CHARLES UNIVERSITY

Faculty of Science

Department of Cell Biology

Study programme: Biology

Branch of study: Developmental and Reproductive biology

Bc. Eliška Macíčková

Roles of acrosomal proteins in fertilization

Role akrozomálních bílkovin v procesu oplození

MASTER'S THESIS

Supervisor: Doc. Mgr. Pavel Stopka, Ph.D.

Prague, 2021

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 8. 8. 2021

Podpis

Acknowledgments

I would like to thank my supervisor doc. Mgr. Pavel Stopka, Ph.D. for his professional guidance, valuable advice and friendly attitude. I am grateful that I could be a part of his lab, where I gained experience mainly in the proteomics and bioinformatics, in which Pavel excels.

Next, I would like to thank Mgr. Tereza Otčenáčková for her endless patience, kindness, and willingness to give advice. Her exceptional sense of order, consistency and systematic work will be an inspiration in my further development.

Many thanks also go to the other lab members, Alica Dodoková, Romana Stopková and Tereza Matějková, who created a friendly environment in the lab.

Last but foremost, I would like to thank my supportive family and friends (D.) for their kind words and engagement, not only in difficult times. Special gratitude goes to my mum, an endlessly kind, wise and good-hearted icon.

Abstract

The acrosome is a crucial organelle for sperm cells. Defects in the biogenesis of the acrosome often lead to the impaired development of sperm cells and consequently, to the inability to fertilize the oocyte. Although it is evident that the acrosome is an essential structure, its main function has not been solved yet. The initial hypothesis was that the acrosome helps sperm to pass through the extracellular layers of the egg. This statement is based on evidence that acrosome is full of proteolytic enzymes, and inhibition of these enzymes reduces the ability to fertilize the oocyte. However, this hypothesis began to be questioned when it was found that (i) acrosome-reacted sperm cells are still able to pass through extracellular layers and fertilize the oocyte (ii) most of the sperm cells undergo premature acrosomal reaction before they reach the oocyte. Therefore, we aimed to characterize proteomes of the acrosome and acrosome-reacted sperm cells, identify important proteins using advanced bioinformatics approaches and discuss the function of the acrosome as a vesicle that serves to store and process metabolites after the previous maturation process. In this work, we revealed new findings that contribute to the hypothesis that acrosome is a lysosome-related organelle.

Key words: sperm, acrosome, lysosome-related organelle, Lipocalins, nLC-MS/MS, random forest

Abstrakt

Akrozom je klíčovou organelou pro spermii. Poruchy v jeho biogenezi vedou často k defektnímu vývoji spermie a neschopnosti úspěšně oplodnit vajíčko. Ačkoli je zřejmé, že akrozom je pro spermii důležitý, stále není vyřešeno v čem spočívá jeho hlavní funkce. Prvotní hypotéza zněla, že akrozom je důležitý pro průchod extracelulárními obaly vajíčka. K této teorii přispívá také fakt, že je akrozom plný proteolytických enzymů a inhibice těchto enzymů, vede ke snížené schopnosti fertilizace. Nicméně tato hypotéza začala být zpochybňována, když bylo zjištěno, (i) že spermie po akrozomální reakci, je stále schopna projít extracelulárními obaly a oplodnit vajíčko (ii) že většina spermií dorazí k vajíčku už bez akrozomálního váčku. Proto jsme si v této práci kladli za cíl charakterizovat proteom akrozomu a spermii po akrozomální reakci, identifikovat důležité proteiny pomocí pokročilých bioinformatických přístupů a diskutovat roli akrozomu, jako váčku, který slouží pro uskladnění a degradaci metabolitů, po předešlém procesu zrání. Přinesli jsme tak nové poznatky, které přispívají k hypotéze, že akrozom je organela příbuzná lysozomu.

Klíčová slova: spermie, akrozom, nLC-MS/MS, random forest, Lipokaliny, lysosomu-příbuzná organela

List of abbrevations

ACR	acrosin
AR	acrosomal reaction
BSPH1	binder of sperm protein homolog 1
CRISP1	cysteine-rich secretory protein 1
CRISP4	cysteine-rich secretory protein 4
ES	equatorial segment
ZP3r	Zona pellucida sperm-binding protein 3
GM130	Golgi matrix protein 130kD
GOPC	Golgi-associated PDZ and coiled-coil motif-containing protein
HYAL5	Hyaluronidase 5
IAM	inner acrosomal membrane
LCN5	lipocalin 5
MT-CO1	cytochrome c oxidase subunit 1
MUP	major urinary protein
nLC-LC/MS	nanoscale liquid tandem chromatography coupled to tandem mass spectrometry
OAM	outer acrosomal membrane
OBP	odorant binding protein
PICK1	PRKCA-binding protein 1
PSMD2	26S proteoasome non-ATPase regulatory subunit 2
RA	retinoic acid receptor
RAR	retinoic acid receptor
RF	random forrest
RXR	retinoic X receptor
SPAM1	hyaluronidase PH-20
STRA8	signalled/stimulated by retinoic acid
TP	transition protein
VAD	vitamin A deficiency
ZP	zona pellucida
ZPbp2	zona pellucida binding protein 2

Table of content

1 Introduction

1.1 Spermatogenesis

Spermatogenesis is a continuous process of formation of haploid spermatozoa from spermatogonia stem cells in the seminiferous tubules. This process is highly regulated and complex. It begins at the base of the epithelium, formed by peritubular myoid cells, and proceeds inwards to tubules, where haploid sperm is released when the cycle is completed. Cycles of meiotic division start when a male is sexually mature and occurs asynchronously along the seminiferous tubule. As a result of these asynchronous cycles, spermatogenesis proceeds along the tubule as a continuous wave (Gewiss *et al.,* 2020). Spermatogenesis covers mitotic division of spermatogonia, meiotic division of spermatocytes and spermatids maturation.

Spermatogonia are cells localized at the base of seminiferous tubules. They form a population of continuously self-renewal cells, which give rise to spermatids during an individual's life. There are two main types of spermatogonia – A, B. Spermatogonia A have some characteristics of stem cells. After division they can be either differentiated or keep the stem cell state and remain spermatogonia A, so-called A_s (single spermatogonia). However, if the spermatogonia differentiate, the cytokinesis does not fully proceed after division and the two daughter cells remain connected with an intracellular bridge – A_{pr} (paired spermatogonia). Similarly, A_{pr} will then undergo 3 rounds of cell division and produce aligned spermatogonia A_{al} (spermatogonia A_{al-4} → A_{al-8} → A_{al-16}) forming a chain. A_{al} differentiates into A_1 spermatogonia. This is again followed by mitotically dividing A_2, A_3, A_4, Intermediate (In), and last B spermatogonia stage, where the meiosis takes over and spermatogonia B becomes a preleptotene spermatocyte. One of the key players in the regulation of spermatogenesis is retinoic acid (RA). RA induces Stra8 gene expression important for the meiotic initiation and progression. This protein is present in the A, Intermediate, B spermatogonia and early meiotic spermatocytes (Zhou *et al.*, 2008).

In general, spermatogonia differentiate into cells called primary spermatocytes. Primary spermatocytes enter meiosis I and undergo the first round of chromosome segregation. Prophase I is a crucial phase subdivided into *Preleptotene, Leptotene, Zygotene, Pachytene, Diakinesis* stages, during which the pairing of homologous chromosomes, establishment of the synaptonemal complex (SC), crossing-over and disassembling of the SC occurs,

respectively. The chromosomal cross-over is a critical event for the meiotic division. Homologous chromosomes are paired tightly with the synaptonemal complex and Spo11 topoisomerase generated double-strand breaks (DSB) (Keeney *et al.,* 1997). DSB are later repaired with homologous recombination to produce new combinations of DNA sequences and thus keep the population genetically diverse. Prophase I is followed by metaphase I, anaphase I, telophase I. Therefore, the first meiotic division ends up with two secondary spermatocytes containing one set of chromosomes. From meiosis I, the spermatocytes go directly to meiosis II. The second meiotic division is similar to mitosis, the four main phases *(Prophase, Metaphase, Telophase, Anaphas*e) are repeated, and sister chromatids separates to daughter cells to produce a total of four haploid spermatids. Spermatids, after the completion of meiosis, enter the process of maturation, called spermiogenesis.

1.1.1 Spermiogenesis

Spermiogenesis is the last step of sperm maturation that takes place in the seminiferous tubules. In this phase, spermatozoa undergo the most significant morphological changes in their development. Round spermatids become motile spermatozoa with a long flagellum and hook-shaped head. However, there are also changes inside the cell. These changes involve DNA condensation, acrosome, and cilium maturation, and in the last step, disposal of the surplus cytoplasm and releasing sperm to the tubule.

DNA condensation of the sperm head is given by several factors, such as posttranslational histone modification, chromatin remodelling, and mainly exchanging most of the histones for protamines. However, the process of DNA condensation seems to begin earlier in spermiogenesis. It has been shown that histones are first hyperacetylated and then are exchanged for protamines in elongating spermatids (Hazzouri et al., 2000). Acetylation masks a positive charge at the N-terminus of histones and thus loosens the packed nucleosome structure. This is supported by Luense *et al.*, (2019) study, where they generated a mouse knockout of histone acetyltransferase that resulted in histone retention and disrupted spermiogenesis. The next step of packaging DNA into the condensed nucleus is carried on by transition proteins (TPs). These TPs with another histone variant create a transition structure that facilitates the eviction of histones and loading protamines on DNA (Barral et al., 2017). Protamines are encoded by two different genes prm1 and prm2 in humans and mice. They are rich in arginine and cysteine residues. Cystine residues can form disulfide bonds between

protamines and thus ensure tight packing of DNA and protection from the negative influence of external factors and nucleases. In addition, arginine masks the negative charge of DNA, which also helps to pack the nucleus tightly. Nevertheless, not all histones are replaced with protamines. 1-10% of all histones remain packed in nucleosomes in the nucleus, mainly around the promotor regions. Some of the promotors are known to be active in embryonic stem cells or other differentiated embryonic cells. It indicates that these genes in sperm are primed with histones for later transcription in the embryo (Jung et al., 2017).

Another crucial event in spermiogenesis involves sperm flagellum formation. Sperm flagellum is a specialized form of the cilium, and thus, some of the findings obtained in the study of cilia can be applied to the sperm tail. In general, the flagellum is composed of axoneme, a structure containing nine microtubules doublets that form a ring around two singlets in the center. The central pair of microtubules (MT) are connected to 9 outer ones by radial spokes. MT doublets slide against each other with the inner and outer dynein arms, ensuring the cilium movement. A nexin-dynein complex stabilizes the MT doublets (Morohoshi et al., 2020). The sperm axoneme arises from the basal body during spermiogenesis, located under the sperm head in the head-to-tail coupling apparatus. It continues along the whole sperm tail through the three main parts - midpiece, principal piece, and end piece.

1.1.2 Retinoic acid in spermatogenesis

Retinoic acid (RA) is an active metabolite of vitamin A and is produced by series of oxidative reactions (Isoherranen & Zhong, 2019). It specifically binds to the nuclear hormone receptors, which directly regulates gene expression (Bastien & Rochette-Egly, 2004). There are two families of nuclear hormones, one which binds all-trans and 9-cis retinoic acid - retinoic acid receptors (RARs) and one which binds 9-cis-retinoic acid – retinoid X receptors (RXRs) (Maire et al., 2019).

RA mediates its effect on spermatogenesis via Stra8 (Signaled by Retinoic Acid 8) gene, which regulates the entry and progression of meiosis (Oulad-Abdelghani et al., 1996). In the testes of Stra8 deficient mouse, there was an accumulation of undifferentiated spermatogonia with no differentiated spermatids (Ma et al., 2018). Injection of RA to testes increased the level of Stra8 and triggered spermatogonial differentiation. Moreover, RA-STRA8 signalling

is periodical, which agrees with the statement from chapter 1.1 that spermatogenesis proceeds through tubules as a continuous wave. Another RA responsive gene is SALL4A transcription factor. RA drives the expression of SALL4A via the heterodimer of RAR/RXR, which results in elevated Kit expression and transition of A_{al} to A_1 spermatogonia (Gely-Pernot et al., 2015). The RA also affects spermiogenesis, where promotes the elongation of spermatids (Endo et al., 2015). Additionally, when the effect of RA is blocked through RAR nuclear hormone, it resulted in an abnormal epididymal phenotype. The integrity of epididymal duct was disrupted with decreased number of sperm cells in tubules (Jauregui et al., 2018). This indicates that RA functions also in the epididymal tissue. If the Vitamin A deficiency (VAD) was induced in rodents, it results in an increased arrest of germ cell differentiation, increased germ cell apoptosis, disorganization of seminiferous tubules (Boucheron-Houston et al., 2013). However, if the VAD rodents were supplemented with vitamin A, spermatogenesis was reinitiated and seminiferous tubules show normal appearance after 62 days of supplementation (Huang & Hembree, 1979). Thus, the effects are reversible.

1.1.2.1 LCN5

Lipocalin 5, also known as the Epididymal Retinoic Acid-Binding Protein (E-RABP), belongs to a protein family of Lipocalins. Lipocalin family contains many proteins with diverse functions. They are evolutionarily conserved across distant organismal taxa and share β-barrel structure, composed of anti-parallel β-sheets, where they bind small hydrophobic molecules. The lipocalin family includes, e.g. Prostaglandin D_2 Synthase, MUPs - Major Urinary Proteins and OBP – Odorant Binding proteins. Besides their role in transporting molecules and chemical communication they also function as carriers of various degradation/toxic products (Grolli et al., 2006; Kwak et al., 2011).

