

Stable bull fertility protein markers in seminal plasma

J. Willforss^{a,*,1}, J.M. Morrell^b, S. Resjö^c, T. Hallap^d, P. Padrik^e, V. Siino^a, D.J. de Koning^f, E. Andreasson^c, F. Levander^{a,*,1}, P. Humblot^{b,**,1}

^a Department of Immunotechnology, Lund University, Lund, Sweden

^b Department of Clinical Sciences, Division of Reproduction, Swedish University of Agricultural Sciences, Uppsala, Sweden

^c Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Alnarp, Sweden

^d Estonian University of Life Sciences, Tartu, Estonia

^e Animal Breeders' Association of Estonia, Raplamaa, Estonia

^f Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

ARTICLE INFO

Keywords:

Bos taurus
Seminal plasma
Fertility
Proteomics
Seasonal variation

ABSTRACT

Bull fertility is an important trait in breeding as the semen of one bull can, potentially, be used to perform thousands of inseminations. The high number of inseminations needed to obtain reliable measures from Non-Return Rates to oestrus creates difficulties in assessing fertility accurately. Improving molecular knowledge of seminal properties may provide ways to facilitate selection of bulls with good semen quality. In this study, liquid chromatography mass spectrometry (LC-MS/MS) was used to analyze the protein content from the seminal plasma of 20 bulls with Non-Return Rates between 35 and 60%, sampled across three seasons. Overall, 1343 proteins were identified and proteins with consistent correlation to fertility across multiple seasons found. From these, nine protein groups had a significant Pearson correlation ($p < 0.1$) with fertility in all three seasons and 34 protein groups had a similar correlation in at least two seasons. Among notable proteins showing a high and consistent correlation across seasons were Osteopontin, a lipase (LIPA) and N-acetylglucosamine-1-phosphotransferase subunit gamma. Three proteins were combined in a multiple linear regression to predict fertility ($r = 0.81$). These sets of proteins represent potential markers, which could be used by the breeding industry to phenotype bull fertility.

Significance: The ability of bull spermatozoa to fertilize oocytes is crucial for breeding efficiency. However, the reliability of this trait from field measures is relatively low and the prediction of fertility given by conventional methods to evaluate sperm quality is currently not very accurate. In this work, we identify sets of proteins in bull seminal plasma from repeated samples collected at different times of the year that correlate to fertility in a consistent way. We combined these individual proteins to build a molecular signature predictive of fertility. This study provides an overview of proteins linked to fertility in seminal plasma, thereby increasing knowledge of the bull seminal plasma proteome. Protein signatures from the latter, potentially related to fertility, may be of use to predict fertility for individual bulls.

1. Introduction

Bull fertility is a crucial trait for breeding companies, with a considerable impact on the dairy herd economy. In cattle, bull fertility is usually determined from the results of artificial inseminations (AI), estimated from Non-Return Rates to oestrus (i.e. the proportion of cows not coming back in heat after insemination and therefore deemed to be pregnant). Due to numerous confounding factors and biases related

mostly to the type of female inseminated and the effects of the environment, high numbers of AIs are necessary to obtain reliable estimates for bull fertility [1]. This was previously a source of limitation for the breeding companies as reliable fertility estimates were obtained from accumulating data over several years, by which time the bull was no longer in use for AI. Thus, it was difficult to define efficient sperm handling or genetic selection strategies. In the context of genomic selection where cohorts of young bulls are used, the number of AIs per

* Correspondence to: Willforss J and Levander F, Department of Immunotechnology, Lund University, Box 118, 221 00 Lund, Sweden.

** Correspondence to: Humblot P, Institution for Clinical Sciences, Swedish University of Agricultural Sciences, Box 7054, 750 07 Uppsala, Sweden.

E-mail addresses: jakob.willforss@immun.lth.se (J. Willforss), fredrik.levander@immun.lth.se (F. Levander), patrice.humblot@slu.se (P. Humblot).

¹ Visiting address: Medicon Village, Scheelevägen 2, Lund

individual are reduced compared to previously. However, since semen from some bulls with a superior genetic potential can still be used for thousands of AI, it is crucial to maintain high fertility standards even for younger bulls. However, for these young individuals, only limited data from AIs are available. This results in fertility estimates of poor predictive value. Identifying key markers of fertility in semen may allow for better reproducibility in handling bulls and bull spermatozoa, and in some cases, improved fertility. They could also be used as phenotypic markers for fertility in genomic selection schemes to avoid possible unfavorable evolution of sire fertility. In the past, differences in fertility between males were often attributed to variation in sperm quality [2]. However, today, the predictive value of conventional sperm quality traits is usually considered to be low [3].

Seminal plasma (SP) has been shown to have an assisting role for spermatozoa and fertility, improving sperm longevity and ability to move within the female tract [4]. In cases where the difference in fertility could not be linked to sperm quality, differences in SP have been observed with effects on the spermatozoa and the genital tract of the female [5,6]. Other effects that have been shown to be associated with seminal plasma proteins include sperm motility, resilience to freezing and survival in the genital tract in the sheep [7,8]. Following the pioneering work of Kilian [9], some specific candidates such as Osteopontin have been associated with fertility [10]. Nowadays, there is an increased interest in seminal plasma proteomics based on global approaches of protein patterns. For instance, in two recent studies, comprehensive profiling of the bull seminal plasma proteomics were established [11,12]. However, analyses were limited to a single time point measurement, and thus did not assess the stability of the markers over time. In another study, proteomic patterns were used to identify the role of seminal plasma proteins added to spermatozoa by comparing epididymal spermatozoa to ejaculated spermatozoa but they were not specifically linked to fertility [13].

Here, we describe proteomic patterns in seminal plasma from bulls of different fertility, using label-free LC-MS/MS measurements carried out on semen samples collected in different seasons. We provide a general proteomic overview, increasing the molecular knowledge of the bull SP proteome. Identifying proteins that are stable in repeated samples obtained in different seasons and correlated to fertility could allow their use as biomarkers of fertility.

2. Methods

2.1. Sample collection and processing

Semen ejaculates (total $n = 109$) were obtained from 20 Holstein bulls at an AI center (Animal Breeders' Association of Estonia, Rapla-maa, Estonia, 58°56'11.4"N 24°53'14.5"E). Ejaculates were collected at three time points: Winter (2016 / 2017), Spring (2017), and Summer (2017). Temperature humidity indices (THI) calculated for the summer indicated no risk for heat stress (temperatures and humidity indices for the collection dates are available in Supplementary Materials S1). Seventeen of these bulls were sampled in all three seasons. At each time point, two ejaculates were collected two weeks apart to allow the study of the stability of proteomic patterns within and through seasons to be assessed. Three samples were lost in the sample handling, leading to a single ejaculate for a certain bull and season. The distribution of samples across season, ejaculates, repeated injections and birth dates are described in the table Supplementary Materials S2 and summarized in Fig. 1. The fertility performance of these bulls (as estimated by non-return rates at 90 days post-AI, based on the outcome from the first insemination) ranged from 35% to 60% based on a total number of inseminations per bull between 42 and 9384 (median 643.5). These values are presented in the table Supplementary Materials S3. Samples from a separate set of Holstein bulls from the same AI center ($n = 17$) were collected in 2018 as a separate cohort for comparison. Estimated non-return rates for these bulls ranged between 34% and 51% based on 316 to 9244 inseminations per bull (median 989).

Bulls were prepared for semen collection by allowing them to follow each other around a circular race one by one for 30–40 min before collection, and making false mounts, to increase libido. Approximately 70% of bulls false mounted once, and 30% twice. The first ejaculates were collected for the experiment. Following collection, ejaculates were centrifuged at 1800g for 10 min to pellet the spermatozoa [2] and aliquots of the resulting SP samples (5 μl) were then deep-frozen ($-196\text{ }^\circ\text{C}$) and later transferred on dry ice to the laboratory for proteomic analyses. Protein extraction was performed by mixing 1 μl of SP with 20 μl of SDS-PAGE sample buffer with DDT (58 mM Tris, pH 6.8, 5% glycerol, 1.7% SDS, 10 mM DDT) followed by a 10 min incubation at $95\text{ }^\circ\text{C}$. Subsequently, iodoacetamide was added to a final concentration of 55 mM and the samples were incubated for 30 min in the dark. The samples were then analyzed on SDS-PAGE until the front had migrated approximately 5 mm into the gel. The pieces were excised, washed and the protein was digested with trypsin in-gel. The tryptic peptides were cleaned with C18 columns (UltraMicro spin columns, Nest Group Inc.,

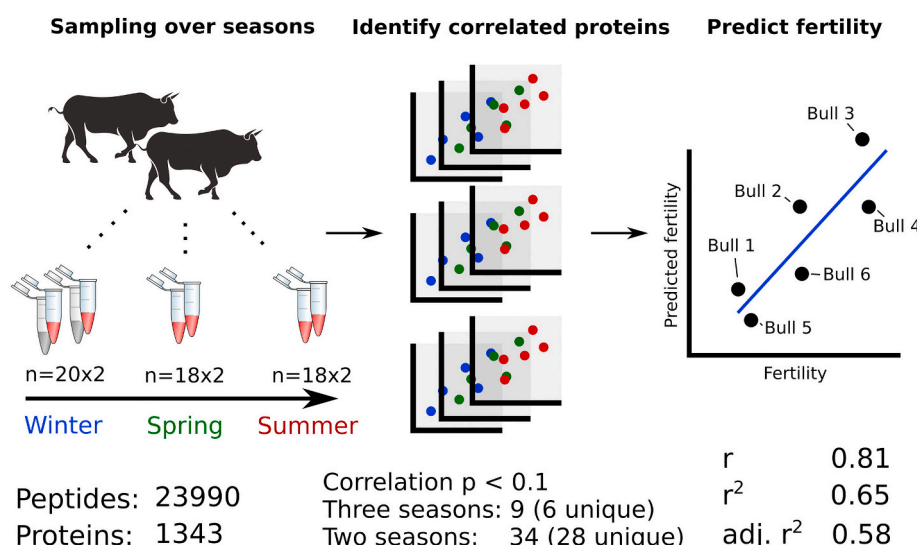


Fig. 1. Study overview. Samples were taken at three different time points from the same set of bulls. For each round, two ejaculates were taken. For the first season, double injections were performed on the mass spectrometer. Next, proteins were selected for having a consistent Pearson correlation across two or three of the seasons ($p < 0.1$). Finally, an optimal subset of these was used to produce a predictor.

Southborough, MA, USA) and resuspended in a solution of 5% formic acid in water.

