of endometriosis. *PROTEOMICS – Clinical Applications* **13**(3), 1800153. doi:10.1002/prca.201800153

Nixon B, Johnston SD, Skerrett-Byrne DA, Anderson AL, Stanger SJ, Bromfield EG, Martin JH, Hansbro PM, Dun MD (2019) Modification of crocodile spermatozoa refutes the tenet that post-testicular sperm maturation is restricted to mammals. *Molecular & Cellular Proteomics: MCP* **18**(Suppl 1), S58–S76. doi:10.1074/mcp.RA118.000904

Pascovici D, Handler DC, Wu JX, Haynes PA (2016) Multiple testing corrections in quantitative proteomics: a useful but blunt tool. *Proteomics* **16**(18), 2448–2453. doi:10.1002/pmic.201600044

Perez-Riverol Y, Bandla C, Kundu Deepti J, Kamatchinathan S, Bai J, Hewapathirana S, John Nithu S, Prakash A, Walzer M, Wang S, Vizcaíno Juan A (2024) The PRIDE database at 20 years: 2025 update. *Nucleic Acids Research* **53**(D1), D543–D553. doi:10.1093/nar/gkae1011

Pfeiffer MJ, Taher L, Drexler H, Suzuki Y, Makałowski W, Schwarzer C, Wang B, Fuellen G, Boiani M (2015) Differences in embryo quality are associated with differences in oocyte composition: a proteomic study in inbred mice. *PROTEOMICS* **15**(4), 675–687. doi:10.1002/pmic.201400334

Pini T, Nixon B, Karr TL, Teperino R, Sanz-Moreno A, da Silva-Buttkus P, Tüttelmann F, Kliesch S, Gailus-Durner V, Fuchs H, Marschall S, Hrabě de Angelis M, Skerrett-Byrne DA (2025) Towards a kingdom of reproductive life - the core sperm proteome. *Reproduction* **169**(6), e250105. doi:10.1530/rep-25-0105

Potiris A, Alyfanti E, Drakaki E, Mavrogianni D, Karampitsakos T, Machairoudias P, Topis S, Zikopoulos A, Skentou C, Panagopoulos P (2024) The contribution of proteomics in understanding endometrial protein expression in women with recurrent implantation failure. *Journal of Clinical Medicine* **13**(7), 2145. doi:10.3390/jcm13072145

Skerrett-Byrne DA, Anderson AL, Bromfield EG, Bernstein IR, Mulhall JE, Schjenken JE, Dun MD, Humphrey SJ, Nixon B (2022) Global profiling of the proteomic changes associated with the post-testicular maturation of mouse spermatozoa. *Cell Reports* **41**(7), 111655. doi:10.1016/j.celrep.2022.111655

Skerrett-Byrne DA, Anderson AL, Teperino R, Burke ND, Bromfield EG, Dun MD, Gailus-Durner V, Fuchs H, Marschall S, Hrabě de Angelis M, Humphrey SJ, Nixon B (2025*a*) The functional maturation of mouse spermatozoa is underpinned by global remodeling of the cellular phosphoproteome. *bioRxiv* 2025.2009.2018.677062. doi:10.1101/2025.09.18.677062

Skerrett-Byrne DA, Pepin AS, Laurent K, Beckers J, Schneider R, Hrabě de Angelis M, Teperino R (2025*b*) Dad's diet shapes the future: how paternal nutrition impacts placental development and childhood metabolic health. *Molecular Nutrition & Food Research* **69**(22), e70261. doi:10.1002/mnfr.70261

Skerrett-Byrne DA, Teperino R, Nixon B (2024) ShinySperm: navigating the sperm proteome landscape. *Reproduction, Fertility and Development* **36**, RD24079. doi:10.1071/rd24079

Skerrett-Byrne DA, Trigg NA, Bromfield EG, Dun MD, Bernstein IR, Anderson AL, Stanger SJ, MacDougall LA, Lord T, Aitken RJ, Roman SD, Robertson SA, Nixon B, Schjenken JE (2021) Proteomic dissection of the impact of environmental exposures on mouse seminal vesicle function. *Molecular & Cellular Proteomics: MCP* **20**, 100107. doi:10.1016/j.mcpro.2021.100107

Trigg NA, Mulhall JE, Nixon B, Laurent K, Pai S, Smyth SP, Burke ND, Beckers J, Bromfield EG, Karr TL, Pleuger C, Schjenken JE, Hrabě de Angelis M, Teperino R, Skerrett-Byrne DA (2026) Curating fertility – proteomic remodelling of sperm during epididymal transit. *Reproduction* 171(2), xaag012. doi:10.1093/reprod/xaag012

Wang J, Huang Y, Lu F, Xu Q, Yang Z, Jiang Y, Shi S, Pan J, Yang Y, Fang Q (2025) Benchmarking informatics workflows for data-independent acquisition single-cell proteomics. *Nature Communications* **16**(1), 10276. doi:10.1038/s41467-025-65174-4

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. doi:10.1038/sdata.2016.18.

Zhang X, Tu H, Zhou X, Wang B, Guo Y, Situ C, Qi Y, Li Y, Guo X (2024) Quantitative phosphoproteomic profiling of mouse sperm maturation in epididymis revealed kinases important for sperm motility. *Molecular & Cellular Proteomics* **23**(8), 100810. doi:10.1016/j.mcpro.2024.100810

Data availability. This article is a viewpoint review of published literature. Consequently, there is no new data relevant to this article.

Declaration of use of AI. ChatGPT (OpenAI) was used to assist in the conceptual development of a schematic figure summarising the manuscript. All scientific content, interpretation and final design were generated, reviewed and approved by the author. The author takes full responsibility for the accuracy and integrity of the work.

Conflicts of interest. David Skerrett-Byrne is an Editor of Reproduction, Fertility and Development but was not involved in the peer review or decision-making process for this paper.

Declaration of funding. National Health and Medical Research Council (NHMRC) Emerging Leadership Fellowship (APP2034392) awarded to D.A.S.B.

Acknowledgements. I thank all my co-authors, mentors, and colleagues over the years to build these amazing databases and advancing our field through proteomic approaches. In particular, I want to acknowledge Dr Heather Murray, Dr Nat Trigg, Jess Mulhall, Dr Liz Bromfield, Dr Nathan Burke, Dr Shannon Smyth, Amanda Anderson, Dr Sean Humphrey, Dr Taylor Pini, Dr Raf Teperino and Prof Brett Nixon for their central roles in generating, curating and interpreting the datasets discussed in this Viewpoint. I am especially grateful to Dr John Schjenken for our many insightful discussions on reproductive proteomics and the importance of developing community standards.

Author affiliations

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

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

[C]Reproductive and Family Health Research Program, Hunter Medical Research Institute, Awabakal Country, New Lambton Heights, NSW, Australia.

[D]School of Biomedical Sciences and Pharmacy, College of Health, Medicine and Wellbeing, University of Newcastle University Drive, Callaghan, Awabakal Country, Newcastle, NSW, Australia.

[E]Mercy Perinatal, Mercy Hospital for Women, Heidelberg, Vic, Australia.

[F]Therapeutics Discovery and Vascular Function in Pregnancy Group, Department of Obstetrics, Gynaecology and Newborn Health, Melbourne Medical School, The University of Melbourne, Heidelberg, Vic, Australia.