LCN5 was first characterized as a protein of epididymal luminal fluid secreted in the initial segment of caput epididymis (Brooks & Higgins, 1980). The crystallography study reveals a typical β-barrel structure of LCN5 with amphipathic-binding pocket. In folded conformation, it binds all-trans and 9-cis isomer of retinoic acid (Newcomer, 1993). An experiment using northern blot method confirmed the localisation of L*cn5* mRNA in caput epididymis and excluded localisation in other parts, like testes, seminal vesicles, vas deferens, brain, liver, spleen and kidney (Lareyre et al., 1998). The promotor of Lcn5 was described as a tissue-specific fragment driving gene expression restricted to caput epididymis (Lareyre et al., 1999)

and thus, the Cre/LoxP system driven by Lcn5 promotor was developed for studying genes in caput epididymis (Xu et al., 2017). The role of LCN5 in the sperm maturation and healthy maintenance of epididymal tissue was proposed based on two evidence (Ong et al., 2000). First, deficiency of vitamin A has a detrimental impact on the reproductive health of individuals and second, LCN5 binds forms of RA and is expressed in rodent's epididymal tissue

1.2 Sperm epididymal maturation

Sperm maturation does not end after completing meiosis in the seminiferous tubules but carries on when spermatozoa pass through epididymis and *vas deferens*. Epididymis consists of long tubules, where spermatozoa undergo a maturation process and is stored until the ejaculation to the female reproductive organ. Sperm leaving testis are incapable of fertilizing a female gamete and have restricted biosynthetic capability due to the packing of DNA during spermiogenesis. In the epididymis, sperm cells gain motility and ability to fertilize ova due to changes in proteome, lipid composition and small nuclear RNA (snRNA) (Pyttel et al., 2014; Sharma et al., 2016; Skerget et al., 2015). Proteins of epididymal epithelia are released to the lumen in two ways - via merocrine secretion when soluble proteins are being released in exocytosis. And via apocrine secretion in epididymosomes, which are small secreted extracellular vesicles (Nixon et al., 2019).

Proteomic studies revealed dramatic changes in sperm proteomes during transition through the epididymis. Sperm cells in the caput contain ~ 1536 proteins, while in corpus the number arise to ~ 1720 and decreases in cauda to ~ 1234 (Skerget et al., 2015). These large numbers represent a vast replacement of proteins during epididymal maturation. It has been described that several groups of proteins participate in sperm maturation. The few listed below represent only a tiny portion of protein participating in epididymal maturation. One group of proteins represent decapacitation factors, e. g. CRISP1 and CRISP4 (Carvajal et al., 2018). Decapacitation factors prevent premature capacitation and thus the acrosomal reaction. Along with the decapacitation factor, the changes in the cytoplasmatic membrane composition also prevent the premature acrosomal reaction. Proteins like BSPH1 and SERPINA16 influence the cytoplasmatic membrane (Zhou *et al.*, 2008; Plante and Manjunath, 2015). Other epididymal-secreted proteins, Beta-defensins, have a dual function. Besides, their anti-microbial properties they also affect sperm motility and maturation (Pujianto et al., 2013;

Zhao et al., 2011). In the past few years, many studies also concentrated on the transport of small nuclear RNA (snRNA) (Conine *et al.*, 2018; Sharma *et al.*, 2016, 2018; Nixon *et al.*, 2019; Trigg *et al*., 2019). These snRNAs have a direct effect on early embryo development and are transmitted to the sperm cells via epididymosomes (Reilly et al., 2016). It was shown that paternally transmitted snRNAs downregulate a set of genes during preimplantation development in mice (Sharma et al., 2018).

1.3 Acrosome

The acrosome (or acrosomal vesicle) is an organelle typical for sperm cells. It is an exocytic vesicle located on the apical region of sperm heads, overlying the nucleus. The acrosomal reaction is a complex process, during which the acrosome undergoes exocytosis, induced in the female genital tract. Binding to zona pellucida was considered as an inductor of AR (Florman & Storey, 1982). However, it is currently known that most of the sperm cells undergo AR before reaching zona pellucida (Jin *et al.*, 2011, La Spina *et al.*, 2016). It is believed that the acrosome has one primary function: allowing the sperm cell to penetrate through extracellular coats of the egg. Acrosome contains several lytic enzymes which are required for the penetration through ZP (Ferrer et al., 2012). Subsequently, the equatorial segment is revealed, which allows the sperm and egg membranes to fuse together. The acrosome evolved from the Golgi apparatus, and it shares some properties also with lysosomal vesicles. Therefore, it has not been clear whether the acrosome is generated from the Golgi or lysosomal vesicles for a long time. Its biogenesis and proper formation are crucial for sperm development. Defects in the acrosome are usually associated with decreased chance to fertilize the oocyte.

The shape of the acrosome varies among species. However, they share several common features. It has a cap-like structure overlying the nucleus (Fig. 1C). The acrosomal content is enveloped by a membrane, distinguished into an inner acrosomal membrane (IAM) and an outer acrosomal membrane (OAM) (Fig. 1D). OAM tightly underlines the plasma membrane covering the acrosome, and they fuse together during the acrosomal reaction. IAM is anchored through to acroplaxome to the nuclear membrane (Kierszenbaum *et al.*, 2003). The place where IAM and OAM gets to the proximity and is called the equatorial segment (ES) and is revealed after acrosome exocytosis. The ES contains several molecules, due to which it fuses with the membrane of the egg (Toshimori *et al.*, 1998, Satouh *et al.*, 2012 , Barbaux *et al.*, 2020).

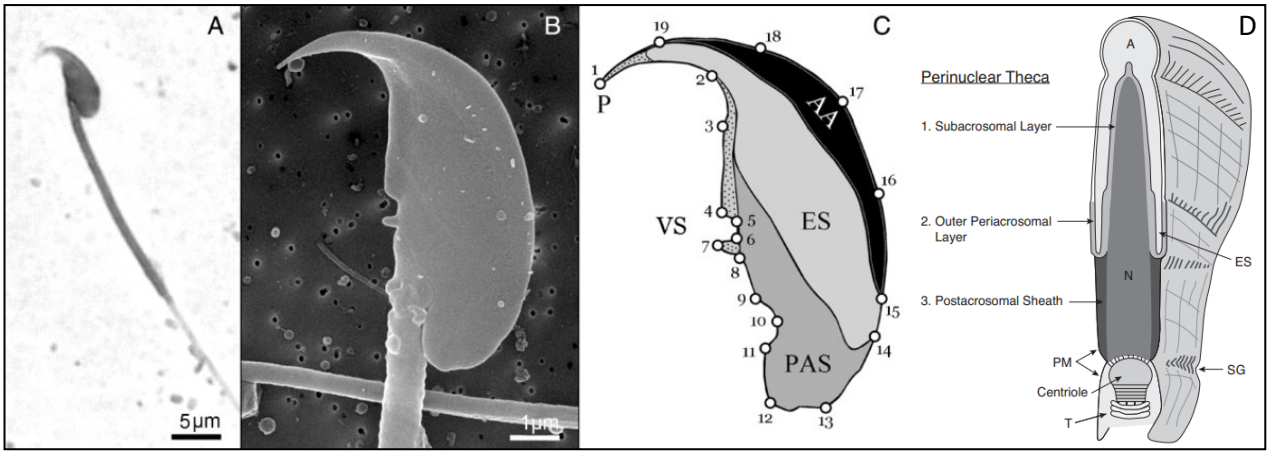

Figure 1: Morphology of the mouse sperm head. (A) Picture from optical microscope depicting the mouse spermatozoon from head to midpiece **(B)** Picture from scanning electron microscope of the mouse sperm head. (C) Diagram of the mouse sperm head with highlighted main parts. (D) Diagram of mid-sagittal section of the mouse sperm head. AA = acrosome, ES =equatorial segment, PAS = postacrosomal segment. Adapted from Oko *et al.*, (2017) and Medarde *et al.*, (2013)

1.3.1 Biogenesis of acrosome

The biogenesis of acrosome can be divided into four major stages: Golgi phase, cap phase, acrosome phase, and maturation phase. The Golgi phase is characterized by a high activity of the Golgi apparatus. The trans-Golgi apparatus secrete many small vesicles, so-called proacrosomal granules. These granules are transported to the top of the nucleus, where they fuse into a single vesicle, a future acrosome. The Golgi phase begins in meiosis during the pachytene stage, when the transcription of acrosomal proteins, such as acrosine, also begins (Kanemori et al., 2016). The vesicle formation, transport and fusion are performed by trafficking machinery and many Golgi proteins are involved in this process (Moreno *et al.*, 2000, Funaki *et al.*, 2013). The proacrosomal granules trafficking is accomplished by myosin motor moving along actin filaments to their destination (Zakrzewski et al., 2020).

During the Cap phase, the acrosome begins to flatten, and its edges start to migrate to the opposite sites enveloping the round nucleus (Fig. 2). The acrosome is anchored to the nucleus through the acroplaxome plate. The acroplaxome is rich in actin and keratin filaments (Kierszenbaum *et al.*, 2003). Furthermore, proteins of the nuclear membrane are also involved in the anchoring of acrosome (Pierre et al., 2012). At this stage, the formation of the acrosomal matrix is evident. The Cap and Acrosomal phases are often joined to one because acrosome forms a cap-like structure during both phases. Finally, the nucleus begins to

condense, and sides of the acrosome start to converge, forming a roof visible from the sagittal section of the head (Fig. 2 – Acrosome phase). Therefore, the morphological changes during the acrosomal phase are closely related to morphological changes in the nucleus.

The Maturation phase takes place during spermiogenesis when the last maturation processes of sperm occur before they are released into the seminiferous tubule. The nucleus condensation continues from the previous stage, forming the acrosome into the familiar crescent shape (Fig. 2).

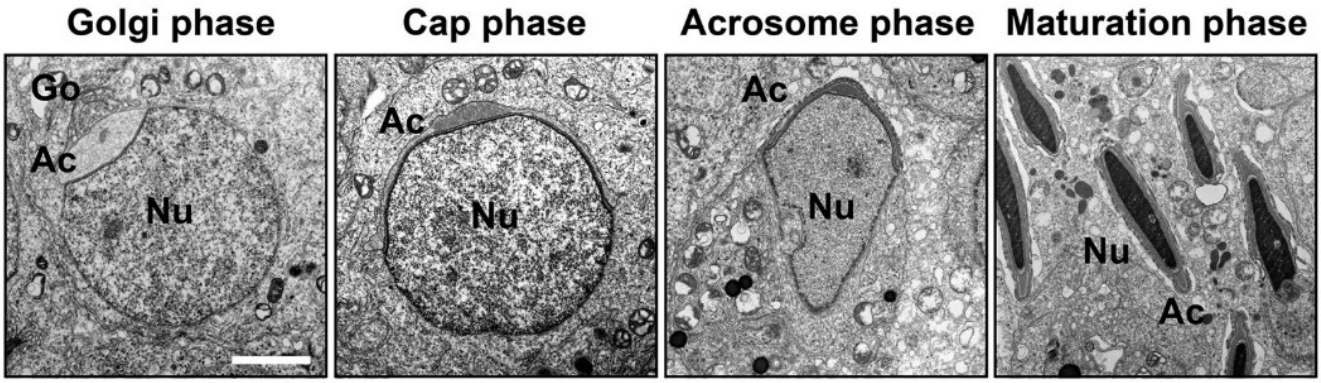

Figure 2: Transmission electron microscopy images depicting the maturation and formation of the acrosomal vesicle during spermatogenesis and spermiogenesis. Go = Golgi apparatus, Ac = Acrosome, Nu = Nucleus, Adapted from Kanemori *et al.*, (2016)

Various defects in acrosomal biogenesis often result in globozoospermia and many proteins have been reported to trigger this state. Globozoospermia is characterized by the presence of round head with severe defects or complete absence of acrosome. The biogenesis of the acrosome can be disrupted in the Golgi trafficking network processes or by the adhesion of the acrosome to the nucleus. Proteins PICK1 (Xiao et al., 2009), GOPC (Yao et al., 2002), GM130 (Han et al., 2017), SMAP2 (Funaki et al., 2013) was reported to be involved in trans-Golgi network trafficking, and knockouts of these genes resulted in defects of proacrosomal granule fusion, abnormal acrosome morphology or the presence of mitochondria in the head. Moreover, PICK1 protein is an interaction partner for GOPC, illustrating the complexness of these events.

1.3.2 Acrosomal content

Many proteins have been so far identified in the acrosome while proteomic analysis reveals more than 1000 of proteins, including proteases, cytoskeletal proteins, transporters, chaperones, etc. (Guyonnet et al., 2012). The content of the acrosomal vesicles can be divided into soluble components and an acrosomal matrix, which differ in the manner of release in the acrosomal reaction. The acrosomal matrix was defined by Buffone *et al.,* (2008) as a membrane-free electron dense material within the lumen, which survives the treatment with non-ionic detergent. The function of the acrosomal matrix is believed to be in a gradual release of acrosomal content and its controlled interaction with zona pellucida of the oocyte (Kim & Gerton, 2003). Additionally, there is an indication that acrosomal content is not distributed homogenously and some proteins can naturally cluster (Westbrook-Case *et al.*, 1995). Apart from this, it seems that acrosomal reaction is not an "all-or-none" event but that acrosomal content is released in a controlled manner, likely due to the self-organisation properties of the acrosomal matrix (Yoshida *et al.*, 2010, Kim *et al.*, 2011).