2.2. Mass spectrometry data acquisition

LC-MS/MS data for the seasonal protein digests were analyzed on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Germany) interfaced with Easy-nLC liquid chromatography system (both Thermo Fisher Scientific, Waltham, MA, USA). Peptides were analyzed using NanoViper Pepmap pre-column (100 $\mu\text{m} \times 2\text{ cm}$, particle size 5 μm , Thermo Fischer Scientific) and an in-house packed analytical column (75 $\mu\text{m} \times 30\text{ cm}$, particle size 3 μm , Reprosil-Pur C18, Dr. Maisch) using a linear gradient from 5% to 35% B over 100 min followed by an increase to 100% B for 5 min, and 100% B for 15 min at a flow of 300 nL/min. Solvent A was 0.2% formic acid in water and solvent B was 0.2% formic acid in 80% acetonitrile.

Precursor ion mass spectra were acquired at 120 K resolution and MS/MS analysis was performed in a data-dependent mode where the 5 most intense precursor ions were selected for fragmentation using CID at a collision energy of 35. Charge states 2 to 4 were selected for fragmentation, and dynamic exclusion was set to 60 s.

For the set of samples corresponding to the first season, replicate injections per sample was performed to assess the degree of technical variation from the mass spectrometer. Four samples from the first seasonal sampling were rerun together with the samples from the second season to assess how much technical variation came from the downstream processing of the samples.

Peptide digests from the follow-up bulls ($n = 17$) were analyzed by LC-MS/MS using an Easy-nano LC system (Thermo Fisher Scientific, Germany) coupled with a QExactive HF-X mass spectrometer (Thermo Fisher Scientific, Germany) operating in positive ion mode for data dependent acquisition (DDA). The analytical column was 15 cm long fused silica capillary (75 μm 16 cm Pico Tip Emitter, New Objective), packed in house with C18 material Reprosil-Pur 1.9 μm (Dr. Maisch GmbH, Germany). Peptides were separated by an 80 min gradient from 5% to 90% solvent B (80% ACN, 0.1% FA) at a constant flow rate of 250 nL/min. The Orbitrap acquired the full MS scan with automatic gain control (AGC) target value of 3×10^6 ions and a maximum fill time of 50 ms. The 20 most abundant peptide ions were selected from the MS for higher energy collision-induced dissociation (HCD) fragmentation (collision energy: 40 V). Fragmentation was performed at 15,000 FWHM resolution for a target of 1×10^5 and a maximum injection time of 20 ms using an isolation window of 1.2 m/z . The software Xcalibur v3.0 (Thermo Fisher Scientific, Germany) was used both to control the nLC system and the MS. Here, it was used to acquire and visualize the raw data.

2.3. Data processing and analysis

Raw MS data were converted to mzML, an open format for mass spectrum, using Proteowizard version 3.0.9220 (3.0.11841 for the follow-up data; [14] with MS-Numpress compression [15], and MS1 peptide features were extracted using Dinosaur [16] version 1.1.4.. Data processing to a peptide abundance matrix was performed in the Proteios software environment ([17], 2.20.0-dev build 4643). X!Tandem ([18], version 2017.2.1.4) and MS-GF+ ([19], v2016.06.29) were used for peptide spectrum matching. Precursor mass error tolerance was set to 7 ppm and for X!Tandem fragment monoisotopic mass error was set to 0.4 Da (0.02 Da for the QExactive data), while for MS-GF+ the instrument setting was used to set the fragment tolerance. Fix carbamidomethylation of cysteine, and variable oxidation of methionine was set as modifications, and protein N-terminal acetylation was set as additional potential modification in MS-GF+. As a search database, proteins annotated as *Bos taurus* were retrieved from UniProt (2017-05-19) and concatenated with the cRAP database containing common contaminant proteins (version 1.0, 116 entries). Reverse entries were added to the

database as decoy sequences, giving a total of 92,984 protein entries. The search results were subsequently combined in Proteios. The built-in algorithm in Proteios was used for matching MS1 peptide features to MS2 identifications (plugin version 0.5.1, peptide spectrum match FDR cutoff 0.001, secondary hit FDR cutoff 0.01, m/z tolerance 0.005, RT tolerance 0.1) and alignment and propagation of feature identifications between samples using default settings. The resulting peptide matrix was used for subsequent analysis. The datasets from different seasons were processed jointly to facilitate the analysis of the same bull individuals across all time points. Raw peptide abundances were log2-transformed and normalized using cyclic Loess normalization in NormalizerDE ([20], version 1.5.4), which is a normalization technique adjusting for systematic differences in abundance between samples at different abundance levels [21]. For a given season, the two measurements from the same bull were combined to provide the median value for each peptide, giving one value per bull and season. This dataset was used for season-specific studies. It was further reduced across batches, to give one value per bull, which was then used to calculate a combined per-bull view. The peptides were summarized into protein groups using an R implementation (<https://github.com/ComputationalProteomics/ProteinRollup>, version 1.0.0) of the DanteR ([22], v1.1.6970) R rollup, retaining proteins supported by at least two peptides. Protein level FDR was estimated by calculating the number of decoy entries passing through the protein rollup and was found to be <0.01 FDR (8 decoys and 1343 targets passing in the seasonal dataset, with 13 decoys and 1640 targets in the follow-up dataset). The mass spectrometry proteomics data and the protein abundance tables have been deposited to the ProteomeXchange Consortium via the PRIDE [23] partner repository with the dataset identifiers PXD021241 and PXD021245.

The dataset was further explored using OmicLoupe ([24], v0.9.0) for a comprehensive exploration of outliers and general trends. One bull was identified as a consistent outlier and was excluded from all the analyses. Further, a single sample had a distorted profile, as identified by density plots, and was also excluded (Supplementary Fig. 1 in Supplementary Materials S4 where Supplementary Figs. 1–5 are presented). The most abundant proteins were estimated within each sampling by retrieving the three peptides with the highest average intensity and calculating the average of these, as proposed by Silva [25]. The 10 most abundant proteins were collected within each seasonal sampling. A gene ontology summary was generated using a Bioconductor based database for Bovine (10.18129/B9.bioc.org.Bt.eg.db, 3.10.0) and enriched using the ClusterProfiler ([26], 3.14.3) R package using FDR < 0.1 (Benjamini-Hochberg corrected; [27]) to determine significantly enriched groups. UpSet plots were generated to show overlapping correlated proteins between seasons in using the UpSetR package [28].

Pearson correlations between protein abundance and fertility estimates were calculated in R (version 3.6.3) for proteins within each of the seasonal batches as well as the validation dataset using the `cor.test` function in the stats R package. Protein groups having a p -value < 0.1 across all three samples and detected in both datasets were selected. Correlations were also weighted for the number of inseminations from which fertility rates were obtained, but results from non-weighted correlations are presented here, as the r^2 values resulting from the two approaches were almost identical. In addition, an extended list of protein groups with Pearson correlation $p < 0.1$ in two of the three seasons with a similar correlation pattern in the third was retrieved. A list of proteins with Pearson correlation $p < 0.05$ in at least one of the seasons was included to facilitate further explorations. Pearson correlations between protein abundances and age were calculated to see if age might have a confounding effect for certain proteins. To determine if the protein differences seen might be related to freezability differences in the spermatozoa, information obtained from a recent study investigating this phenomenon in a comprehensive proteomics dataset [12] was summarized for the top-listed proteins.

For the study including repeated times / seasonal effects, repeated-measures ANOVA was carried out using the `aov` function in the stats R

package, with batch as a fixed effect and bull as a random effect, thus accounting for the individual variation in the model. The resulting p -values were adjusted to FDR values using the Benjamini-Hochberg procedure [27]. These values were filtered using $FDR < 0.05$ and a minimum log2 fold change between any two of the three groups.

Ways to build signatures combining multiple proteins or peptides were explored. Multiple linear regression models were computed using the regsubsets command from the leaps R package (version 3.1, <https://cran.r-project.org/web/packages/leaps/index.html>). The calculations were based on the proteins with Pearson correlation $p < 0.1$ across all seasons, selecting the protein variant with the lowest median p -value in cases where multiple had the same annotation, resulting in a reduced list of 6 protein groups consistently correlated across seasons, as illustrated in Table 1. This reduced list was produced by filtering redundant proteins matching to the same annotation, selecting the protein with the lowest median p -value. Individual protein abundances (previously Loess-normalized and log2 transformed on sample-basis) were normalized further to mean 0 and standard deviation 1. Subsequently, all combinations of predictors were tested systematically. The model with the highest adjusted r^2 value was selected. Further, the same procedure was carried out in the extended list of protein candidates (as illustrated in Table 2), limiting the maximum number of independent proteins to three to avoid overfitting. Finally, we explored with the same procedures a combination of proteins from the extended list showing a similar correlation pattern in the additional set of bulls, resulting in a final list of 11 proteins. The full R analysis, including code for generating most of the figures present in this article, is available as R notebook documents and HTML-files in the Supplementary Materials S5-S8.

3. Results

3.1. Seminal plasma proteome overview

An overview of the seminal plasma proteome was obtained from the analysis of proteomics data from the three seasons. Globally, 23,990 different peptides were found, representing 1343 protein groups that were each identified by two or more peptides (Fig. 1) at $FDR < 0.01$. At a quantitative level, the technical replicates run in the first of the three seasons showed high technical reproducibility in the mass spectrometry (Supplementary Fig. 2). Furthermore, in Principal Component Analysis (PCA), samples from the rerun of the first season with the second season grouped primarily with the first season, indicating that the major part of the variation between seasonal datasets is not due to the preparation process or measurement, but due to variation between biological samples (Supplementary Fig. 3). PCA from the full dataset revealed that season has the strongest main effect. A separation related to fertility measurement can be seen from further principal components (PC) with axis explaining less variation as PC7 and PC8 (Fig. 2). This indicates a systematic relationship between the fertility and the overall expression

seen in the proteomic data but also shows that this effect is overshadowed by other effects such as season of sampling and other sources of variation related to sample biological heterogeneity.

Gene ontology distributions of identified proteins were generally similar to the gene ontology distribution of the predicted *Bos taurus* proteome, with certain differences which were further highlighted by a gene ontology enrichment analysis. The highest enriched cellular component term was as expected “extracellular region part”. Further, the highest enriched biological process was “carbohydrate metabolic processes” and “hydrolase activity acting on ester bonds”. The overall distribution of gene ontology terms compared to the predicted proteome and enrichments comparing proteins observed in the seminal plasma proteome to the general background are displayed in Fig. 3.

Proteins estimated among the top 10 most abundant in at least one sampling (14 in total) are illustrated in Supplementary Fig. 4 with the top abundant proteins being Seminal plasma protein BSP-30 K, Spermadhesin-1 and Seminal plasma protein PDC-109, as shown in the table Supplementary Materials S9.

3.2. Proteins variable over seasons

To detect the proteins that were most variable between seasons, a comparison of seasonal variation using repeated measures ANOVA was performed. The analysis yielded 403 proteins when features with FDR below 0.05 and a log2 fold difference of 1 between the two groups with the greatest difference were selected. These protein groups are presented in the table Supplementary Materials S10, and the 30 top hits are illustrated in Supplementary Fig. 5.

3.3. Correlation to fertility reveals protein markers stable over seasons and biological repeated samples

Nine protein groups showed consistent correlation trends ($p < 0.1$) across the three seasons (Fig. 4), out of which 3 had a significant negative correlation with fertility estimates with $p < 0.05$, namely Placenta-expressed transcript 1 protein (PLET1), Lipase (LIPA) and *N*-acetylglucosamine-1-phosphotransferase subunit gamma (GNPTG). All of these had similar correlation trends across all three seasons. The extension of this list, made by including proteins with correlation $p < 0.1$ in at least two seasons, resulted in 34 protein groups. Out of these, all but two protein groups displayed the same correlation direction across all three seasons. These proteins were not among the most abundant ones, and the ten most abundant proteins within each season showed either a low or no correlation with fertility in subsequent tests. Principal component plots using proteins with p -value < 0.1 in at least one of the three seasonal samples show clustering in all three batches, as well as in the combined dataset. These trends are generally clear, except for bull 16. Correlation overlaps and PCAs are illustrated in Fig. 5. The overlapping correlations are illustrated as an UpSet plot (panel A) which

Table 1
Consistently correlated protein groups.

Description	S1 corr	S2 corr	S3 corr	S1 P-val	S2 P-val	S3 P-val	AgeSig	Freeze
Osteopontin	0.387	0.443	0.415	0.092	0.066	0.087	0	NonSig
Osteopontin (*)	0.518	0.419	0.423	0.019*	0.084	0.08	0	NonSig
Osteopontin	0.379	0.551	0.441	0.1	0.018*	0.067	0	NonSig
Osteopontin	0.381	0.421	0.532	0.097	0.082	0.023*	0	NonSig
Placenta-expressed transcript 1 protein	-0.491	-0.584	-0.537	0.028*	0.011*	0.021*	2	NonSig
Lipase	-0.605	-0.582	-0.548	0.005*	0.011*	0.019*	0	Missing
EGF like repeats and discoidin domains 3	-0.523	-0.444	-0.409	0.018*	0.065	0.092	0	Missing
Cartilage acidic protein 1	0.394	0.433	0.504	0.086	0.073	0.033*	0	Missing
<i>N</i> -acetylglucosamine-1-phosphotransferase subunit gamma	-0.781	-0.746	-0.641	0*	0*	0.004*	0	NonSig

List of protein groups identified as correlated with fertility (Pearson's correlation $p < 0.1$) across all three seasons. Proteins with $p < 0.05$ are indicated with *. For cases with multiple proteins matching the same annotation, the one used for building multiple linear regressions is indicated with (*). The AgeSig column indicates for how many seasons the protein was found correlated with age (Pearson, $p < 0.1$). The Freeze column indicates whether the protein was found to be linked to different freezability of sperm in a recent study by Gomes et al. [12] (NonSig = non significant, Missing = not detected).

Table 2
Protein groups with correlation across two seasons.

Description	S1 corr	S2 corr	S3 corr	S1 P-Val	S2 P-Val	S3 P-val	AgeSig	Freeze
Osteopontin	0.412	0.453	0.227	0.071	0.059	0.365	0	NonSig
Osteopontin	0.387	0.443	0.415	0.092	0.066	0.087	0	NonSig
Osteopontin	0.31	0.483	0.414	0.184	0.042*	0.088	0	NonSig
Osteopontin (*)	0.518	0.419	0.423	0.019*	0.084	0.08	0	NonSig
Osteopontin	0.379	0.551	0.441	0.1	0.018*	0.067	0	NonSig
Osteopontin	0.381	0.421	0.532	0.097	0.082	0.023*	0	NonSig
Placenta-expressed transcript 1 protein	-0.491	-0.584	-0.537	0.028*	0.011*	0.021*	2	NonSig
C-C motif chemokine	0.018	0.544	-0.586	0.94	0.02*	0.011*	1	NegRel
Lipase	-0.605	-0.582	-0.548	0.005*	0.011*	0.019*	0	Missing
C11H9ORF9 protein (Sperm acrosome associated 9)	-0.017	-0.443	-0.535	0.945	0.066	0.027*	0	Missing
LOC504248 protein	0.423	0.741	0.009	0.071	0.002*	0.973	2	NegRel
Heat shock 70 kDa protein 1A	-0.475	-0.389	-0.584	0.034*	0.111	0.011*	1	NonSig
Cullin-associated NEDD8-dissociated protein 1	0.248	0.432	0.477	0.291	0.074	0.045*	0	NegRel
Protein arginine N-methyltransferase 5	0.497	-0.117	0.478	0.026*	0.643	0.045*	0	NegRel
Prostaglandin D2 synthase 21 kDa	-0.329	-0.601	-0.432	0.157	0.008*	0.074	0	NonSig
Alpha-galactosidase	-0.436	-0.512	-0.277	0.054	0.03*	0.266	0	Missing
Olfactomedin like 2A	-0.48	-0.391	-0.421	0.032*	0.109	0.082	0	Missing
Uncharacterized protein	-0.313	-0.585	-0.437	0.179	0.011*	0.07	1	Missing
Alpha-(1,3)-fucosyltransferase	-0.537	-0.369	-0.506	0.015*	0.132	0.032*	1	Missing
EGF like repeats and discoidin domains 3	-0.523	-0.444	-0.409	0.018*	0.065	0.092	0	Missing
Vesicle-associated membrane protein 2	0.516	0.537	-0.348	0.071	0.059	0.244	0	NonSig
Cartilage acidic protein 1	0.394	0.433	0.504	0.086	0.073	0.033*	0	Missing
Phospholipase B-like	-0.493	-0.535	-0.071	0.027*	0.022*	0.781	0	NonSig
Protein OS-9	-0.497	-0.402	-0.265	0.026*	0.098	0.288	0	NonSig
Lactadherin (*)	-0.356	-0.664	-0.463	0.123	0.003*	0.053	1	NegRel
Lactadherin	-0.288	-0.467	-0.5	0.219	0.051	0.035*	1	NegRel
Apolipoprotein M	-0.398	-0.461	NA	0.082	0.054	NA	0	Missing
Succinate-CoA ligase	-0.474	-0.049	-0.599	0.035*	0.848	0.014*	0	NonSig
N-acetylglucosamine-1-phosphotransferase subunit gamma	-0.781	-0.746	-0.641	0*	0*	0.004*	0	NonSig
Brain ribonuclease (BRB)	-0.417	-0.416	-0.317	0.068	0.086	0.2	2	NonSig
Bifunctional purine biosynthesis protein PURH	0.41	0.312	0.46	0.072	0.208	0.055	0	NonSig
L-lactate dehydrogenase A-like 6B	-0.41	0.059	-0.484	0.073	0.823	0.057	0	NegRel
Proteasome subunit alpha type-4	0.416	-0.446	0.061	0.068	0.064	0.809	1	NonSig
Inactive ribonuclease-like protein 10 (Protein Train A)	-0.384	-0.446	-0.258	0.094	0.063	0.302	1	NonSig

List of protein groups identified as correlated with fertility (Pearson's correlation $p < 0.1$) across at least two out of three seasons. Proteins with $p < 0.05$ are indicated with *. For cases with multiple proteins matching the same annotation, the one used for building multiple linear regressions is indicated with (*). The AgeSig column indicates for how many seasons the protein was found correlated with age (Pearson, $p < 0.1$). The Freeze column indicates whether the protein was found to be linked to different freezability of sperm in a recent study by Gomes et al. [12] (NegRel = negative relation, NonSig = non significant, Missing = not detected).

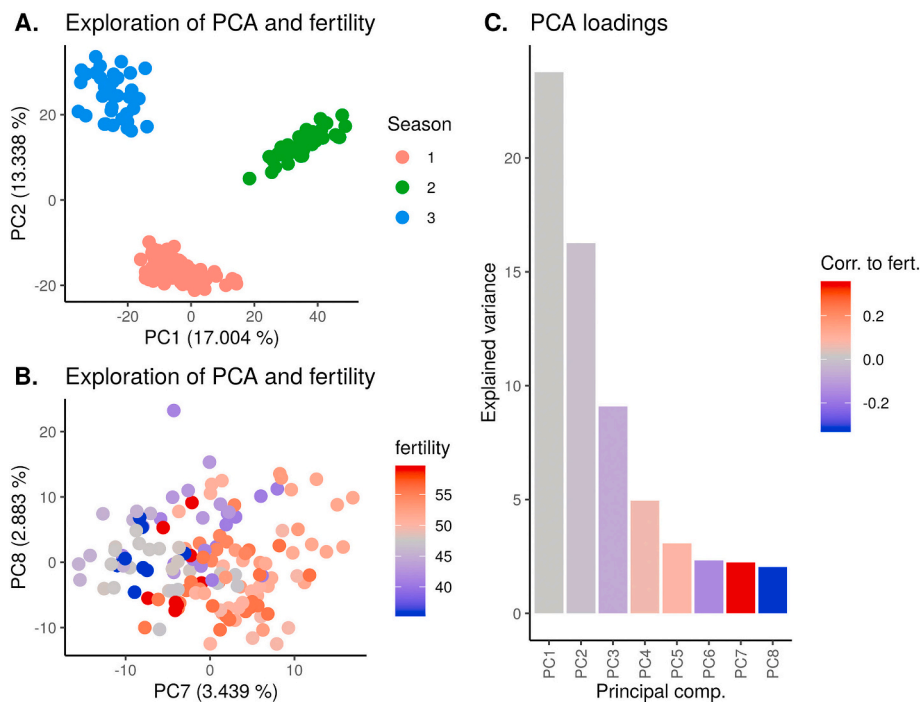


Fig. 2. Principal component on season and fertility. Principal component illustration of samples from all three seasons. A) illustrates the first two principal components and is colored by the season they were sampled. B) is colored on fertility and illustrates principal components 7 and 8, which were selected based on a strong correlation to fertility. C) illustrates the amount of variation explained by the principal components and is colored on their respective correlation to fertility.