Among various hypothesised proteins, amyloids are involved in the stability of the acrosomal content. Amyloids are self-assembling proteins, which aggregate to beta-sheet fibrils. They are often associated with pathological conditions, such as amyloidosis, Parkinson's disease, or Alzheimer's disease. However, it seems that amyloids have structural roles in the acrosome due to their self-assembling properties (Guyonnet *et al*., 2014). It was shown that amyloids are stable at lower acidic pH, and upon alkalization at pH 7, they start to disperse. These conditions mimic the changes in pH during capacitation and the acrosomal reaction in the acrosome. Therefore, it was proposed that amyloids contribute to the formation of a stable core of the acrosomal matrix and may associate with other acrosomal matrix proteins. Moreover, they participate in gradual dispersion of proteins during the acrosomal reaction and the transient interaction between zona pellucida and sperm head (Guyonnet *et al*., 2014).

Another group of proteins that occur in the acrosome are zona interacting proteins, namely Zonadhesin (Tardif et al., 2010), ZP3r/sp56 (Bleil & Wassarman, 1990), Zona pellucida binding protein 1 (Zpbp1), and Zona pellucida binding protein 2 (Zpbp2) (Lin et al., 2007b). Although these proteins had predicted function in the binding of zona pellucida, it seems that they are also involved in proper acrosomal matrix formation (Guyonnet *et al.*, 2014; Kim *et al*., 2001; Lin *et al.*, 2007).

Proteases are highly abundant enzymes in the acrosome. Well-documented and significantly abundant proteins include sperm-specific acrosine (ACR) (Ferrer et al., 2012) and several hyaluronidases, e.g. (HYAL2, HYAL3, SPAM1) (Kimura et al., 2009; Modelski et al., 2014; Reese et al., 2010). However, there are many more identified proteases, like metalloprotease (Gottlieb and Meizel, 1987), serin proteases different from acrosine (Mbikay *et al.*, 1997, Talbot and Dicarlantonio, 1985, Ohmura *et al.*, 1999), lysosomal protease Cathepsin D (Srivastava & Ninjoor, 1982) and more. Another extensively discussed proteolytic system present in the acrosome is the proteasomal complex. It was previously believed that proteasomal complex could be involved in protein degradation of zona pellucida and thus in sperm penetration through the extracellular coat of the oocyte (Sutovsky et al., 2004). But lately, the research has been more shifted to its role during sperm maturation and capacitation. Proteolytic enzymes listed above are believed to have various functions during sperm development, starting with the physiological turnover of proteins during maturation (Yi et al., 2011), continuing with specific processing of proteins during capacitation (Zigo et al., 2019), ending with the protein cleavage during AR (Buffone et al., 2009). This indicates that the acrosome is a vital organelle and is constantly formed during the maturation of sperm until the accumulation of changes culminates in AR.

2 Aims of the thesis

In this work, we aimed to investigate the function of acrosomal vesicle in the sperm cells. In particular, we aimed to elucidate whether the acrosome is an organelle related to the lysosome and can store and degrade metabolites during the maturation process of sperm cells. In order to do so, we used a spectrum of experimental and bioinformatics methods.

To address this fundamental question, the subtasks were following:

- Obtaining proteome samples of acrosome and acrosome-reacted sperm from 10 mice
- Characterization of tested proteomes
- Identification of important proteins using machine learning method
- Detecting differentially abundant proteins within tested samples
- Verification of selected acrosomal proteins using indirect immunofluorescence

3 Materials and methods

3.1 Materials

3.1.1 Animals used for sperm collection

In our experiments, we worked with the house mice *Mus musculus musculus*. They were housed under controlled conditions with food and water provided *ad libitum*. Animals were sacrificed by cervical dislocation in all experiments. All animal procedures were carried out in strict accordance with the law of the Czech Republic paragraph 17 no. 246/1992 and the local ethics committee of the Faculty of Science of Charles University in Prague.

3.1.2 Reagents

- Phosphate buffer solution – NaCl 137 mmol/L, KCl 2.7 mmol/L, Na_2HPO_4 10 mmol/L, KH_2HPO_4 1.8 mmol/L; pH = 7.4
- M2 medium (Sigma-Aldrich)
- Triton X-100 (Electron Microscopy Sciences)
- Vectashield with DAPI (Vector Laboratories)
- Paraffin Oil (Roth)
- Formaldehyde (Electron Microscopy Sciences)
- Calcium ionophore (Sigma-Aldrich)
- Bovine serum albumin (Sigma-Aldrich)
- Tissue-Tek® O.C.T ™ (Sakura)
- Primary antibodies: Rabbit polyclonal anti-LCN5 (MyBioSource); Mouse monoclonal anti-PSMD2 (SantaCruz technologies); Mouse monoclonal anti-MT-CO1 (ThermoFisher)
- Secondary antibodies: Goat anti-rabbit IgG (H+L) Alexa Fluor 488 (Life technologies)
  Goat anti-mouse IgG (H+L) Abberior STAR 635P (Abberior GmbH)

3.1.3 Laboratory material

- Petri dishes (Eppendorf)
- Pippet tips 10 µl, 200 µl and 1000 µl (Eppendorf)
- Eppendorf tube 1.5 ml, 0.5 ml (Eppendorf)
- Microscopy slides (ThermoFisher)

- SuperFrost adhesions slides (ThermoFisher)
- Microscopy cover slides
- Parafilm (Bemis Company)

3.1.4 Equipment/Instruments

- Analytical scale
- Lab shaker
- CO_2 incubator
- Pipettes
- Centrifuge
- Vortex
- nLC-MS/MS
- Inverted confocal microscope Leica TSC SP8
- Inverted confocal microscope Carl Zeiss LSM 880 NLO
- Cryostat Leica CM1950

3.1.5 Softwares for data analysis

- Fiji-ImageJ (Schindelin et al., 2012)
- R studio (R Core Team, 2020)

3.2 Methods

3.2.1 Proteomic analysis

For proteomic profiling we used sperm isolated from *cauda epididymis* from 10 adult males of the house mouse *(Mus musculus)*, n = 10. The subsequent protocol is described only for 1 mouse, which is easier to follow.

First, we prepared one preheated at 37°C Petri dish with two 200 µL drops of M2 medium covered with paraffin oil, where we placed *cauda epididymis* after dissection from the mouse. Each *cauda* to one drop. *Cauda epididymis* were cut in half and placed to CO_2 incubator for 10 minutes to let spermatozoa swim out from the tubules.

3.2.1.1 Protein extraction from sperm cells and acrosomal vesicle

This part of the protocol was done only with one 200 µL drop, i.e., with the spermatozoa released from one *cauda epididymis*. Prior to isolation of *cauda epididymis*, we have prepared 8 petri dishes each with 3 drops of 100 µL covered with paraffin oil and placed to CO_2 incubator at 37°C.

To each 100 µL drop we pippeted 5 µL from 200 µL drop where spermatozoa were released from *cauda epididymis*, thus the final volume of each drop was 105 µL. These Petri dishes were placed to CO_2 incubator for 90 minutes while sperm undergo capacitation. Then we aspirated maximal volume of each drop from 4 of the Petri dishes and released them to tube. For the other 4 of Petri dishes, we did the same. Each tube thus contained approximately 1200 µL M2 media with spermatozoa. Tubes were centrifuged for 5 minutes at 2 500 rpm room temperature, supernatant discarded, and the pellet was resuspended in 600 µL of sterile PBS and centrifuged once more. Then, we equally divided the pellet from one tube to 4 new tubes with 200 µL preheated PBS and to each tube add 2 µL of Calcium ionophore (inductor of acrosomal reaction). The samples were placed in CO_2 incubator for 90 minutes during which the acrosomal content has been released. Next, the content of the two tubes were mixed together and centrifuged for 10 min, 500G at 4 °C. After centrifugation, the proteins of the acrosome were released to the supernatant while acrosome-reacted sperm cells remained in the pellet. The supernatant was placed to new tube and left on ice. We added 100 µL of the isolation buffer (0.63 g urea, 0.225 g thiourea, 0.015 g chaps, 0.00465 g DDT, 750 µL dH2O) to the pellets and incubated them for 60 min at RT and vortexed few times. Proteins were precipitated with ice cold acetone, dried for several minutes and then we added the digestion buffer (1% SDC, 100 mM TEAB—pH = 8.5) as described in Stopkova *et al.*, (2017). We used tube with M2 media as the negative control in all experiments.

3.2.1.2 nLC-MS^2

We used the tandem mass spectrometry (nano LC-MS/MS) for peptide mass-fingerprinting and for deriving the individual peptide sequence in non-targeted proteomics.

The nano reversed phase columns were used. Mobile phase buffer A contained 2% acetonitrile and 0.1% formic acid in water. Mobile phase buffer B was composed of 80% acetonitrile, and 0.1% formic acid in water. Our samples were loaded onto a trap column for 4 min at 15µL/min in loading buffer (2% acetonitrile and 0.1% trifluoroacetic acid in water).

Then the ventile was switched and Mobile phase B increased from 4 to 35% at 60 min, 75% B at 61 min, hold for 8 min, and 4% B at 70 min, hold for 15 min until the end of run. Eluting peptide cations were converted to gas-phase ions by electrospray ionization and analysed on a Thermo Orbitrap Fusion. For more detailed description see Stopkova *et al.*, (2017).

3.2.1.3 Protein data analysis

For the generation of proteomic datasets we used the MaxQuant software (Cox et al., 2014). All possible contaminations were removed. For assigning protein names from UniProt *Mus musculus* database to mass spectra we used Andromeda search engine, already integrated in MaxQuant software (Cox et al., 2011).

We used R software for all statistical analysis (R Core Team, 2020) and packages deposited within the Biocondutor repository. Our analysis included several steps. First, we checked the title distribution whether we could use the parametric test using the package mixtools (mixmdl). Next step included data exploration with sPLS-DA from mixOmics package (Lê Cao *et al.,* 2009; Rohart *et al.*, 2017). We used machine learning algorithm within randomForest package (Breiman, 2001; Liaw & Wiener, 2002). To detect differentially abundant proteins within the samples we used Power Law Global Error Model – PLGEM (Pavelka et al., 2004). To visualise results we used particular functions within the ggplot2 package. For calculation the Identity likelihood ratio (IL), we used following formula:

$$IL = \frac{S - P}{S + P} \quad ; \quad IL = (-1; +1)$$

$$S \dots supernatant; P \dots pellet$$

3.2.2 Imunnocytochemistry and imunnohistochemistry

3.2.2.1 Sperm extraction

The spermatozoa were collected from cauda epididymis. Dissection of whole epididymis was performed and then tissue was placed on Petri dish, where it was cut into three parts. Each part was placed to pre-tempered at 37°C 200 µL drop of M2 medium covered by paraffin oil and additionally cut in half to allow sperm cells swim out from the tubules. Petri dishes were placed to CO_2 incubator for 10 minutes.

Smear preparation: 4 µl drops of sperm were smeared onto the microscopic slides and let dried on air. The smear area was outlined using diamond marking pen.

3.2.2.2 Cryosectioning of tissue

Whole tissues *(testes* and *epididymis)* were dissected from the sacrificed mice and immediately placed in a mold prepared from aluminium foil, where they were embedded in Tissue-Tek® O.C.T ™ compound. At room temperature OCT is viscous but freeze at – 20°C. The molds were placed above a liquid nitrogen for 5 minutes to let the tissue freeze. The whole procedure took from 5 to 15 minutes after dissection, to prevent tissue degradation. In this step, we stored frozen tissues in freezer for longer time at - 80°C.

The tissues were cut in 6 µm thick sections using cryostat set at – 19°C and collected on an adhesive microscopy slide. Half of the samples were then fixed in ice-cold methanol-aceton (1:1) for 5 minutes and the other half in 3.7 % formaldehyde for 10 minutes for later optimization. We then performed 5-minute wash in PBS and slides were placed in the freezer set at – 20°C for further usage, described in chapter 3.2.2.3.

3.2.2.3 Immunostaining

Microscopic slides were then fixed with 3.7 % formaldehyde and washed once with PBS. Fixation was followed by blocking and permeabilization step, which we performed using BSA and Triton X-100 solution. For tissue section, which were first hydrated for 5 min in PBS, we used 10% BSA, 0.3 % Triton-X in PBS incubated 1h and for sperm smears 5% BSA, 0.3 % Triton-X in PBS, incubated 45 min in humified chamber at room temperature. Next, samples were incubated with primary antibody, diluted in 1% BSA, 0.3 % Triton X-100 over night at 4°C, in humified chamber in the dark. Dilution ratio for tissue section was 1:60 and for smears 1:80. Then we performed 4x5min washing step in PBS on lab shaker. Secondary antibody was diluted 1:500 in 1% BSA, 0.3% Triton X-100 and incubated with samples for 1 h at room temperature, in humified chamber in the dark and washed 4x5min in PBS on lab shaker. Prior to mounting specimens in Vectashield, we treated sperm cells with Lectin PNA conjugated with Alexa Fluor 568 in dilution ratio 1:500 for 30 min at RT, in humified chamber in the dark. This lectin specifically stains the outer acrosomal membrane and thus, is used as a common acrosome vesicle marker. Eventually, samples were mounted in a drop of Vectashield mounting media and coverslip was sealed with nail polish to prevent drying of

the specimen. Samples were stored at 4°C until imaging. The negative controls were performed to test whether the observed signal is not due to non-specific binding of secondary antibody. The samples were incubated with PBS instead of primary antibody.

3.2.2.4 Image data processing

All data obtained from microscopes were manually processed in Fiji software. Contrast was adjusted remove noise and highlighted observed patterns. For the same purpose was applied Gaussian blur.

4 Results

4.1 Proteomic profiling

In the first step, we used explorative methods to reveal specific patterns in our proteomic dataset. For this purpose, we also used the machine learning algorithm called Random Forest. Next, we continued with another analysis to reveal differentially abundant proteins in our tested samples. The last step involved verifying the localisation of selected proteins by indirect immunofluorescence.

4.1.1 Proteome data mining and predictive exploration

Methods of multivariate predictive exploration are typically used in large biological datasets, such as proteomes and transcriptomes. These exploratory approaches aim to unravel the correlation structure between tested samples and were designed for the classification in highly dimensional datasets with possibly many correlated variables in relatively few samples. To see if our samples have unique protein composition, we ran the sparse partial least squares discriminant analysis (sPLS-DA). sPLS-DA is an extension of sparse PLS which was designed to identify subsets of correlated variables in two datasets (Lê Cao et al., 2008). sPLS-DA enables the selection of the most predictive or discriminative features in the data that help classify the samples (Lê Cao *et al.*, 2011). In large datasets, there are many uninformative proteins which do not contribute to the characterisation of the different classes. The sPLS-DA analysis identifies the subset of proteins that best discriminate the classes.

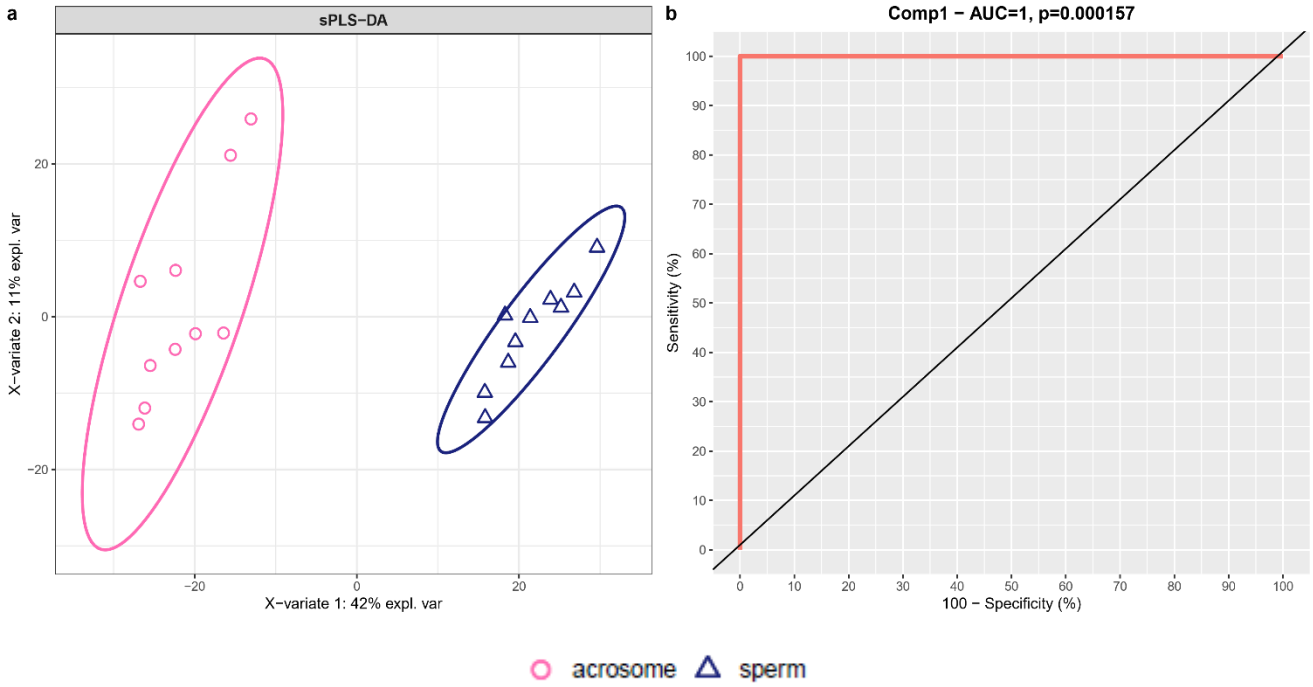

Figure 3: Graphical visualisation of proteome variability between tested groups by sPLS-DA. (a) sPLS-DA reveals a distinct signature in the acrosome and acrosome-reacted sperm cells. This shows that proteomes of these structures have unique protein composition. Clouds of individual mice cluster together, which indicates less variability between individual mice. The variability between individual mice was higher in the protein composition of acrosome. **(b)** AUC ROC curve shows 100% specificity and 100% sensitivity meaning that used classifier (sPLS-DA) is reliable. AUC=1, p=0.000157 reveals an ideal model with high confidence of correctness.

We used 5 components for sPLS discrimination (ncomp = 5). In Fig. 3a X-variate 1 explains 42% variation, while X-variate 2 explains 11% of detected variation. The discrimination results in clear separation between the two proteomes. The data show 'perfect' discrimination, significant already in the first dimension (p = 0.000157). AUC is abbreviation for the Area Under Curve, which is an index used in the classification analysis which determines the probability that a positive event has a higher probability given by the model than a negative event. It ranges from 0 to 1. A model which made 100% prediction false has the AUC value 0. A model is considered good when the AUC value is higher than 0.7. And the classification accuracy is perfect when the model has 100% specificity and 100% sensitivity (Fig. 3b).

Perfect discrimination in our samples, means that the proteome of the acrosome and acrosome-reacted sperm cells are unique structures in terms of protein composition. Much lower variability was observed on the level of individuals.

Identification of proteins important for acrosome and acrosome-reacted sperm cells

To further understand which proteins in particular are responsible for the discrimination of the proteomes, we used the machine learning algorithm called Random Forest. It belongs to a group of tools used for the classification of high dimensional data and can handle both categorical and numerical data. For detailed description of this method see Breiman (2001).

In brief, the random forest consists of several mathematical algorithms that make decision trees which evaluate the importance of the affiliation of protein to one or the other group. The random forest grows many decision trees. Then the forest chooses a protein with most votes as more important. Put it in another way, the random forest takes also into account the variance between individual samples of acrosomal proteins and sperm cell proteins. The protein that has a high variability in abundancy within samples, will have lower importance than a protein with low variability. The more features (proteins) are in the dataset, the smaller the value of importance will be. The strength of random forest analysis is that decisions/predictions are not made by individual trees but by the whole forest. Thus, the trees protect each other from their individual errors.

The number of trees that were generated in our forest is 500 and we used 2000 permutations, (nPerm = 2000). Random forest analysis identified 34 proteins on the scale of importance ranging from 0.05 to 0.15 (Fig. 4). This relatively low number of importance is due to a large number of proteins in the dataset. From the group of acrosome-reacted sperm cells it identified a total of 16 proteins and 18 proteins from the acrosome. SPINK2 – Kazal-type 2 acrosin inhibitor was identified as the most important for sperm cells. In the next step, we annotated these proteins using UniProt Knowlegdebase (The UniProt Consortium et al., 2021). UniProtKB is a freely accessible database of protein sequences and functional information which integrates and interprets data from multiple resources. The UniprotKB has two sections: one with the manually annotated records providing information from literature and curator supervised computational analysis and the second one with automated analysed records awaiting manual annotation. The manual annotation makes the UniProt a reliable source of information. For the annotation of our data, we extracted information only from the first section – manually annotated. Greater emphasis was put on the acrosomal proteins.

Figure 4: Graphical representation of 34 top important proteins evaluated by random forest analysis. Proteins are listed by gene names. The random forest determines 34 proteins with diverse functions on the scale of importance ranging from 0.05 to 0.15. 18 proteins were chosen from acrosome and 16 proteins from acrosome-reacted sperm cell. SPINK2 (Kazal type 2 serin peptidase inhibitor) and ADSS (Adenylosuccinate synthetase) were identified as the most important in the sperm cell and in the acrosome, respectively. Interestingly, many proteins from acrosome, e. g. PGK2 (Phosphoglycerate kinase 2), TKT (Transketolase), HK1 (Hexokinase 1) are involved in carbon metabolism according to the UniProt database. Acrosome = pink, Acrosome-reacted sperm cell = blue.

Some of the annotated proteins from acrosome (Supplementary material, Table 1) have a proteolytic function, e. g. PREP, PSMB7, or are involved in the associated ubiquitin-dependent catabolic pathway, like ARIH1, CUL3 and deubiquitination – OTUB1. Another two proteins, HK1 and PGK2, mediate steps of the glycolytic pathway. Additionally, TKT takes part in the pentose-phosphate pathway of the carbon metabolism. Proteins of stress response were also identified – HSPA4, LTF and PARK7. LCN5 and PTGDS belong to the Lipocalin protein family with various functions. Moreover, structural protein (AP1G1), pH regulation protein (CA2), protein of purine metabolism (ADSS), regulatory subunit of protein phosphatase (PP2R1A) and a protein with unknown function (PEBP4) were identified.

From the table of annotated proteins of acrosome (Supplementary material, Table 1) and acrosome-reacted sperm cells (Supplementary material, Table 2), we can conclude that random forest analysis determines many proteins with diverse functions. It is no surprise that many identified sperm-cell proteins are involved in the flagellum movement/assembly or oxidative phosphorylation in mitochondria. On the other hand, the proteins of acrosome have much broader spectrum of functions.

4.1.2 Identification of diferentially abundant proteins

In the next step, we identified all differentially abundant (DA) proteins in sperm and acrosome. For this purpose, we used the analytic method called The Power Law Global Error Model (PLGEM) (Pavelka et al., 2004). PLGEM represents an improvement in the evaluation of DA proteins or expressed transcripts in transcriptomics. This model favours proteins which have a lower variance in one group and high in the other and penalizes those, which have a high variance in each of the groups. Standard deviations are calculated from the mean signal intensity. In the framework of statistical hypothesis testing, the model tests against the null hypothesis that between the tested groups there are no differentially abundant proteins.

PLGEM was applied to our dataset and the results of DA proteins between our tested groups are visualized in MA plot - Fig. 5a. MA plots are a standard way to visualize genomic and proteomic data. The values of fold difference (log ratio) are plotted against mean signal intensity (log ratio) between our measured samples. Those proteins that were identified with random forest are highlighted with their gene names in Fig.5a. Non-significant DA proteins have p-value > 0.05 and are represented by grey dots. The significant DA proteins are ranging from green to blue with p-value < 0.05 and fold difference (FD) < -2 and > 2.

Similarly, we calculated the identity likelihood ratio for non-significant DA and significant DA proteins (Fig.5b). The histogram density plot reflects standardise data distribution (-1;1) in acrosome and acrosome-reacted sperm cells. Most non-significant DA proteins have their maximum density distribution around zero, while the significant DA proteins have its maximum in the extreme values: -1 and 1. Negative values score for the acrosome and positive for the sperm cells. Interestingly, the density of significant DA proteins was much higher in the acrosome when compared to sperm cells.