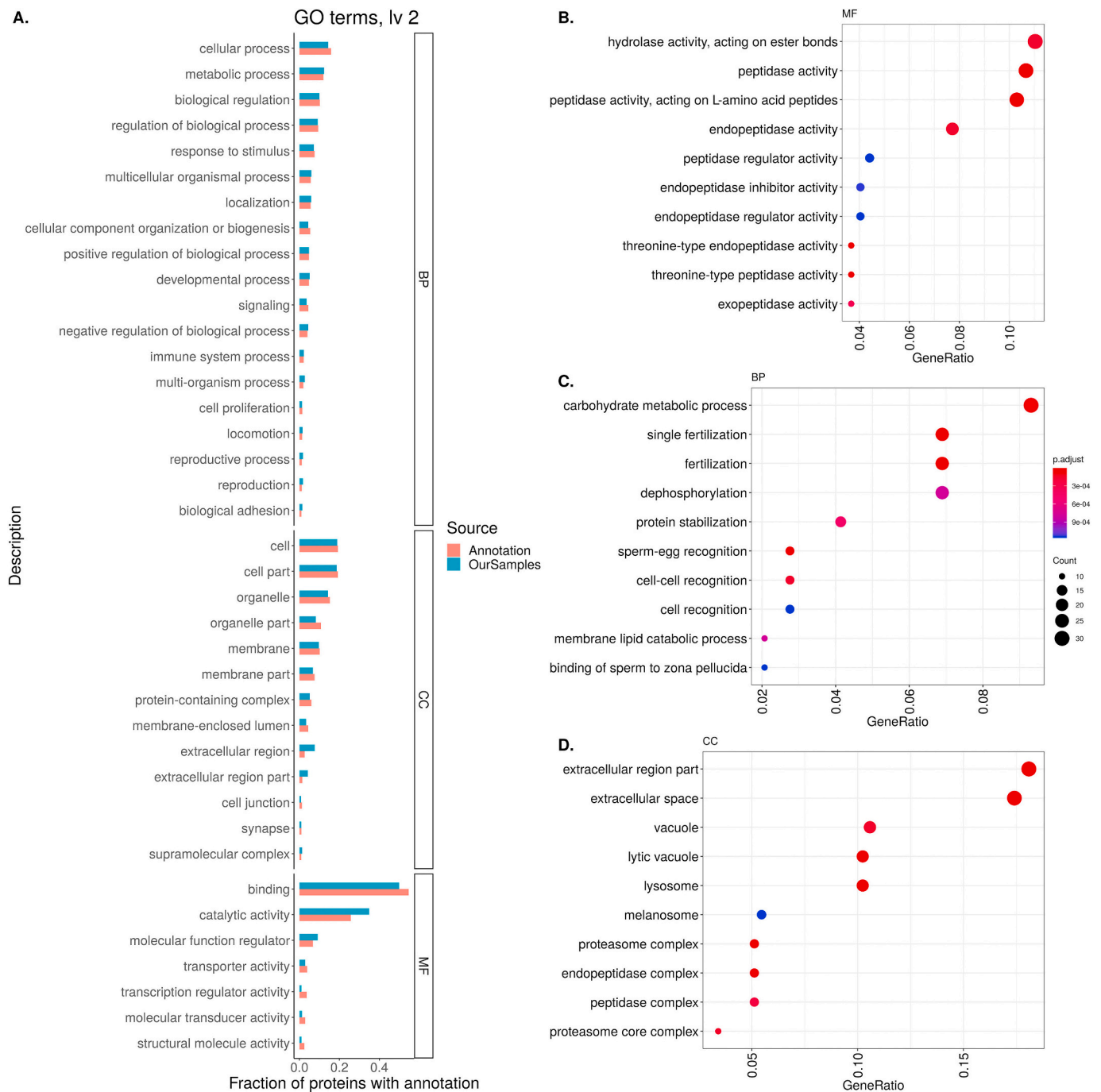


Fig. 3. Gene ontology and enrichment.

Gene Ontology (GO) overview based on second-highest level GO terms, categorized into biological process (BP), cellular compartment (CC), and molecular function (MF). Panel A illustrates the distribution of functions of proteins observed in the dataset compared to the distribution seen in the full predicted proteome. B–D displays top enriched terms in the different ontology categories when comparing observed proteins with the full annotation. A strongly enriched term indicates comparably more proteins related to this function is present in the observed data. The gene ratio indicates the fraction of proteins annotated with that GO term in the annotation that is also identified in the proteomics data.

displays the overlaps among features, divided into groups with no overlap, overlapping in at least two conditions and the same correlation direction, and overlapping in at least two conditions with different correlation direction. It appears that many proteins are identified as correlated with fertility in only one season, which could be due to random associations related to sample variability. However, in cases where a protein group is found to be correlated with fertility in two or more seasons, we observe that their correlation with fertility tend to be similar. At the peptide level, 28 peptides were found with a consistent

correlation ($p < 0.1$) across all three seasons (batches of analysis), matching to 19 different protein groups. Protein groups with consistent correlation to fertility in at least two seasons are illustrated in Table 1, Table 2 and Supplementary Materials S11. The proteins in these tables were also checked for potential correlation with age and annotated for potential relation to freezability, as found in the recent study by Gomes et al. [12]. Among the top-listed proteins, one was found related to age differences in the shortlisted proteins and 11 in the extended list. Furthermore, seven proteins in the extended list have been shown to be

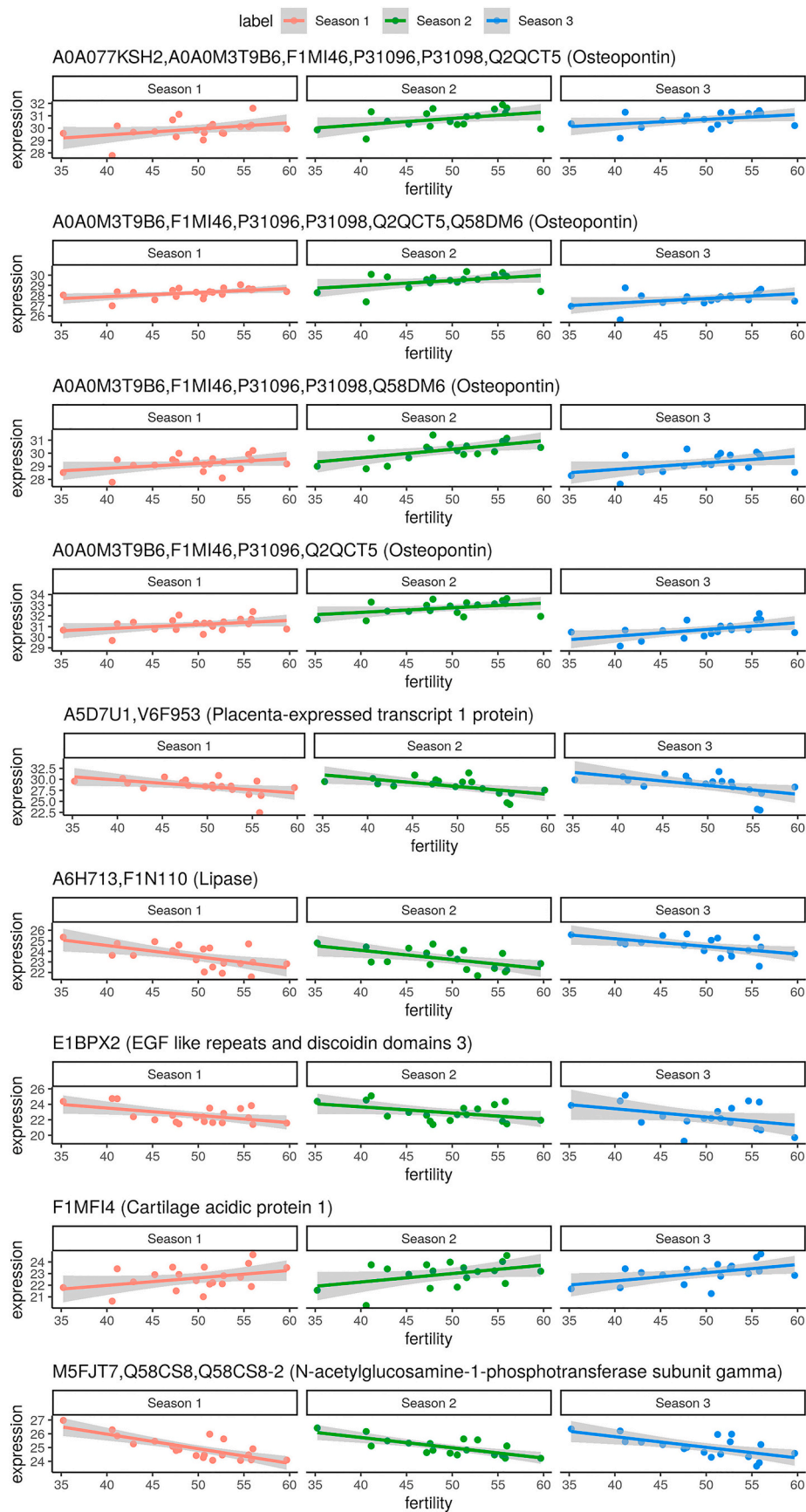


Fig. 4. Consistently correlated protein profiles.

Protein groups found to be correlated to fertility across all three seasons (Pearson's correlation $p < 0.1$ in all cases). The multiple combinations of IDs mapping to the annotation Osteopontin could be due to protein isoforms.

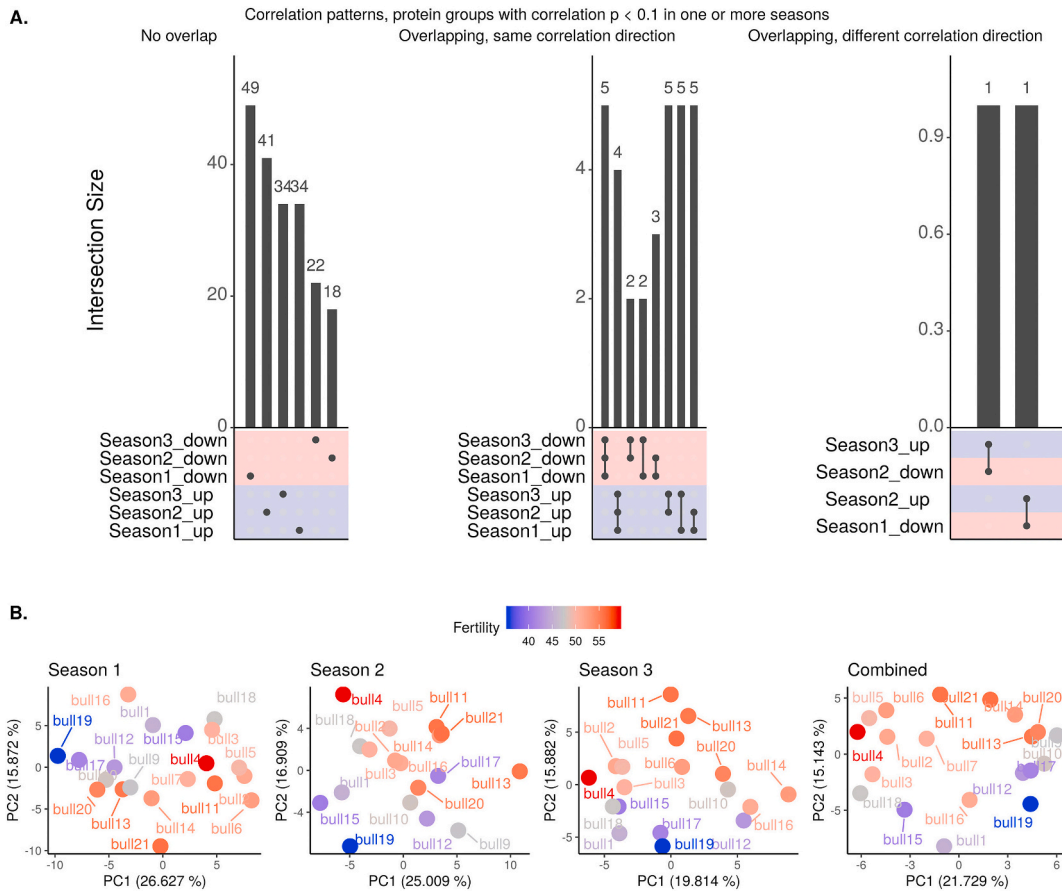


Fig. 5. Correlation pattern overlaps.