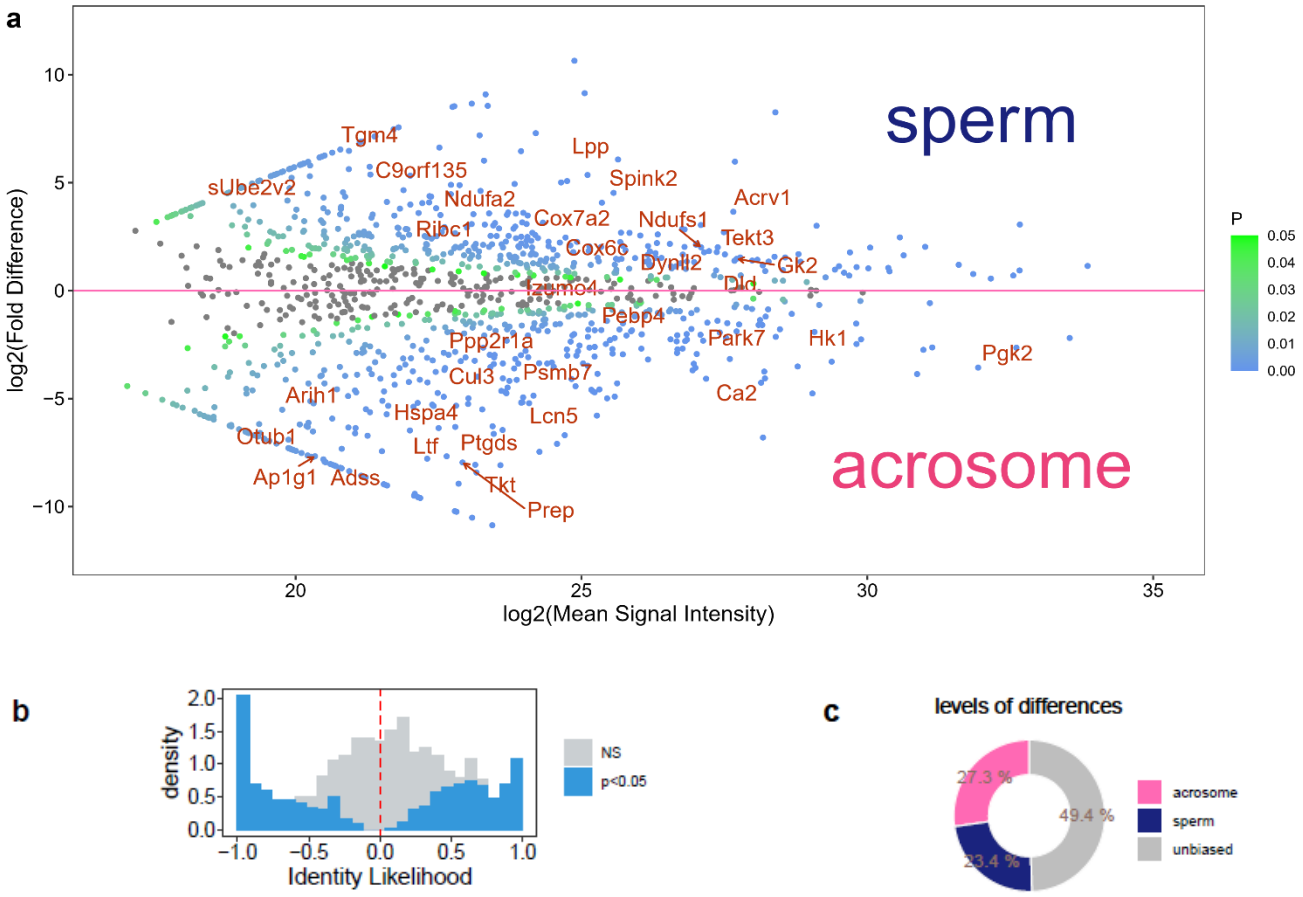

Figure 5: The graphical visualisation of differentially abundant proteins in the acrosome and acrosome-reacted sperm cells. (a) MA plot of differentially abundant proteins shows the differences on the protein level between acrosome and acrosome-reacted sperm cells. Fold difference (log ratio, y-axis) is plotted against mean signal intensity (log ratio, x-axis). Individual proteins are represented by dots. The p-value is color-coded: $p > 0.05$ grey, $p = (0.05; 0.03)$ green, $p < 0.03$ blue. The non-differentially abundant proteins (FD < -2 and > 2) clustered along the pink baseline. Differentially abundant proteins cluster further from the baseline ($y = 0$). The proteins previously identified with random forest analysis are marked with their gene names. **(b) Histogram of identity likelihood ratio**. Data distribution of non-significantly DA proteins are in grey, whereas the significant DA ($p < 0.05$) protein in blue. **(c) Overall balance of data distribution in pie chart.** 49.4 % proteins were present in both samples with no significant differential abundance. 50.6 % was differentially abundant (27.3 % were enriched in the acrosome, 23.4 % enriched in the acrosome-reacted sperm cells).

The summarization of these results is also depicted in the pie chart (Fig. 5c). We detected a total of 1229 proteins. 73 proteins were uniquely detected in the acrosome and 81 in the sperm cells. A total of 49.4 % proteins were non-significant, and equally occurred in both types of the samples. Whereas significantly abundant proteins represent more than 50%, a total of 23.4% were enriched in the sperm cells, and even more, 27.3% in the acrosome.

Following these results, we can repeatedly conclude that acrosome is a unique structure in terms of protein composition.

4.1.3 Localisation of detected proteins

In order to provide evidence that the sample preparation procedure of acquiring proteomes was carried out correctly and our data are valid, we continued with series of experiments using immunofluorescence probes. For this purpose, we selected three proteins: LCN5, PSMD2 and MT-CO1. The immunolocalisation also included series of protocol optimalization, such as the antibody titration and fixative testing (data not shown), always using the negative controls. The experiments were performed on the mouse sperm cells released from the cauda epididymis and on the sperm cells after acrosomal exocytosis to confirm that selected proteins are released during acrosomal reaction. To assess whether the proteins of interest are in the acrosome, we used peanut agglutinin (PNA) which is a common acrosomal marker. PNA is a lectin from *Arachis hypogaea* and binds specific carbohydrates in the outer membrane of the acrosome.

LCN5 is also known as specific epididymal retinoic acid-binding protein. It carries all-trans and 9-cis retinoic acids. We selected LCN5 not only because it was evaluated by random forest analysis as important protein for acrosome but also because the family of lipocalins is important in modulation of mammalian reproduction and chemical communication. So far, there is no mention of its function in the sperm acrosome. However, this protein is primarily assigned a function in the maintenance of healthy endothelia of epididymal tissue. Thus, it is a question whether it is also the case of sperm cells.

PSMD2 is a subunit of 19S regulatory particle of 26S proteasomal complex. Although the random forest analysis identified PSMB7 subunit as important for acrosome, we possessed specific antibody only against PSMD2, which we also detected in the acrosome (based on signal from LC-MS/MS). Thus, we selected PSMD2 as a suitable replacement for PSMB7.

MT-CO1/COX1 was not identified by random forest analysis. However, we found interesting that the subunit of Cytochrome c oxidase I, the complex IV of electron transport chain in mitochondria was significantly abundant in the acrosome and not in the sperm cells. Therefore, we selected this protein for immunolocalisation. We expected that this could serve as a robust indicator to strengthen our hypothesis that acrosomal vesicle serve to chelating the catabolic waste during maturation process.

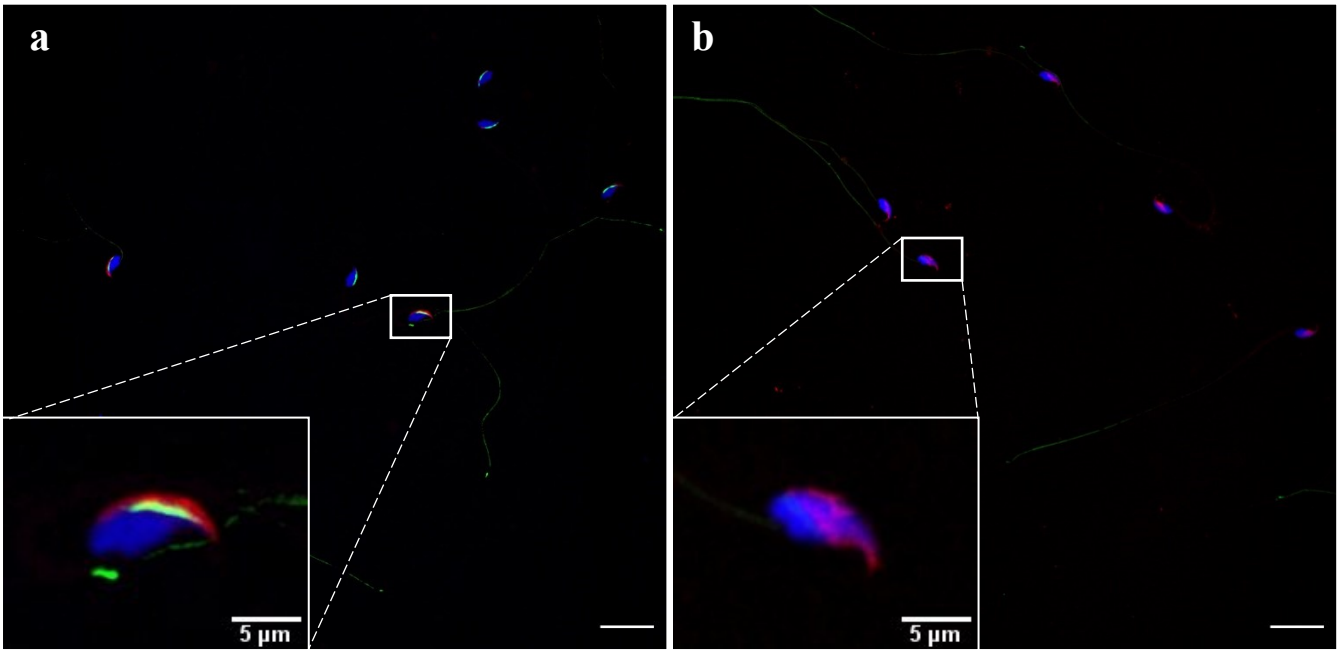

Figure 6: Immunolocalisation of LCN5 in the mouse spermatozoa released from cauda epididymis **(a)** Localisation of LCN5 in released spermatozoa. **(b)** Spermatozoa after acrosomal exocytosis. LCN5 = green, PNA = red, DAPI = blue, scale bar = 20 µm, objective: 63x oil immersion, NA = 1.4

As shown in Fig. 6-8a we detected all three proteins in the acrosome. The green signal from proteins of interest was overlapping with the red signal from PNA and showed a reliable degree of colocalisation. Signal intensity from MT-CO1 was weaker, which is probably due to a lower concentration when compared to LCN5 and PSMD2 (based on the signal from LC-MS/MS). Interestingly, MT-CO1 was also detected in the connecting piece under the sperm head. The release of detected proteins during AR is verified by the loss of green signal (Fig. 6-8b). These findings are in agreement with the proteomic data.

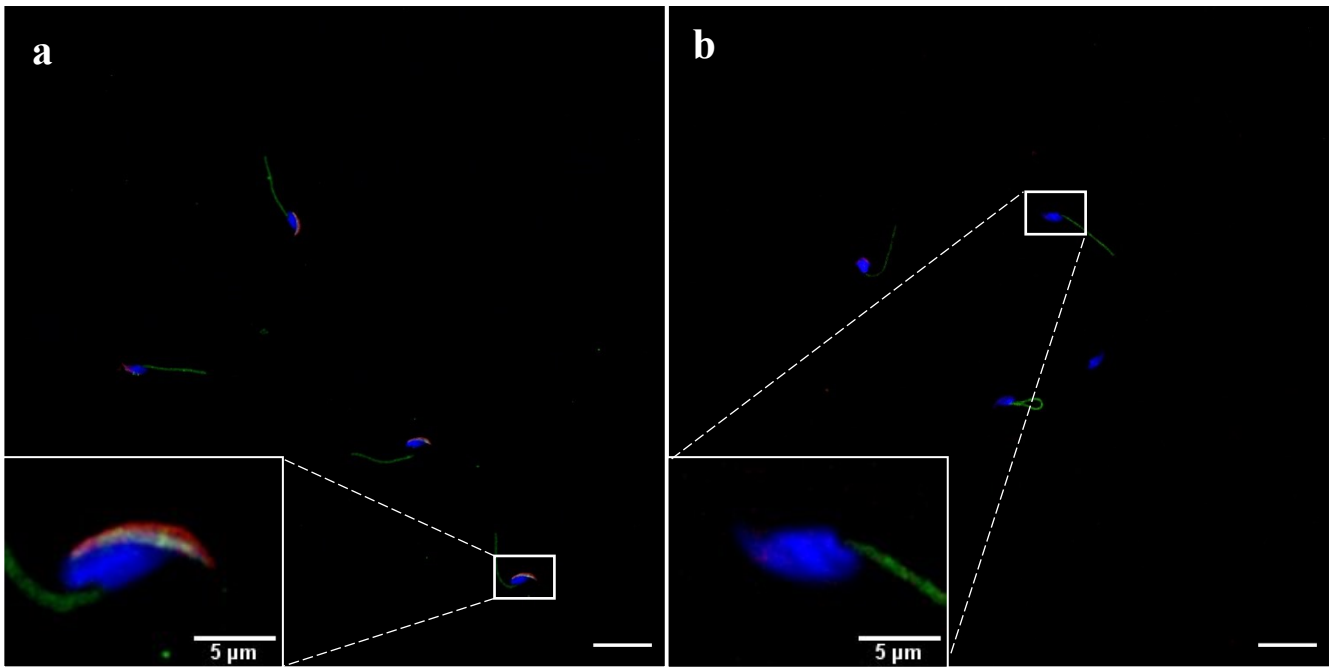

Figure 7: Immunolocalisation of PSMD2 in the mouse spermatozoa released from cauda epididymis. (a) Localisation of PSMD2 in released spermatozoa. **(b)** Spermatozoa after acrosomal exocytosis. PSMD2 = green, PNA = red, DAPI = blue, scale bar = 20 µm, objective: 63x, oil immersion, NA = 1.4