Study of features found to be correlated to fertility in at least one season ($p < 0.1$). The A) UpSet plots illustrate how these features are distributed between seasons and further between positive and negative correlation to fertility. The top bars show the number found in each overlap, and the linked dots illustrate which season/direction of correlation is involved. As can be seen, all except two proteins (94% of overlapping proteins) found to be correlated across multiple conditions had the same correlation direction.

The B) panel shows the principal component analysis for the same set of protein groups ($p < 0.1$ correlation with fertility in at least one season), both for individual seasons and after combining all data for each bull.

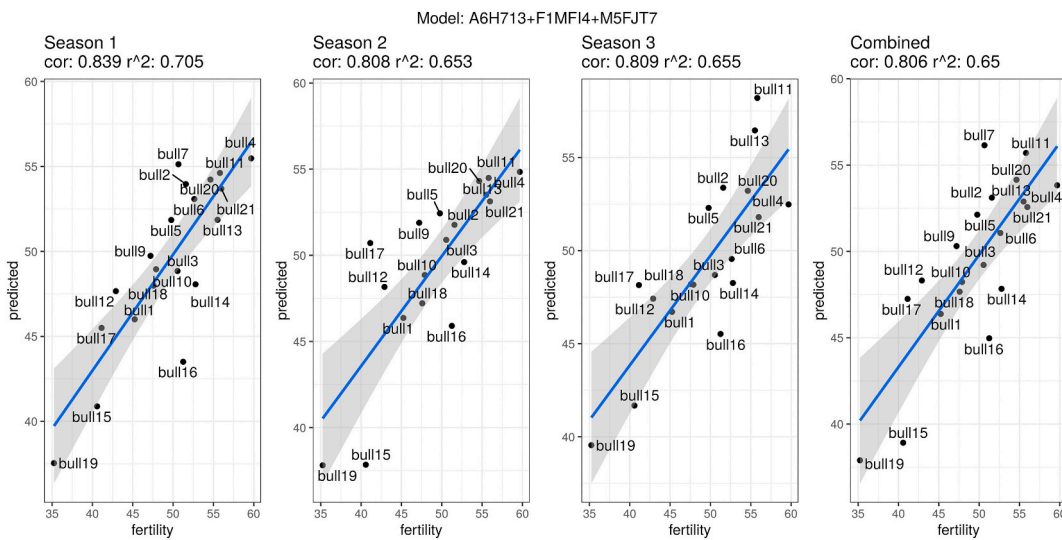


Fig. 6. Predictor results.

The best performing models' predictions illustrated both on the combined dataset, and its performance on individual seasons. The model demonstrates a consistent predictive ability across all seasons, while not fully capturing individual variation (most notably seen in bull 16). The correlation is calculated as Pearson's r .

linked to differences in freezability. Peptides with consistent correlation to fertility across all seasons are shown in the table Supplementary Materials S12. 105 protein groups had $p < 0.05$ in at least one season. These are presented in the table Supplementary Materials S13.

3.4. Single- and multiple- regression models for predicting fertility

A predictor for fertility was built based on both the list of six protein groups consistently correlated to fertility ($p < 0.1$ across seasons) and the extended list of 28 protein groups ($p < 0.1$ across at least two seasons), identifying the combination of proteins producing the predictor with the highest adjusted r^2 value with a maximum of three predictive variables. The adjusted r^2 penalizes the inclusion of many features in the model by only increasing when the added proteins improve the model more than by chance. Values are reported as unadjusted r^2 . The best performing combination of predictors based on six protein groups reached an r^2 value of 0.65 ($r = 0.81$) and was built from three proteins – lipase (LIPA), a cartilage acidic protein 1 (CRTAC1), and an *N*-acetylglucosamine-1-phosphotransferase subunit gamma (GNPTG). Fertility predictions obtained from this model for the combined seasonal bull values are illustrated in Fig. 6. The best performing combination of predictors based on the extended list reached an r^2 value of 0.85 ($r = 0.92$), but had worse predictive performance when applied to individual seasons, and was thus deemed as less robust.

A follow-up set of seminal plasma samples from an independent cohort of bulls with mainly mid-range non-return rates (34–51%) was further analyzed, resulting in 23,423 peptides, and 1640 protein groups identified (FDR < 0.01). The linear regression model generated for the seasonal samples was applied to these samples but showed low predictive ability in this fertility range. However, the lipase (LIPA) used in the predictor showed a similar trend in both datasets, both on the protein

group level and for all shared four underlying peptides identified in both datasets. Trends of all eight (four shared) peptides identified for LIPA are illustrated in Fig. 7.

4. Discussion

In the current study, the bull seminal plasma proteome was explored using a label-free single-shot LC-MS/MS approach. The overall profiling of gene ontology terms revealed a rich proteomic landscape covering a wide range of functions. Gene ontology enrichment against the general background revealed enrichment of extracellular proteins, which is to be expected when studying an extracellular substrate. Furthermore, as expected, there was an enrichment of proteins involved in biological processes related to fertilization and spermatozoa. As for molecular function, peptidase action was enriched, which has been discussed as potentially involved in seminal plasma physiology [29]. The label-free approach allowed the comparison of many samples, and the proteome depth of 1343 proteins reached is similar to a recent study where 1445 proteins were detected [12], and slightly higher than previously reported [11,13], where 685 and 1159 proteins, respectively, were detected. This increased depth provides opportunities for exploration of proteins beyond the most highly abundant and readily measured. This is a key point that stands in contrast to prior studies with less powerful methods, for instance, those carried out using 2D gel electrophoresis coupled to MS, which provides a valuable view of protein patterns, but due to technical limitations, addresses a limited number of proteins [30,31]. Inspection of the most highly abundant proteins shows similarity with previous literature. Soleilhavoup carried out a proteomic study of seminal plasma identifying over 700 proteins, where BSPs belonging to the Spermadhesin family were described as the predominant proteins in bovine SP, with LEG1/C6orf58 also being prominent

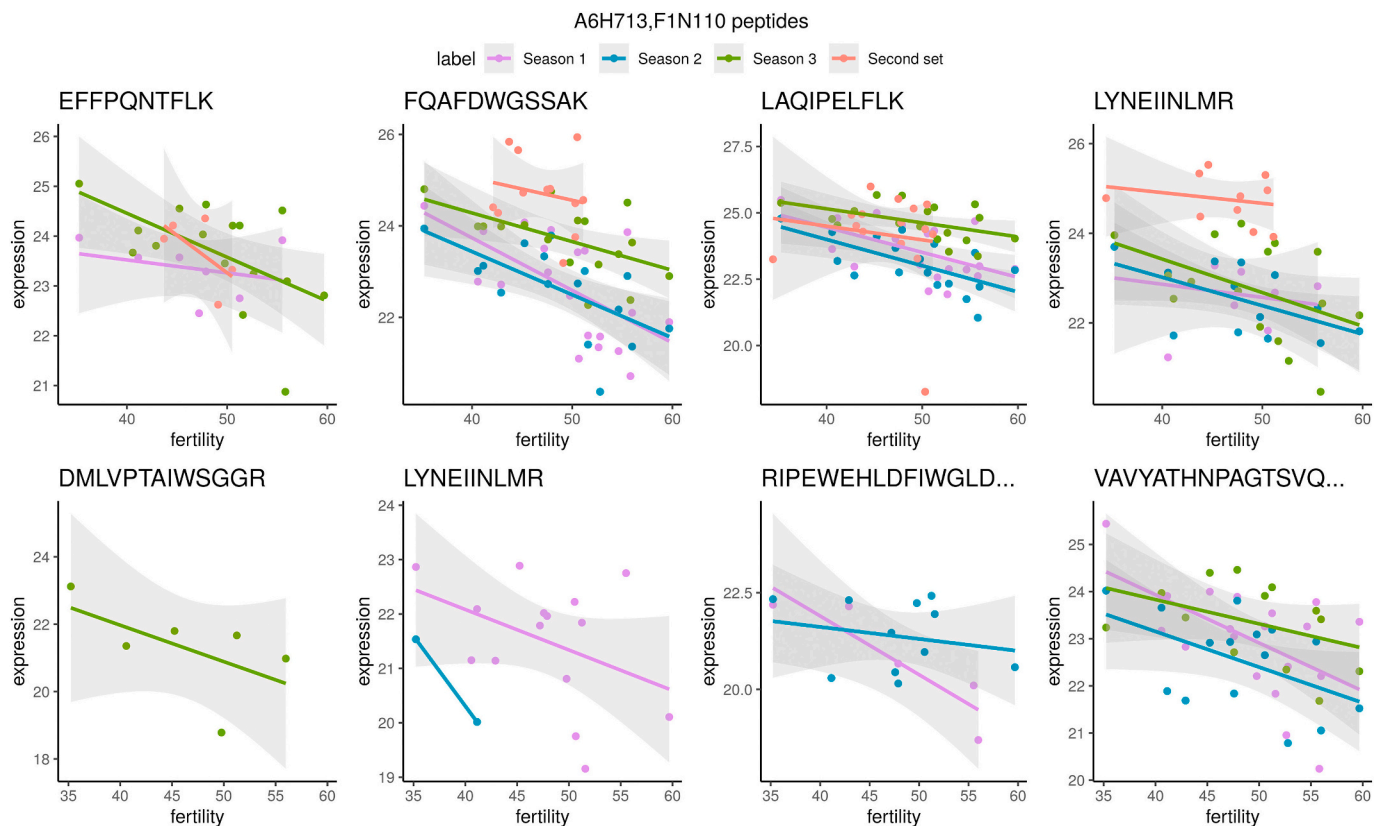


Fig. 7. LIPA peptides.

Illustration of individual peptide correlations to fertility in the lipase protein. Peptide abundances from the second set of bulls are illustrated in orange. All underlying peptides show a negative correlation to fertility in both the three seasonal datasets and in the validation dataset.

[31,32]. Another recent study found BSP-1 and Spermadhesin-1 to have the highest abundances [12]. In our study, the seminal plasma protein BSP-30 K, Spermadhesin-1 and Seminal plasma protein PDC-109 were found to be the most abundant.

Our results confirm that the technical variation associated with repeated analyses by mass spectrometry is relatively small. This is in line with other proteomic studies, where protein extraction and digestion were identified as the major sources of variability [33].

In the present work, the NRR was calculated without adjusting for environmental factors that may influence fertility. The fertility rates from bulls with low numbers of AI are the most susceptible to the above limitation. However, when attempts were made to weight the correlations by the number of inseminations performed for each bull, there was only a very minor impact on the observed protein signatures correlated to fertility. Furthermore, attempts to account for age were performed by calculating the correlation of the top proteins to age. One of the short-listed proteins (A5D7U1, Placenta-expressed transcript 1 protein) had a strong correlation with age, which should be considered when interpreting the data. Any possible inaccuracies in fertility estimates due to other factors would be likely to contribute by increasing the background noise, reducing the sensitivity of the study and leading to missed proteins rather than misleading findings.

To our knowledge, this is the first time that relationships between SP proteomic patterns and fertility were approached from multiple measurements performed from several ejaculates from the same bulls collected at different times of the year. Our results show that using only one time point may lead to misinterpretation as some of the candidates display opposite relationships with fertility estimates at different time points. Season has been reported to impact sperm quality [34], and freezability [35] and our observations suggest that some of the above differences may be related to variations in protein content of ejaculates/seminal plasma between seasons. Although our study was not initially designed to focus on the effect of season, we could identify a series of proteins for which abundance had a relationship to this factor. These results show the importance of analyzing different ejaculates at different time points when aiming at identifying protein markers for fertility.

4.1. Proteins showing consistent relationships with fertility across samples

This study allowed the identification of several proteins or protein groups, where respective abundance was significantly correlated with fertility consistently across all samples. Some of the proteins have been reported previously to have a role in reproductive function and/or fertility. Osteopontin (<https://www.uniprot.org/uniprot/P31096>), is a glycoposphoprotein which was shown to be involved in many different functions [36]. This protein takes part in a mineralized matrix by binding hydroxyapatite and is important for cell-matrix interactions. Another function of this protein which could be of importance for the interactions between spermatozoa and the female genital tract, is its immune function through acting as a cytokine enhancing the production of interferons, correlating with TNF α and being an attractant for macrophages and T cells [36]. In addition, Osteopontin has been reported to play a positive role in sperm-egg binding and embryo development as reported in several studies [37–40]. High concentrations of Osteopontin in SP were also positively related to the freezability of bull semen [41] and Osteopontin gene polymorphism was associated with semen production traits in the buffalo [42]. Our results showing a positive correlation between several peptides of this protein and fertility may be related to the above-mentioned functions. These results are consistent with results previously obtained in bulls [10,43], horses [44] and camels [45].

In humans, EDIL3 is involved in the adhesion of endothelial cells through interaction with a receptor (<https://www.uniprot.org/uniprot/O43854>). This protein may also have an inhibitory role while regulating the formation of vascular-like structures. However, in cattle, no annotation has been reported so far (<https://www.uniprot.org/uniprot>

/E1BPX2), and there is some uncertainty in the annotation due to missing conserved residues compared to other similar proteins.

Leahy et al. [46] mention that this membrane protein limits leukocyte recruitment during inflammation, and speculate that the addition of this protein to the ram sperm membrane at ejaculation [13] may help to modulate the female immune system and aid sperm survival in the ewe reproductive tract. So far, no relationship between EDIL3 abundance and fertility has been reported. Its negative correlation with fertility, found consistently across seasons in our study, and the above regulatory role on the immune response of the female genital tract deserves further investigation.

A lipase (LIPA, long name: Lysosomal acid lipase (LAL)/cholesteryl ester hydrolase) consistently showed a negative correlation to fertility in both our seasonal dataset and in the follow-up dataset. This protein is annotated in the human species (<https://www.uniprot.org/uniprot/P38571>) and is known to be involved in intracellular hydrolysis of cholesteryl esters and triglycerides internalized via endocytosis of lipoprotein particles. It is also reported to be important in mediating the effect of LDL (low-density lipoprotein) uptake on suppression of 'hydromethylglutaryl-CoA' and on the activation of endogenous cellular cholesteryl ester formation. In both cases, it is involved in pathways related to the control of cholesterol levels. In the human, LAL is encoded by the LIPA gene, and low activity of this enzyme causes lipid accumulation and reduction of free fatty acids and cholesterol in the cytosol. This reduction influences numerous downstream genes via transcription factors, resulting in higher expression of the low-density lipoprotein receptor, acetyl-coenzyme A acetyltransferase. This effect results in amplified lysosomal lipid accumulation, increased levels of serum very low-density lipoproteins, and decreased serum high-density lipoproteins, also influencing free cholesterol.

This phenomenon may be of importance for fertility as the male reproductive function is influenced by cholesterol homeostasis [47]. Loss of sperm membrane cholesterol is involved in the process of capacitation [48]. Membrane cholesterol content has been related positively to sperm freezability [49,50]. Low cholesterol in the sperm tail has been associated with poor motility parameters as observed in samples collected during the summer when compared to samples collected during the winter [51], and cholesterol has been presented as a predictive tool for semen quality evaluation [52]. However, despite LAL having been shown to have an essential role in lipid metabolism [53], very little information exists on the relationship between LIPA and male fertility. Other lipases such as the Hormone Sensitive Lipase (HSL/LIPE) has been favorably associated to sperm traits [54]. To explain the negative relationships observed here with LIPA in light of the above information would need further investigation. It would be interesting to document further the precise impact of high LIPA on the distribution of cholesterol in different sperm compartments, especially in the membrane, to explain the negative relationship to fertility found here.

4.2. Proteins with trends for relationships

Among proteins found correlated to fertility in two seasons, attention could be given to Prostaglandin D synthase. This protein is an extracellular transport protein with a high binding affinity for specific cell receptors and small hydrophobic ligands [55]. Its isoform lipocalin-type prostaglandin D synthase was previously thought to be related to fertility [56]. Prostaglandin D synthase was reported to be associated with poor semen parameters such as decreased sperm count, motility, and normal morphology [57,58]. It has been used as a biomarker for 'obstructive azoospermia' in men, a condition where spermatozoa are absent from the ejaculate despite normal spermatogenesis, and an assay diagnostic purposes has been developed [59]. The results of our study, showing a strong and consistent trend across seasons for a negative correlation of this protein with fertility, fit well with the above-mentioned studies and also with previous results obtained in camels [45]. Still, Prostaglandin D synthase has also been reported to have a

positive trend with fertility (as discussed by Saito et al. [60]), showing the complexity of the system. Further studies would be needed to explain the mechanisms behind these trends in different biological systems.

Two other proteins identified in the extended list correlated with fertility in two out of three seasons are GNPTG and CRTAC1. The negative correlation between GNPTG and fertility is consistent with the work of Liu et al. [61], showing a higher abundance of this protein in semen from poorly performing bulls. Furthermore, we found here that CRTAC1 had a positive correlation with fertility but more investigations about its role may be needed as different trends were previously found when studying its relationship with freezing resilience and liquid preservation [8,31].

Numerous studies have described the relationships between the most abundant proteins in SP and fertility, such as BSPs and Spermadhesins. The BSPs (1, 3,5) have been either negatively or positively correlated with fertility [4] or freezability [12], whereas for Spermadhesins (SPADH1 and SPADH2) which may have a protective role against oxidative stress, positive correlations have been reported with freezability [62] and fertility of frozen semen [63,64].

Beyond the proteins directly included in the shortlist, other proteins of interest showing trends for a correlation with fertility were ZAG, showing a negative correlation with fertility, as in previous studies [31], and Spermadhesin Z13, which in our study, showed a trend for a positive relationship with fertility, similar to previous observations [65].

4.3. Prediction of fertility from a combination of protein markers

Due to its importance for the breeding industry, the ability to predict bull fertility has long been sought using both conventional and molecular methods. One early attempt using proteomics was carried out by Killian [9], where it was found that bovine seminal plasma contained proteins associated with bull fertility. A regression model with a positive correlation with fertility was developed ($r = 0.89$), but its robustness when confronted to multiple samplings was not tested. In a recent study also using liquid chromatography-mass spectrometry, more than 1000 proteins were identified, yielding a predictor for fertility (Spearman's rho of 0.94 / Pearson's correlation of 0.83) [11]. However, predictors based on proteomics in linear regression models could be subject to overfitting [66] as the number of proteins used in the models by Viana et al. was high (six proteins) when compared to the number of individuals from which the models were developed (10 bulls). Here, to limit overfitting to the dataset at hand, the protein candidates included in predictive models were chosen by selecting a limited number (in this case, a maximum of three) of those showing a consistent correlation to fertility across repeated samples collected in different seasons, and using adjusted r^2 as selective criteria. This lowers the r^2 value of the prediction brought by our model when compared to other studies. This effect was observed in the data presented here where the inclusion of more proteins led to a model with improved performance in the combined data, but with reduced performance when applied to individual seasons. The development of a predictive signature is difficult due to the many sources of variation influencing molecular data, stemming from both biological factors (season) and technical factors (sample preparation, sample storage, mass spectrometry processing).

To increase the robustness of the signature, we explored further their pertinence by using an independent set of bulls. A limiting factor in this approach was that the fertility range of this second set of bulls was considerably narrower than the first. This contributed to reducing the significance of relationships previously established for most of the proteins, which showed a consistent correlation with fertility in the first study. Despite this reduction, our correlations with fertility, for instance, the negative correlation found between fertility and the amount of lipase (LIPA), and a trend for a positive correlation with one of the Osteopontin protein groups, were confirmed by this additional set of bulls. Moreover, in the extended list of bulls, 11 out of 34 showed similar trends in the independent dataset, albeit not significant, and would need further

verification. We also attempted to explore the behavior of our predictor proteins when applied to the proteins in the study by Viana et al. Here, two of the proteins were identified (Osteopontin and *N*-acetylglucosamine-1-phosphotransferase) but showed no significant trends in the dataset by Viana et al, and lack of quantitative data did not permit further explorations. Many factors may influence the results, such as various software and normalization strategies used in the different studies. Future work would ideally involve testing the signature in a larger set of bulls with a wide range of fertility. This is complicated by the low availability of bulls of low fertility. Due to the progression of genomic selection, only sufficient semen is collected from each for a small number of inseminations. Still, this type of validation would be essential to confirm the reliability of the identified markers.