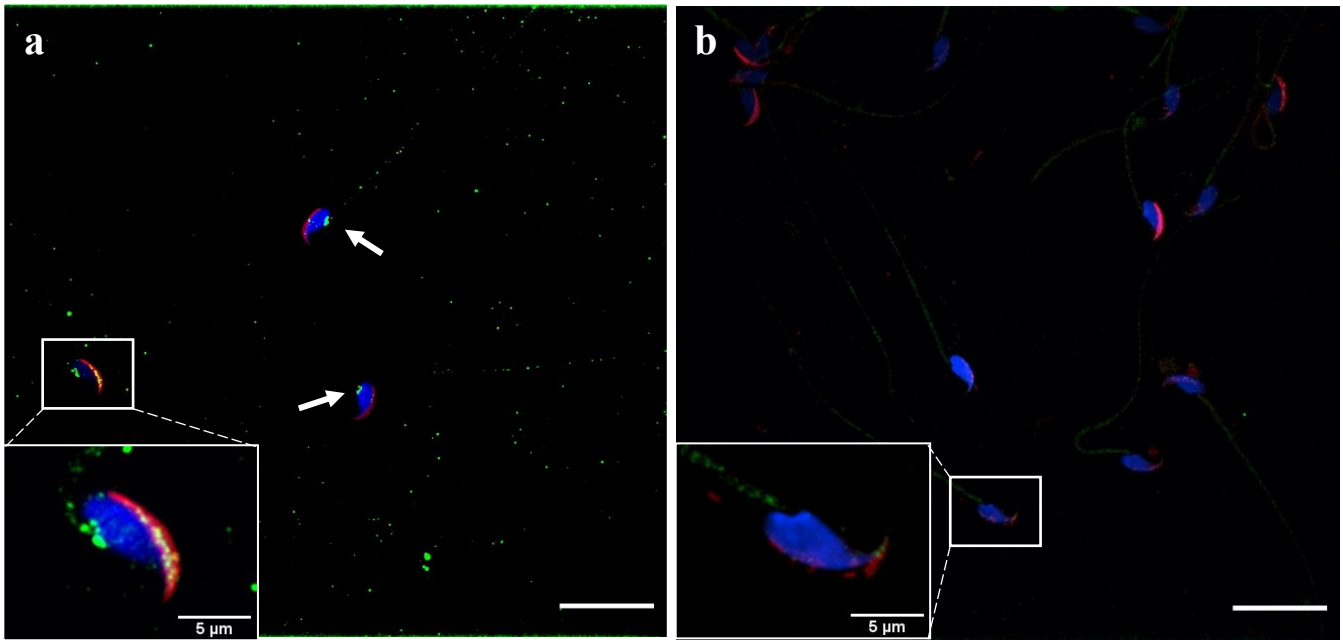

Figure 8: Immunolocalisation of MT-CO1 in the mouse spermatozoa released from cauda epididymis. (a) Localisation of MT-CO1 in released spermatozoa. The white arrows point to the connecting piece under head. **(b) S**permatozoa after acrosomal exocytosis. MT-CO1 = green, PNA = red, DAPI = blue, scale bar = 20 µm, objective: 100x oil immersion, NA = 1.4

4.1.3.1 LCN5 in cryosections of testes and epididymis

LCN5 was characterised with in situ hybridization and northern blot as epididymal specific protein, and is expressed mainly in the caput epididymis (Lareyre et al., 1998). Because both methods test only the occurrence of transcript, we aimed to detect the folded protein during sperm maturation in the epididymis. Moreover, proteomic data reveals its localisation in the acrosome, which was subsequently confirmed in experiment using immunofluorescence probe. Thus, we begun to question whether LCN5 could be transported to acrosome from epididymal secretion or is already expressed in the testes. For this purpose, we localised this protein in cryosections from caput, corpus, cauda epididymis and testes.

With this experiment, we confirmed that LCN5 is a secretory protein of the epithelia of mouse epididymal tissue (Fig. 9b, c, d). We have detected this protein in and on the surface of principal cells in the epididymal endothelia. The signal intensity decreased from caput to cauda, meaning that LCN5 is mainly expressed in caput epididymis. Moreover, we detected LCN5 already in the seminiferous tubules of the testes, forming cap-like structures in spermatocytes (Fig. 9a).

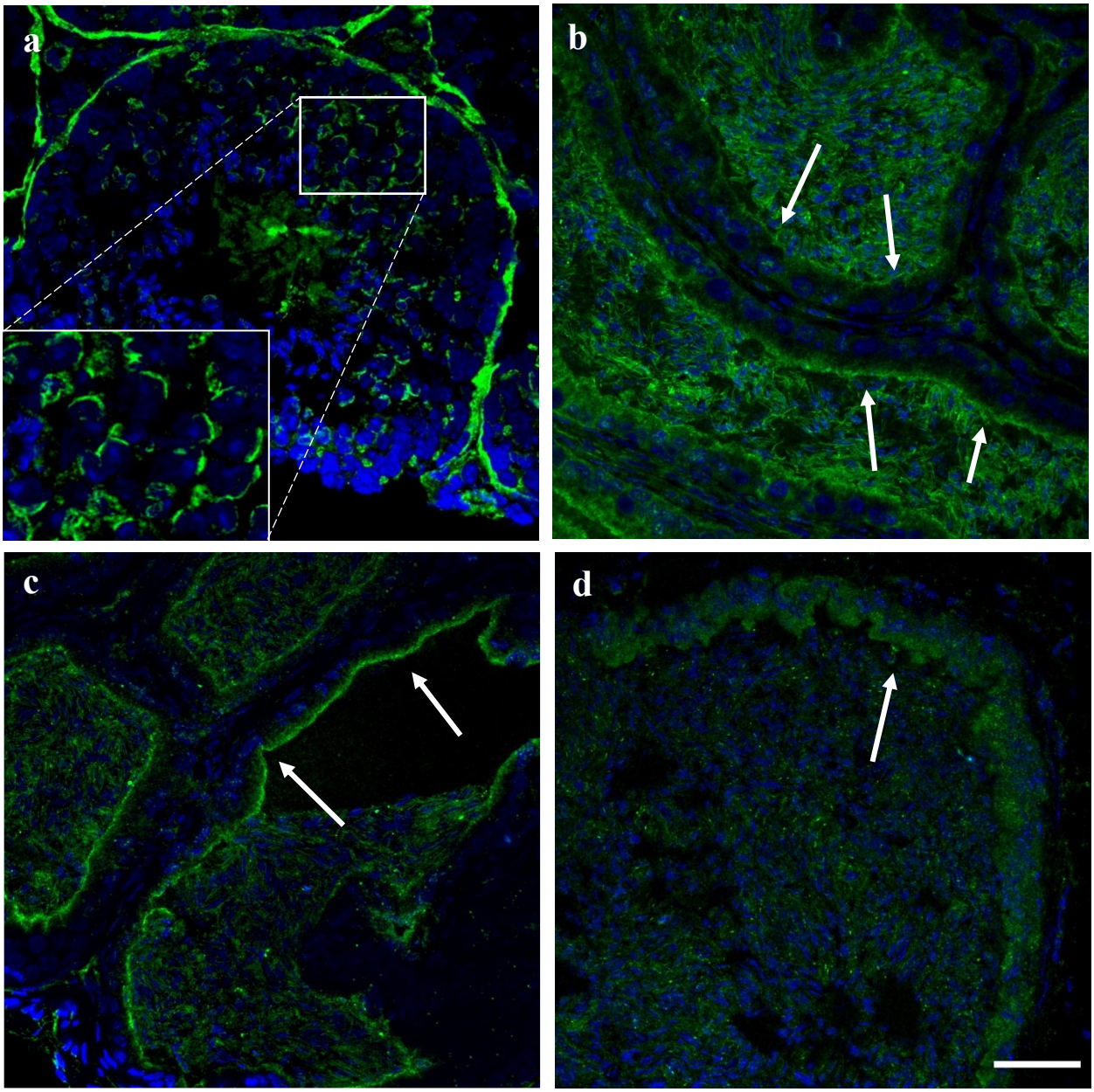

Figure 9: Immunolocalisation of LCN5 in cryosections of mouse testes (a), caput (b), corpus (d) and cauda epididymis (d). In the testes are highlighted spermatocytes with detected LCN5 forming cap-like structure. The white arrows point to the epithelia of epididymal tissue, where the LCN5 is secreted. LCN5 = green, DAPI = blue, scale bar = 40 um, objective: 63x oil immersion, NA = 1.4

5 Discussion

Localisation of particular proteins within the cells that were identified with LS-MS/MS is a great challenge in biology because it furthers the understanding of their functions in given cell components. In my master's thesis I have also touched a fundamental question in biology, namely, how to extract important candidate proteins in large biological datasets based on their importance. Here we combined relevant complementary approaches that increase the chance that detected proteins have important function in particular biological process, such as the fertilisation. These complementary methods included random forest for classification, analysis of differentially abundant proteins and gene ontology searches. To verify the localisation of candidates proteins we used indirect immunofluorescence.

5.1 Biological aspects

We have detected a total of 1229 proteins, of which 1148 proteins were detected in the acrosome and 1156 in the sperm cells. This is slightly more than in the study of Guyonnet *et al.*, (2012) who analysed the acrosomal matrix in the laboratory mouse. This number may differ due to a different protocol for acquiring acrosomal content and possibly due to a different number of mice. Guyonnet *et al.,* (2012) extracted samples from two caput epididymis and three from cauda epididymis. Three repeats (n = 3) of the experiment are considered the minimum. We used a sample from 10 mice (n = 10), which allowed us to perform deeper analysis and increase the signal-to-noise ratio.

The acrosome is an exocytic vesicle that helps sperm pass through the extracellular layer of the egg. The acrosomal reaction occurs after the sperm contact zona pellucida. Then, the acrosomal content, full of proteases, is released on the extracellular layers. The proteolytic enzyme digests the extracellular coat; thus, the sperm can easily pass through and fuse with the cytoplasmatic membrane of the egg. However, it has been demonstrated that most of the spermatozoa undergo acrosomal exocytosis before reaching extracellular layers of the egg and they are able to pass through and fertilize the oocyte (Jin et al., 2011; La Spina et al., 2016). Thus, the original hypothesis started to be questioned. In this work, we aimed to contribute with some findings to challenge this hypothesis and discuss the parallel function of acrosome as a waste storage vesicle. This statement is based on the evidence discussed below.

Currently, acrosome is described as a lysosome-related organelle (LRO). Previously, the acrosome was considered as an organelle related to the Golgi apparatus. And undoubtedly, the Golgi has an essential role in the acrosome biogenesis, especially when generating and trafficking proacrosomal granules during the Golgi-phase of the acrosome maturation. Nevertheless, several markers were revealed contributing to the hypothesis that acrosome is LRO. One of them is the acidic pH, which is about the same level as in the lysosome. The pH of the acrosome was estimated to be 4 - 5.3 and increasing during capacitation (Meizel & Dreamer, 1978; Nakanishi et al., 2001). The lysosomal pH ranges from 4 – 5, the values differ among studies (Li et al., 2019; Ohkuma & Poole, 1978). Additionally, several enzymes were detected in both structures – Cathepsin D (Srivastava & Ninjoor, 1982), hyaluronidase (Baba et al., 2002; Fiszer-Szafarz & Nadal, 1977; Reese et al., 2010; Strobl et al., 1998). Also, LAMP-2 and LAMP-1 (Lysosomal-associated membrane protein 2 and 1) were detected with microscopy in early and late spermatids, respectively. However, these two proteins were not detected in mature spermatozoa (Moreno, 2003). It is also considered that acrosome could arise from autophagosome, which fused with a lysosome in an early stage of acrosomal biogenesis and give rise to LRO. This would explain why the knockout of Atg7 (a protein involved in the autophagy induction) resulted in disruption of acrosomal biogenesis (Wang et al., 2014). Interestingly, the Random Forest analysis in our work identified Carbonic Anhydrase 2 (CAII) as an important enzyme for acrosome (Supplementary material, Tab. 1). CAII promotes acidification of the environment and was detected in the lysosome in human red blood cells (Rikihisa, 1985). We suppose that this enzyme could contribute to the acidic pH also in the acrosome. Moreover, RF also identified the AP1G1 protein important for acrosome. This protein is involved in the granules trafficking, containing lysosomal Cathepsin D, between Golgi and lysosome in humans (Supplementary material, Tab. 1) (Hirst et al., 2005). Therefore, if the acrosomal vesicle is LRO, it would explain why the RF detected these proteins as important for the acrosome.

Surprisingly, we have detected a group of proteins of the mitochondrial electron transport chain in the acrosome. Mitochondria are located in the mid-piece and function as a source of energy after the hyperactivation of the sperm cells. One of the detected mitochondrial proteins was MT-CO1, also known as a COX1 - the subunit of cytochrome c oxidase, complex IV of electron transport chain. To our surprise, this protein was more abundant in the acrosome when compared to sperm-cell – 7-fold difference (based on LC-MS/MS signal). Although the depletion/mutation of this gene is connected to pathologies, one study also suggests a

physiological absence of the MT-CO1 (Bernstein et al., 2010). The Mt-co*1* gene is often mutated in the colonic crypt of the gastrointestinal tract and the cells with mutated Mt-co1 colonize the entire crypt. Authors referred that mitochondrion with mutated Mt-co1 must have a positive selection bias to expand in the entire cell. Possible explanation is that mitochondria without MT-CO1 could lead to a lower production of ROS and oxidative damage of the cell (Bernstein et al., 2010). Although we do not have any evidence of the mutated form of Mt-co1 in the mitochondrion of the sperm cells, the protein subunit was present in the acrosome and in less number in the acrosome-reacted sperm cells in our work. Interestingly, we detected the signal from MT-CO1 in released spermatozoa also under the head, in the connecting piece. This signal was no longer present in the capacitated spermatozoa after acrosomal exocytosis. It is known that paternal mitochondria are being degraded after fertilization, and only maternal mitochondria are transmitted through generations. Although this seems very unlikely, it could be a mark that the degradation of redundant mitochondrial protein begun much earlier. It would be interesting to see if the signal intensity is changing during the maturation and capacitation of spermatozoa. This would need further testing.