5. Conclusions

We identified sets of proteins robustly correlated with fertility and established a model providing consistent results for this group of bulls across repeated samples collected at different seasons. It was explored in an independent set of bulls and a reduced number of candidates were found to show consistent results for the two sets of bulls. However, considering the relatively low number of bulls, this model and its associated protein candidates would need further validation, ideally from a set of bulls with a wider range of fertility. Still, we find consistent relationships between the abundance of most of the protein candidates and fertility, as well as other features of male reproductive function. Moreover, there were consistent correlations across samplings. Thus, these results pave the way for functional studies to further decipher the mechanisms underlying male fertility.

Acknowledgements

The authors would like to thank the Proteomics Core Facility at Sahlgrenska Academy, University of Gothenburg for performing the proteomic analysis, and the Animal Breeders' Association of Estonia, Kehta AI station personnel for providing samples. The authors gratefully acknowledge the Swedish Foundation for Strategic Environmental Research (Mistra Biotech) for financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2021.104135>.

References

- [1] P. Humblot, G. Decoux, T. Dhorne, Effects of the sire and district of A1 on cow fertility, *Reprod. Domest. Anim.* 26 (1991) 225–234, <https://doi.org/10.1111/j.1439-0531.1991.tb01533.x>.
- [2] T. Nongbua, Y. Guo, A. Edman, P. Humblot, J.M. Morrell, Effect of bovine seminal plasma on bovine endometrial epithelial cells in culture, *Reprod. Domest. Anim.* 53 (2018) 85–92, <https://doi.org/10.1111/rda.13069>.
- [3] H. Rodriguez-Martinez, Semen evaluation: can we forecast fertility? *Vet. Stanica.* 50 (2019) 293–305.
- [4] X. Druart, J.P. Rickard, G. Tsikis, S.P. de Graaf, Seminal plasma proteins as markers of sperm fertility, *Theriogenology.* 137 (2019) 30–35, <https://doi.org/10.1016/j.theriogenology.2019.05.034>.
- [5] W.M.C. Maxwell, I. Parrilla, I. Caballero, E. Garcia, J. Roca, E.A. Martinez, J. M. Vazquez, D. Rath, Retained functional integrity of bull spermatozoa after double freezing and thawing using PureSperm® density gradient centrifugation, *Reprod. Domest. Anim.* 42 (2007) 489–494, <https://doi.org/10.1111/j.1439-0531.2006.00811.x>.
- [6] S.A. Robertson, Seminal plasma and male factor signalling in the female reproductive tract, *Cell Tissue Res.* 322 (2005) 43–52, <https://doi.org/10.1007/s00441-005-1127-3>.
- [7] J.P. Rickard, T. Pini, C. Soleilhavoup, J. Cogne, R. Bathgate, G.W. Lynch, G. Evans, W.M.C. Maxwell, X. Druart, S.P. De Graaf, Seminal plasma aids the survival and cervical transit of epididymal ram spermatozoa, *Reproduction.* 148 (2014) 469–478, <https://doi.org/10.1530/REP-14-0285>.
- [8] J.P. Rickard, T. Leahy, C. Soleilhavoup, G. Tsikis, V. Labas, G. Harichaux, G. W. Lynch, X. Druart, S.P. de Graaf, The identification of proteomic markers of

- sperm freezing resilience in ram seminal plasma, *J. Proteome* 126 (2015) 303–311, <https://doi.org/10.1016/j.jprote.2015.05.017>.
- [9] G.J. Killian, D.A. Chapman, L.A. Rogowski, Fertility-associated proteins in holstein bull seminal plasma, *Biol. Reprod.* 49 (1993) 1202–1207, <https://doi.org/10.1095/biolreprod49.6.1202>.
- [10] A.M. Cancel, D.A. Chapman, G.J. Killian, Osteopontin is the 55-kilodalton fertility-associated protein in holstein bull seminal plasma, *Biol. Reprod.* 57 (1997) 1293–1301, <https://doi.org/10.1095/biolreprod57.6.1293>.
- [11] A.G.A. Viana, A.M.A. Martins, A.H. Pontes, W. Fontes, M.S. Castro, C.A.O. Ricart, M.V. Sousa, A. Kaya, E. Topper, E. Memili, A.A. Moura, Proteomic landscape of seminal plasma associated with dairy bull fertility, *Sci. Rep.* 8 (2018) 1–13, <https://doi.org/10.1038/s41598-018-34152-w>.
- [12] F.P. Gomes, R. Park, A.G. Viana, C.F. Costa, E. Topper, A. Kaya, E. Memili, J.R. Y. Iii, A.A. Moura, Protein signatures of seminal plasma from bulls with contrasting frozen - thawed sperm viability, *Sci. Rep.* (2020) 1–14, <https://doi.org/10.1038/s41598-020-71015-9>.
- [13] T. Pini, T. Leahy, C. Soleilhavoup, G. Tsikis, V. Labas, L. Combes-Soia, G. Harichaux, J.P. Rickard, X. Druart, S.P. De Graaf, Proteomic investigation of ram spermatozoa and the proteins conferred by seminal plasma, *J. Proteome Res.* 15 (2016) 3700–3711, <https://doi.org/10.1021/acs.jproteome.6b00530>.
- [14] M.C. Chambers, B. MacLean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egerton, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, F. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R. L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, *Nat. Biotechnol.* 30 (2012) 918–920, <https://doi.org/10.1038/nbt.2377>.
- [15] J. Telemán, A.W. Dowsey, F.F. Gonzalez-Galarza, S. Perkins, B. Pratt, H.L. Röst, L. Malmström, J. Malmström, A.R. Jones, E.W. Deutsch, F. Levander, Numerical compression schemes for proteomics mass spectrometry data, *Mol. Cell. Proteomics* 13 (2014) 1537–1542, <https://doi.org/10.1074/mcp.O114.037879>.
- [16] J. Telemán, A. Chawade, M. Sandin, F. Levander, J. Malmström, Dinosaur: a refined open source peptide MS feature detector, *J. Proteome Res.* 15 (2016) 2143–2151, <https://doi.org/10.1021/acs.jproteome.6b00016>.
- [17] J. Häkkinen, G. Vincic, O. Månsson, K. Wårell, F. Levander, The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data, *J. Proteome Res.* 8 (2009) 3037–3043, <https://doi.org/10.1021/pr900189c>.
- [18] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* 20 (2004) 1466–1467, <https://doi.org/10.1093/bioinformatics/bth092>.
- [19] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis* 20 (1999) 3551–3567, [https://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
- [20] J. Willfors, A. Chawade, F. Levander, NormalizerDE: online tool for improved normalization of Omics expression data and high-sensitivity differential expression analysis, *J. Proteome Res.* 18 (2019) 732–740, <https://doi.org/10.1021/acs.jproteome.8b00523>.
- [21] K.V. Ballman, D.E. Grill, A.L. Oberg, T.M. Therneau, Faster cyclic loess: normalizing RNA arrays via linear models, *Bioinformatics* 20 (2004) 2778–2786, <https://doi.org/10.1093/bioinformatics/bth327>.
- [22] A.D. Polpitiya, W.J. Qian, N. Jaitly, V.A. Petyuk, J.N. Adkins, D.G. Camp, G. A. Anderson, R.D. Smith, DANTE: a statistical tool for quantitative analysis of -omics data, *Bioinformatics* 24 (2008) 1556–1558, <https://doi.org/10.1093/bioinformatics/btn217>.
- [23] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* (2018) 1–9, <https://doi.org/10.1093/nar/gky1106>.
- [24] J. Willfors, V. Siino, F. Levander, OmicLoupe: facilitating biological discovery by interactive exploration of multiple omic datasets and statistical comparisons [PREPRINT], *BioRxiv.* (2020), <https://doi.org/10.1101/2020.10.22.349944>, 2020.10.22.349944.
- [25] J.C. Silva, M.V. Gorenstein, G.Z. Li, J.P.C. Vissers, S.J. Geromanos, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition, *Mol. Cell. Proteomics* 5 (2006) 144–156, <https://doi.org/10.1074/mcp.M500230-MCP200>.
- [26] G. Yu, L.G. Wang, Y. Han, Q.Y. He, ClusterProfiler: an R package for comparing biological themes among gene clusters, *Omi. A J. Integr. Biol.* 16 (2012) 284–287, <https://doi.org/10.1089/omi.2011.0118>.
- [27] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *R. Stat. Soc.* 57 (1995) 289–300.
- [28] J.R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties, *Bioinformatics* 33 (2017) 2938–2940, <https://doi.org/10.1093/bioinformatics/btx364>.
- [29] D. Fernández, A. Valdivia, J. Irazueta, C. Ochoa, L. Casis, Peptidase activities in human semen, *Peptides* 23 (2002) 461–468, [https://doi.org/10.1016/S0196-9781\(01\)00622-2](https://doi.org/10.1016/S0196-9781(01)00622-2).
- [30] R.K. Kasimanickam, V.R. Kasimanickam, A. Arangasamy, J.P. Kastelic, Sperm and seminal plasma proteomics of high- versus low-fertility Holstein bulls, *Theriogenology* 126 (2019) 41–48, <https://doi.org/10.1016/j.theriogenology.2018.11.032>.
- [31] C. Soleilhavoup, G. Tsikis, V. Labas, G. Harichaux, P.L. Kohnke, J.L. Dacheux, Y. Guérin, J.L. Gatti, S.P. de Graaf, X. Druart, Ram seminal plasma proteome and its impact on liquid preservation of spermatozoa, *J. Proteome* 109 (2014) 245–260, <https://doi.org/10.1016/j.jprote.2014.07.007>.
- [32] X. Druart, S. de Graaf, Seminal plasma proteomes and sperm fertility, *Anim. Reprod. Sci.* 194 (2018) 33–40, <https://doi.org/10.1016/j.anireprosci.2018.04.061>.
- [33] P.D. Piehowski, V.A. Petyuk, D.J. Orton, F. Xie, M. Ramirez, A. Engel, A. P. Lieberman, R.L. Albin, D.G. Camp, R.D. Smith, A.J. Myers, Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis, *J. Proteome Res.* 12 (2013) 2128–2137, <https://doi.org/10.1021/pr301146m>.
- [34] H.C.A. Teixeira, N.V. Nascimento, C. McManus, A.A. Egito, A. da Mariante, A. F. Ramos, Seasonal influence on semen traits and freezability from locally adapted Curraleiro bulls, *Anim. Reprod. Sci.* 125 (2011) 56–61, <https://doi.org/10.1016/j.anireprosci.2011.04.002>.
- [35] I. Barranco, M.D. Ortega, M.J. Martínez-Alborcia, J.M. Vazquez, E.A. Martínez, J. Roca, Season of ejaculate collection influences the freezability of boar spermatozoa, *Cryobiology* 67 (2013) 299–304, <https://doi.org/10.1016/j.cryobiol.2013.09.001>.
- [36] M.A. Icer, M. Gezmen-Karadağ, The multiple functions and mechanisms of osteopontin, *Clin. Biochem.* 59 (2018) 17–24, <https://doi.org/10.1016/j.clinbiochem.2018.07.003>.
- [37] A.A. Moura, Seminal plasma proteins and fertility indexes in the bull: the case for osteopontin, *Anim. Reprod.* 2 (2005) 3–10, <http://www.cbra.org.br/pages/publicacoes/animalreproduction/issues/download/AR025.pdf>.
- [38] D.W. Erikson, A.L. Way, D.A. Chapman, G.J. Killian, Detection of osteopontin on Holstein bull spermatozoa, in cauda epididymal fluid and testis homogenates, and its potential role in bovine fertilization, *Reproduction* 133 (2007) 909–917, <https://doi.org/10.1530/REP-06-0228>.
- [39] R.F. Gonçalves, D.A. Chapman, R.P. Bertolla, I. Eder, G.J. Killian, Pre-treatment of cattle semen or oocytes with purified milk osteopontin affects in vitro fertilization and embryo development, *Anim. Reprod. Sci.* 108 (2008) 375–383, <https://doi.org/10.1016/j.anireprosci.2007.09.006>.
- [40] E. Monaco, B. Gasparini, L. Boccia, A. De Rosa, L. Attanasio, L. Zicarelli, G. Killian, Effect of osteopontin (OPN) on in vitro embryo development in cattle, *Theriogenology* 71 (2009) 450–457, <https://doi.org/10.1016/j.theriogenology.2008.08.012>.
- [41] J.P.A. Rego, J.M. Martins, C.A. Wolf, M. van Tilburg, F. Moreno, R.A. Monteiro-Moreira, D.O. Santos, A. Moura, Proteomic analysis of seminal plasma and sperm cells and their associations with semen freezability in Guzerat bulls, *Am. Soc. Anim. Sci.* 94 (2016) 5308–5320, <https://doi.org/10.2527/jas2016-0811>.
- [42] S.T. Rolim Filho, H.F.L. Ribeiro, G.M.F. de Camargo, D.F. Cardoso, R.R. Aspilcueta-Borquis, H. Tonhati, K.B. Nunes, W.G. Vale, E.M. Barbosa, K.C. de Sousa, Identification of polymorphisms in the osteopontin gene and their associations with certain semen production traits of water buffaloes in the Brazilian Amazon, *Reprod. Domest. Anim.* 48 (2013) 705–709, <https://doi.org/10.1111/rda.12144>.
- [43] P. Kumar, M. Saini, D. Kumar, A. Bharadwaj, P.S. Yadav, Estimation of endogenous levels of osteopontin, total antioxidant capacity and malondialdehyde in seminal plasma: application for fertility assessment in buffalo (*Bubalus bubalis*) bulls, *Reprod. Domest. Anim.* 52 (2017) 221–226, <https://doi.org/10.1111/rda.12882>.
- [44] C.I. Brandon, G.L. Heusner, A.B. Caudle, R.A. Fayer-Hosken, Two-dimensional polyacrylamide gel electrophoresis of equine seminal plasma proteins and their correlation with fertility, *Theriogenology* 52 (1999) 863–873.
- [45] M.M. Waheed, I.M. Ghoneim, A.K. Alhaider, Seminal plasma and serum fertility biomarkers in dromedary camels (*Camelus dromedarius*), *Theriogenology* 83 (2015) 650–654, <https://doi.org/10.1016/j.theriogenology.2014.10.033>.
- [46] T. Leahy, J.P. Rickard, T. Pini, B.M. Gadella, S.P. Graaf, 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* 1900289 (2020) 1–15, <https://doi.org/10.1002/pmic.201900289>.
- [47] L. Sèdes, L. Thirouard, S. Maqdashy, M. Garcia, F. Cairra, J.M.A. Lobaccaro, C. Beaudoin, D.H. Volle, Cholesterol: a gatekeeper of male fertility? *Front. Endocrinol. (Lausanne)* 9 (2018) 1–13, <https://doi.org/10.3389/fendo.2018.00369>.
- [48] I. Thérien, R. Moreau, P. Manjunath, Major proteins of bovine seminal plasma and high-density lipoprotein induce cholesterol efflux from epididymal sperm, *Biol. Reprod.* 59 (1998) 768–776, <https://doi.org/10.1095/biolreprod59.4.768>.
- [49] J.S. Rajoriya, J.K. Prasad, S.K. Ghosh, P. Perumal, A. Kumar, S. Kaushal, M. Singh, Effects of seasons on enzymatic changes and cholesterol efflux in relation to freezability in Tharparkar bull semen, *Asian Pacific J. Reprod.* 2 (2013) 280–288, [https://doi.org/10.1016/S2305-0500\(13\)60163-3](https://doi.org/10.1016/S2305-0500(13)60163-3).
- [50] M. Singh, S.K. Ghosh, J.K. Prasad, A. Kumar, R.P. Tripathi, S.K. Bhure, N. Srivastava, Seminal PDC-109 protein vis-à-vis cholesterol content and freezability of buffalo spermatozoa, *Anim. Reprod. Sci.* 144 (2014) 22–29, <https://doi.org/10.1016/j.anireprosci.2013.10.016>.
- [51] N. Argov-Argaman, K. Mahgrefthe, Y. Zeron, Z. Roth, Season-induced variation in lipid composition is associated with semen quality in Holstein bulls, *Reproduction* 145 (2013) 479–489, <https://doi.org/10.1530/REP-12-0498>.
- [52] B. Beer-Ljubić, J. Aladrović, T.S. Marenjak, R. Laskaj, I. Majić-Balić, S. Milinković-Tur, Cholesterol concentration in seminal plasma as a predictive tool for quality semen evaluation, *Theriogenology* 72 (2009) 1132–1140, <https://doi.org/10.1016/j.theriogenology.2009.07.009>.
- [53] F. Li, H. Zhang, Lysosomal acid lipase in lipid metabolism and beyond, *Arterioscler. Thromb. Vasc. Biol.* 39 (2019) 850–856, <https://doi.org/10.1161/ATVBAHA.119.312136>.

- [54] L. Hermo, S. Chung, M. Gregory, C.E. Smith, P.W. Shu, M. El-Alfy, D.G. Cyr, G. A. Mitchell, J. Trasler, Alterations in the testis of hormone sensitive lipase-deficient mice is associated with decreased sperm counts, sperm motility, and fertility, *Mol. Reprod. Dev.* 75 (2008) 565–577, <https://doi.org/10.1002/mrd.20800>.
- [55] L. Samanta, R. Parida, T.R. Dias, A. Agarwal, The enigmatic seminal plasma: a proteomics insight from ejaculation to fertilization, *Reprod. Biol. Endocrinol.* 16 (2018) 1–11, <https://doi.org/10.1186/s12958-018-0358-6>.
- [56] G. Killian, R. Gerena, D. Chapman, Y. Cao, A fertility-associated protein in bull seminal plasma is prostaglandin D2 synthase, *Prostaglandins*. 58 (1998) 826–833, [https://doi.org/10.1016/0090-6980\(98\)86801-8](https://doi.org/10.1016/0090-6980(98)86801-8).
- [57] E.P. Diamandis, W.P. Arnett, G. Foussias, H. Pappas, S. Ghandi, D.N. Melegos, B. Mullen, H. Yu, J. Srigley, K. Jarvi, Seminal plasma biochemical markers and their association with semen analysis findings, *Urology*. 53 (1999) 596–603, [https://doi.org/10.1016/S0090-4295\(98\)00550-0](https://doi.org/10.1016/S0090-4295(98)00550-0).
- [58] S.M. Heshmat, J.B. Mullen, K.A. Jarvi, A. Soosaipillai, E.P. Diamandis, R. J. Hamilton, K.C. Lo, Seminal plasma lipocalin-type prostaglandin D synthase: a potential new marker for the diagnosis of obstructive azoospermia, *J. Urol.* 179 (2008) 1077–1080, <https://doi.org/10.1016/j.juro.2007.10.070>.
- [59] A.P. Drabovich, K. Jarvi, E.P. Diamandis, Verification of male infertility biomarkers in seminal plasma by multiplex selected reaction monitoring assay, *Mol. Cell. Proteomics* 10 (2011), <https://doi.org/10.1074/mcp.M110.004127>.
- [60] S. Saito, H. Tsuda, T. Michimata, Prostaglandin D2 and reproduction, *Am. J. Reprod. Immunol.* 47 (2002) 295–302.
- [61] X. Liu, W. Wang, P. Zhu, J. Wang, Y. Wang, X. Wang, J. Liu, N. Li, X. Wang, C. Lin, F. Liu, In-depth quantitative proteome analysis of seminal plasma from men with oligoasthenozoospermia and normozoospermia, *Reprod. BioMed. Online* 37 (2018) 467–479, <https://doi.org/10.1016/j.rbmo.2018.06.025>.
- [62] M.I.M. Jobim, E.R. Oberst, C.G. Salbego, D.O. Souza, V.B. Wald, F. Tramontina, R. C. Mattos, Two-dimensional polyacrylamide gel electrophoresis of bovine seminal plasma proteins and their relation with semen freezability, *Theriogenology*. 61 (2004) 255–266, [https://doi.org/10.1016/S0093-691X\(03\)00230-9](https://doi.org/10.1016/S0093-691X(03)00230-9).
- [63] L. Somashekar, S. Selvaraju, S. Parthipan, S.K. Patil, B.K. Binsila, M. M. Venkataswamy, S. Karthik Bhat, J.P. Ravindra, Comparative sperm protein profiling in bulls differing in fertility and identification of phosphatidylethanolamine-binding protein 4, a potential fertility marker, *Andrology*. 5 (2017) 1032–1051, <https://doi.org/10.1111/andr.12404>.
- [64] L. Somashekar, S. Selvaraju, S. Parthipan, J.P. Ravindra, Profiling of sperm proteins and association of sperm PDC-109 with bull fertility, *Syst Biol Reprod Med* 61 (2015) 376–387, <https://doi.org/10.3109/19396368.2015.1094837>.
- [65] E.B. Menezes, R.V. de Oliveira, M.F. van Tilburg, E.A. Barbosa, N.V. Nascimento, A. L.M.C.S. Velho, F.B. Moreno, R.A. Moreira, A.C.O. Monteiro-Moreira, G.M. C. Carvalho, A.F. Ramos, E. Memili, A.A. Moura, Proteomic analysis of seminal plasma from locally-adapted “Curraleiro Pé-Duro bulls” (*Bos taurus*): identifying biomarkers involved in sperm physiology in endangered animals for conservation of biodiversity, *Anim. Reprod. Sci.* 183 (2017) 86–101, <https://doi.org/10.1016/j.anireprosci.2017.05.014>.
- [66] A. Schneider, G. Hommel, M. Blettner, Linear regression analysis, *Dtsch. Arztebl. Int.* 107 (2010) 776–782, <https://doi.org/10.3238/arztebl.2010.0776>.