Furthermore, we have detected several members of lipocalin family to be abundant in the acrosome. As mentioned in chapter 1.1.2.1, lipocalins have various functions in mammals, such as chemical communication or binding a toxic waste. For instance, MUPs are expressed mainly in the liver where they bind and transport volatile organic compounds - pheromones and excrete them to urine to ensure chemical communication in mice. This is the case of DARCIN/MUP20, which carries signals underlying dominance status in mice in males (Nelson et al., 2015). However, it was shown that MUPs can also bind xenobiotics and chemicals used in the industry and subsequently, they are excreted into the urine (Kwak et al., 2011). Similarly, OBP can bind 4-hydroxynonenal (HNE) – a toxic product derived from lipid peroxidation (Grolli et al., 2006). This protein was detected in high concentrations in the nasal mucosa, which is constantly exposed to harmful substances inhaled from air. Therefore, these harmful metabolic products may bind to lipocalins before they are degraded by relevant enzymes. This early trapping can prevent chemical modifications and subsequent damage caused by these harmful substances. Based on these studies, we predict that lipocalins could have a similar function in the sperm. We detected several lipocalins in our study, such as FABP9, MUP9, MUP10, MUP20, MUP3, MUP17, LCN5, LCN12, APOE, APOO, PTGDS, and PTGDS1 (data not shown). Random forest identified LCN5 and PTGDS as important for acrosomal vesicle. LCN5 binds all-trans and 9-cis RA. PTGDS2 is an isomerase, which

catalyses the conversion of prostaglandin H2 to prostaglandin D2. The findings of existing literature do not indicate that these two proteins could serve as a scavenger of harmful waste. However, we do not know whether these proteins localised to the acrosome with a bounded substrate or they were involved in another biological process in the sperm cell and are stored in the acrosome as waste. Moreover, we detected the LCN5 in the testes and epididymal tissue, which makes LCN5 no longer a specific epididymal protein. We assume that the lipocalins could function as scavengers for metabolic/toxic waste, which is subsequently stored and metabolized in the acrosome. However, this hypothesis should be further tested.

As is previously mentioned, the maturation of spermatozoa does not take place only in the testes. The sperm cells undergo a maturation process in the epididymis, during ejaculation in the seminal plasma and in the urogenital tract of the females. Inevitably, organic waste must be generated during these maturation steps. The metabolic products of somatic cells are degraded in the lysosome, autophagosome or released in the exosome to the extracellular environment. We assume that the sperm cells can store this waste in the acrosome, which can temporarily serve as a protective structure of the sperm head during ejaculation. Proteins essential for the fusion of the gametes are revealed after the acrosomal exocytosis (Toshimori *et al.*, 1998; Satouh *et al.*, 2012; Barbaux *et al.*, 2020). By this mechanism, it can be ensured that sperm cells bring to the egg only the necessary material – genetic information. This hypothesis is supported by the evidence mentioned above. Moreover, it has been described that in the promiscuous mice, such as *Apodemus sylvaticus*., the acrosomal exocytosis occurs much earlier than in less promiscuous, *Apodemus falvicollis* and *Mus musculus musculus* (Johnson et al., 2007). The advantage of releasing acrosomal content earlier could be in the reduced mass of sperm cells and thus, a faster movement towards the egg. It is not an exception that females of promiscuous *Apodemus sylvaticus* copulate with many males in a short time and have sperm cells from different males in her urogenital tract. Therefore, the premature acrosomal reaction could have evolved due to the competition of sperm at the level of individuals. To add, we assume that the acrosome could serve as a vesicle, where the spermatozoa place various degradation/toxic products during the maturation process.

In addition, there is a place to discuss the SPINK2 – Kazal type inhibitor of trypsin/acrosin, which was identified by random forest analysis with the highest importance value. Although this protein localises in the acrosome of mature sperm cells according to Kherraf *et al.*, (2017), RF analysis classifies it as an important sperm cell protein. Their finding is contrary

to our data; we identified the SPINK2 in higher concentration in sperm cells with a 4.5-fold difference compared to the acrosome. Our finding is in agreement with Zigo *et al.*, (2019). Zigo *et al.*, (2019) detected the SPINK2 in the post-acrosomal region of 90% of bull spermatozoa whereas only 10% of spermatozoa in the acrosome. Different localisation of the SPINK2 could be given by different stages of spermatozoa maturation. However, it is not clear where is the final destination of SPINK2 and where it fulfills its role as a protease inhibitor. Kherraf *et al.*, (2017) suppose that SPINK2 inhibits the protease activity during transport through Golgi and proacrosomal granules to acrosome in maturating spermatids because the deficiency of SPINK2 resulted in the fragmentation of Golgi apparatus. Nevertheless, this does not explain its localisation in the post-acrosomal region in the maturated sperm cells and further research of SPINK2 function is needed.

5.2 Technical aspects

We used non-targeted proteomics and bioinformatics of proteomes of the acrosome and acrosome-reacted sperm cells. Approaches of non-targeted proteomics have many advantages and allow us to study structures as whole units on the level of proteins, transcripts, or metabolites. However, if we do not analyse whole cells or tissues but subcellular structures, protocols for obtaining these structures tend to be long and demanding and one should think about every step of the protocol in the context of cellular processes. For example, if we consider the process of acrosomal exocytosis, during which the cytoplasmic membrane fuses with an outer acrosomal membrane in several places. Thus, smaller lipid vesicles are formed during this process containing non-acrosomal proteins. In the protocol, we separated the acrosomal content from the acrosome-reacted sperm cells by centrifugation. It can be assumed that these vesicles are concentrated between the pellet and supernatant after the centrifugation due to their different weight/density. And thus, we could introduce bias to the proteomic dataset. We let the pellet (acrosome-reacted sperm cell) with a small amount of supernatant to avoid this error, where we assume these residual lipid vesicles could be settled. However, proteins of the inner acrosomal membrane remained in the pellet and they are included in the sperm proteome in the analysis. The second discussed pitfall is the experimental conditions where spermatozoa undergo the maturation process before acrosomal exocytosis. In mice, sperm cells cannot be collected any other way than after killing the animal and isolating them from excised cauda epididymis. However, the sperm maturation process does not end in the

cauda epididymis but continues during ejaculation. Sperm cells are exposed to secretion from other accessory glands, such as the bulbourethral glands, prostate, seminal vesicles, and in the female reproductive tract, to an acidic uterovaginal environment. This is mimicked by M2 medium, which to some extent simulates the natural conditions. Nevertheless, we assume that since sperm cells collected from the cauda epididymis and capacitated in vitro are able to fertilize oocytes and give birth to healthy pups, these conditions are still supportive enough to help sperm accomplish its goal. Therefore, we assume that our results are biologically relevant.

In this work, we used the machine learning approach. Besides its other uses, machine learning approaches are powerful tools for detecting and extracting patterns in large datasets, which become typical in proteomics, transcriptomics, or metabolomics and other fields. We used Breiman's Random Forest algorithm based on decision trees (Breiman, 2001). These decision trees try to find in each iteration a variable that would divide the group of features (in this work proteins) into the smaller homogeneous groups with the highest possible information gain. To follow variables of one decision tree is relatively easy and possible to interpret, as its decision rules can be visualized to observe the dividing variables. But to follow the variables of hundreds of trees is much harder. Thus, the forest of trees sacrifices the intrinsic interpretability present in one decision tree. However, the number of decision trees in the random forest offers an advantage. Because the choice does not stand on one tree but on the whole forest, they protect each other from individual error. If one tree makes the wrong decision, its mistake will be overruled by the decisions of other trees. Together, they try to select those features (proteins), which are more statistically important. Furthermore, one should be careful about interpreting the variable Importance, which is one of the resulting values of the random forest analysis. As was said earlier, in each decision split, the tree is searching for the variable that will divide the group into two smaller homogeneous groups with some internal entropy. Then the importance of every feature (protein) is calculated as a value of impurity reduction/increasing entropy of the divided group due to the feature.

6 Conclusion

In my thesis, I aimed to further the understanding of the acrosomal function. Overall, the proteomic analysis revealed 1229 proteins, 50.6 % differentially abundant within our tested samples. 27.3 % were enriched in the acrosome and 23.4 % in the acrosome-reacted sperm cells. Indeed, this enrichment of proteins shows that the acrosome is a unique structure. From these great number of proteins, 73 were uniquely detected in the acrosome and 81 in the acrosome-reacted sperm cells. Next, we identified important proteins within our tested proteomes using the machine learning approach, the random forest. Important proteins of acrosome-reacted sperm cells were mainly involved in sperm motility and oxidative phosphorylation, whereas acrosomal proteins have a much broader spectrum of functions. Finally, the localisation of proteins in the acrosome was verified with immunofluorescence probes. Furthermore, this study provides new evidence contributing to the relatively novel hypothesis that acrosome is a lysosome-related organelle.

7 References

Baba, D., Kashiwabara, S., Honda, A., Yamagata, K., Wu, Q., Ikawa, M., Okabe, M., & Baba, T. (2002). Mouse Sperm Lacking Cell Surface Hyaluronidase PH-20 Can Pass through the Layer of Cumulus Cells and Fertilize the Egg *. *Journal of Biological Chemistry*, *277*(33), 30310–30314.

Barbaux, S., Ialy-Radio, C., Chalbi, M., Dybal, E., Homps-Legrand, M., Do Cruzeiro, M., Vaiman, D., Wolf, J. P., & Ziyyat, A. (2020). Sperm SPACA6 protein is required for mammalian Sperm-Egg Adhesion/Fusion. *Scientific Reports*, *10*(1).

Barral, S., Morozumi, Y., Tanaka, H., Montellier, E., Govin, J., de Dieuleveult, M., Charbonnier, G., Couté, Y., Puthier, D., Buchou, T., Boussouar, F., Urahama, T., Fenaille, F., Curtet, S., Héry, P., Fernandez-Nunez, N., Shiota, H., Gérard, M., Rousseaux, S., … Khochbin, S. (2017). Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Molecular Cell*, *66*(1), 89-101.

Bastien, J., & Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene*, *328*(1–2), 1–16.

Bernstein, C., Facista, A., Nguyen, H., Zaitlin, B., Hassounah, N., Loustaunau, C., Payne, C. M., Banerjee, B., Goldschmid, S., Tsikitis, V. L., Krouse, R., & Bernstein, H. (2010). Cancer and age related colonic crypt deficiencies in cytochrome c oxidase I. *, 2*(12), 429–442.

Bleil, J. D., & Wassarman, P. M. (1990). Identification of a ZP3-binding protein on acrosome-intact mouse sperm by photoaffinity crosslinking. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(14), 5563–5567.

Boucheron-Houston, C., Canterel-Thouennon, L., Lee, T.-L., Baxendale, V., Nagrani, S., Chan, W.-Y., & Rennert, O. M. (2013). Long-Term Vitamin-A-Deficiency Induces Alteration of Adult Mouse Spermatogenesis and Spermatogonial Differentiation: direct effect on spermatogonial gene expression and indirect effects via somatic cells. *The Journal of Nutritional Biochemistry*, *24*(6), 1123.

Breiman, L. (2001). Random Forests. *Machine Learning 2001 45:1*, *45*(1), 5–32.

Brooks, D. E., & Higgins, S. J. (1980). Characterization and androgen-dependence of proteins associated with luminal fluid and spermatozoa in the rat epididymis. *Journal of Reproduction and Fertility*, *59*(2), 363–375.

Buffone, M. G., Foster, J. A., & Gerton, G. L. (2008). The role of the acrosomal matrix in fertilization. *International Journal of Developmental Biology*, *52*(5–6), 511–522.

Buffone, M. G., Kim, K.-S., Doak, B. J., Rodriguez-Miranda, E., & Gerton, G. L. (2009). Functional consequences of cleavage, dissociation and exocytotic release of ZP3R, a C4BP-related protein, from the mouse sperm acrosomal matrix. *Journal of Cell Science*, *122*(17), 3153.

Cao, K.-A. L., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics 2011 12:1*, *12*(1), 1–17.

Carvajal, G., Brukman, N. G., Muñoz, M. W., Battistone, M. A., Guazzone, V. A., Ikawa, M., Haruhiko, M., Lustig, L., Breton, S., & Cuasnicu, P. S. (2018). Impaired male fertility and abnormal epididymal epithelium differentiation in mice lacking CRISP1 and CRISP4. *Scientific Reports*, *8*(1).

Conine, C. C., Sun, F., Song, L., Rivera-Pérez, J. A., & Rando, O. J. (2018). Small RNAs gained during epididymal transit of sperm are essentialfor embryonic development in mice. *Developmental Cell*, *46*(4), 470.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics : MCP*, *13*(9), 2513.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*, *10*(4), 1794–1805.

Endo, T., Romer, K. A., Anderson, E. L., Baltus, A. E., Rooij, D. G. de, & Page, D. C. (2015). Periodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), E2347.

Ferrer, M., Rodriguez, H., Zara, L., Yu, Y., Xu, W., & Oko, R. (2012). MMP2 and acrosin are major proteinases associated with the inner acrosomal membrane and may cooperate in sperm penetration of the zona pellucida during fertilization. *Cell and Tissue Research*, *349*(3), 881–895.

Fiszer-Szafarz, B., & Nadal, C. (1977). Lysosomal enzyme activities in the regenerating rat liver. *Cancer Research*, *37*(2), 354–357.

Florman, H. M., & Storey, B. T. (1982). Mouse gamete interactions: The zona pellucida is the site of the acrosome reaction leading to fertilization in vitro. *Developmental Biology*, *91*(1), 121–130.

Funaki, T., Kon, S., Tanabe, K., Natsume, W., Sato, S., Shimizu, T., Yoshida, N., Wong, W. F., Ogura, A., Ogawa, T., Inoue, K., Ogonuki, N., Miki, H., Mochida, K., Endoh, K., Yomogida, K., Fukumoto, M., Horai, R., Iwakura, Y., … Satake, M. (2013). The Arf GAP SMAP2 is necessary for organized vesicle budding from the trans-Golgi network and subsequent acrosome formation in spermiogenesis. *Molecular Biology of the Cell*, *24*(17), 2633–2644.

Gely-Pernot, A., Raverdeau, M., Teletin, M., Vernet, N., Féret, B., Klopfenstein, M., Dennefeld, C., Davidson, I., Benoit, G., Mark, M., & Ghyselinck, N. B. (2015). Retinoic Acid Receptors Control Spermatogonia Cell-Fate and Induce Expression of the SALL4A Transcription Factor. *PLoS Genetics*, *11*(10), 1005501.

Gewiss, R., Topping, T., & Griswold, M. D. (2020). Cycles, waves, and pulses: Retinoic acid and the organization of spermatogenesis. *Andrology*, *8*(4), 892–897.

Gottlieb, W., & Meizel, S. (1987). Biochemical Studies of Metalloendoprotease Activity in the Spermatozoa of Three Mammalian Species. *Journal of Andrology*, *8*(1), 14–24.

Grolli, S., Merli, E., Conti, V., Erika, S., & Ramoni, R. (2006). Odorant binding protein has the biochemical properties of a scavenger for 4-hydroxy-2-nonenal in mammalian nasal mucosa. *The FEBS Journal*, *273*(22), 5131–5142.

Guyonnet, B., Egge, N., & Cornwall, G. A. (2014). Functional Amyloids in the Mouse Sperm Acrosome. *Molecular and Cellular Biology*, *34*(14), 2624–2634.

Guyonnet, Benoit, Zabet-Moghaddam, M., SanFrancisco, S., & Cornwall, G. A. (2012). Isolation and proteomic characterization of the mouse sperm acrosomal matrix. *Molecular and Cellular Proteomics*, *11*(9), 758–774.

Han, F., Liu, C., Zhang, L., Chen, M., Zhou, Y., Qin, Y., Wang, Y., Duo, S., Cui, X., Bao, S., & Gao, F. (2017). Globozoospermia and lack of acrosome formation in GM130-deficient mice. *Cell Death and Disease*, *8*(1).

Hazzouri, M., Pivot-Pajot, C., Faure, A. K., Usson, Y., Pelletier, R., Sèle, B., Khochbin, S., &

Rousseaux, S. (2000). Regulated hyperacetylation of core histones during mouse spermatogenesis: Involvement of histone-deacetylases. *European Journal of Cell Biology*, *79*(12), 950–960.

Hirst, J., Borner, G. H., Harbour, M., & Robinson, M. S. (2005). The aftiphilin/p200/gamma-synergin complex. *Molecular Biology of the Cell*, *16*(5), 2554–2565.

Huang, H. F. S., & Hembree, W. C. (1979). Spermatogenic Response to Vitamin A in Vitamin A Deficient Rats. *Biology of Reproduction*, *21*, 891–904.

Isoherranen, N., & Zhong, G. (2019). Biochemical and Physiological Importance of the CYP26 Retinoic Acid Hydroxylases. *Pharmacology & Therapeutics*, *204*, 107400.

Jauregui, E. J., Mitchell, D., Topping, T., Hogarth, C. A., & Griswold, M. D. (2018). Retinoic acid receptor signaling is necessary in steroidogenic cells for normal spermatogenesis and epididymal function. *Development (Cambridge, England)*, *145*(13).

Jin, M., Fujiwara, E., Kakiuchi, Y., Okabe, M., Satouh, Y., Baba, S. A., Chiba, K., & Hirohashi, N. (2011). Most fertilizing mouse spermatozoa begin their acrosome reaction before contact with the zona pellucida during in vitro fertilization. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 4892–4896.

Johnson, P. M., Clift, L. E., Andrlikova, P., Jursova, M., Flanagan, B. F., Cummerson, J. A., Stopka, P., & Dvorakova-Hortova, K. (2007). Rapid sperm acrosome reaction in the absence of acrosomal CD46 expression in promiscuous field mice (Apodemus). *Reproduction, 134(6), 739-747.*

Jung, Y. H., Sauria, M. E. G., Lyu, X., Cheema, M. S., Ausio, J., Taylor, J., & Corces, V. G. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Reports*, *18*(6), 1366–1382.

Kanemori, Y., Koga, Y., Sudo, M., Kang, W., Kashiwabara, S. I., Ikawa, M., Hasuwa, H., Nagashima, K., Ishikawa, Y., Ogonuki, N., Ogura, A., & Baba, T. (2016). Biogenesis of sperm acrosome is regulated by pre-mRNA alternative splicing of acrbp in the mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *113* (26) E3696-E3705

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, *88*(3), 375–384.

Kherraf, Z., Christou-Kent, M., Karaouzene, T., Amiri-Yekta, A., Martinez, G., Vargas, A. S., Lambert, E., Borel, C., Dorphin, B., Aknin-Seifer, I., Mitchell, M. J., Metzler-Guillemain, C., Escoffier, J., Nef, S., Grepillat, M., Thierry-Mieg, N., Satre, V., Bailly, M., Boitrelle, F., … Arnoult, C. (2017). SPINK 2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Molecular Medicine*, *9*(8), 1132–1149.

Kierszenbaum, A. L., Rivkin, E., & Tres, L. L. (2003). Acroplaxome, an F-Actin-Keratin-containing Plate, Anchors the Acrosome to the Nucleus during Shaping of the Spermatid Head. *Molecular Biology of the Cell*, *14*(11), 4628–4640.

Kim, K.-S., Cha, M. C., & Gerton, G. L. (2001). Mouse Sperm Protein sp56 Is a Component of the Acrosomal Matrix. *Biology of Reproduction*, *64*(1), 36–43.

Kim, K. S., Foster, J. A., Kvasnicka, K. W., & Gerton, G. L. (2011). Transitional states of acrosomal exocytosis and proteolytic processing of the acrosomal matrix in guinea pig sperm. *Molecular Reproduction and Development*, *78*(12), 930–941.

Kim, K. S., & Gerton, G. L. (2003). Differential release of soluble and matrix components: Evidence for intermediate states of secretion during spontaneous acrosomal exocytosis in mouse sperm. *Developmental Biology*, *264*(1), 141–152.

Kimura, M., Kim, E., Kang, W., Yamashita, M., Saigo, M., Yamazaki, T., Nakanishi, T., Kashiwabara, S., & Baba, T. (2009). Functional Roles of Mouse Sperm Hyaluronidases, HYAL5 and SPAM1, in Fertilization1. *Biology of Reproduction*, *81*(5), 939–947.

Kwak, J., Josue, J., Faranda, A., Curran Opiekun, M., Preti, G., Osada, K., Yamazaki, K., & Beauchamp, G. K. (2011). Butylated hydroxytoluene is a ligand of urinary proteins derived from female mice. *Chemical Senses*, *36*(5), 443–452.

La Spina, F. A., Puga Molina, L. C., Romarowski, A., Vitale, A. M., Falzone, T. L., Krapf, D., Hirohashi, N., & Buffone, M. G. (2016). Mouse sperm begin to undergo acrosomal exocytosis in the upper isthmus of the oviduct. *Developmental Biology*, *411*(2), 172–182.

Lareyre, J.-J., Thomas, T. Z., Zheng, W.-L., Kasper, S., Ong, D. E., Orgebin-Crist, M.-C., & Matusik, R. J. (1999). A 5-Kilobase Pair Promoter Fragment of the Murine Epididymal Retinoic Acid-binding Protein Gene Drives the Tissue-specific, Cell-specific, and Androgen-regulated Expression of a Foreign Gene in the Epididymis of Transgenic Mice *. *Journal of Biological Chemistry*, *274*(12), 8282–8290.

Lareyre, J.-J., Zheng, W.-L., Zhao, G.-Q., Kasper, S., Newcomer, M. E., Matusik, R. J., Ong, D. E., & Orgebin-Crist, M.-C. (1998). Molecular Cloning and Hormonal Regulation of a Murine Epididymal Retinoic Acid-Binding Protein Messenger Ribonucleic Acid. *Endocrinology*, *139*(6), 2971–2981.

Lê Cao, K.-A., Rossouw, D., Robert-Granié, C., & Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology*, *7*(1).

Lê Cao, K., González, I., & Déjean, S. (2009). integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics*, *25*(21), 2855.

Li, S.-S., Zhang, M., Wang, J.-H., Yang, F., Kang, B., Xu, J.-J., & Chen, H.-Y. (2019). Monitoring the Changes of pH in Lysosomes during Autophagy and Apoptosis by Plasmon Enhanced Raman Imaging. *Analytical Chemistry*, *91*(13), 8398–8405.

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*(3). http://www.stat.berkeley.edu/

Lin, Y.-N., Roy, A., Yan, W., Burns, K. H., & Matzuk, M. M. (2007a). Loss of Zona Pellucida Binding Proteins in the Acrosomal Matrix Disrupts Acrosome Biogenesis and Sperm Morphogenesis. *Molecular and Cellular Biology*, *27*(19), 6794.

Lin, Y.-N., Roy, A., Yan, W., Burns, K. H., & Matzuk, M. M. (2007b). Loss of Zona Pellucida Binding Proteins in the Acrosomal Matrix Disrupts Acrosome Biogenesis and Sperm Morphogenesis. *Molecular and Cellular Biology*, *27*(19), 6794–6805.

Luense, L. J., Donahue, G., Lin-Shiao, E., Rangel, R., Weller, A. H., Bartolomei, M. S., & Berger, S. L. (2019). Gcn5-Mediated Histone Acetylation Governs Nucleosome Dynamics in Spermiogenesis. *Developmental Cell*, *51*(6), 745-758.

Ma, H.-T., Niu, C.-M., Xia, J., Shen, X.-Y., Xia, M.-M., Hu, Y.-Q., & Zheng, Y. (2018). Stimulated by retinoic acid gene 8 (Stra8) plays important roles in many stages of spermatogenesis. *Asian Journal of Andrology*, *20*(5), 479.

Maire, A. le, Teyssier, C., Balaguer, P., Bourguet, W., & Germain, P. (2019). Regulation of RXR-RAR Heterodimers by RXR- and RAR-Specific Ligands and Their Combinations. *Cells*, *8*(11).

Mbikay, M., Tadros, H., Ishida, N., Lerner, C. P., De Lamirande, E., Chen, A., El-Alfy, M., Clermont, Y., Seidah, N. G., Chrétien, M., Gagnon, C., & Simpson, E. M. (1997).

Impaired fertility in mice deficient for the testicular germ-cell protease PC4. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 6842–6846.

Medarde, N., Martínez-Vargas, J., Sánchez-Chardi, A., López-Fuster, M. J., & Ventura, J. (2013). Effect of Robertsonian translocations on sperm head form in the house mouse. *Biological Journal of the Linnean Society*, *110*(4), 878–889.

Meizel, S., & Dreamer, D. W. (1978). The pH of the hamster sperm acrosome. *The Journal of Histochemistry and Cytochemistry : Official Journal of the Histochemistry Society*, *26*(2), 98–105.

Modelski, M. J., Menlah, G., Wang, Y., Dash, S., Wu, K., Galileo, D. S., & Martin-DeLeon, P. A. (2014). Hyaluronidase 2: A Novel Germ Cell Hyaluronidase with Epididymal Expression and Functional Roles in Mammalian Sperm. *Biology of Reproduction*, *91*(5), 1–11.

Moreno, R. D. (2003). Differential expression of lysosomal associated membrane protein (LAMP-1) during mammalian spermiogenesis. *Molecular Reproduction and Development*, *66*(2), 202–209.

Moreno, R. D., Ramalho-Santos, J., Sutovsky, P., Chan, E. K. L., & Schatten, G. (2000). Vesicular Traffic and Golgi Apparatus Dynamics During Mammalian Spermatogenesis: Implications for Acrosome Architecture1. *Biology of Reproduction*, *63*(1), 89–98.

Morohoshi, A., Miyata, H., Shimada, K., Nozawa, K., Matsumura, T., Yanase, R., Shiba, K., Inaba, K., & Ikawa, M. (2020). Nexin-Dynein regulatory complex component DRC7 but not FBXL13 is required for sperm flagellum formation and male fertility in mice. *PLoS Genetics*, *16*(1).

Nakanishi, T., Ikawa, M., Yamada, S., Toshimori, K., & Okabe, M. (2001). Alkalinization of Acrosome Measured by GFP as a pH Indicator and Its Relation to Sperm Capacitation. *Developmental Biology*, *237*(1), 222–231.

Nelson, A. C., Cunningham, C. B., Ruff, J. S., & Potts, W. K. (2015). Protein pheromone expression levels predict and respond to the formation of social dominance networks. *Journal of Evolutionary Biology*, *28*(6), 1213.

Newcomer, M. E. (1993). Structure of the epididymal retinoic acid binding protein at 2.1 Å resolution. *Structure*, *1*(1), 7–18.

Nixon, B., Iuliis, G. N. De, Hart, H. M., Zhou, W., Mathe, A., Bernstein, I. R., Anderson, A.

L., Stanger, S. J., Skerrett-Byrne, D. A., Jamaluddin, M. F. B., Almazi, J. G., Bromfield, E. G., Larsen, M. R., & Dun, M. D. (2019). Proteomic Profiling of Mouse Epididymosomes Reveals their Contributions to Post-testicular Sperm Maturation. *Molecular & Cellular Proteomics : MCP*, *18*(Suppl 1), S91.

Ohkuma, S., & Poole, B. (1978). Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(7), 3327.

Ohmura, K., Kohno, N., Kobayashi, Y., Yamagata, K., Sato, S., Kashiwabara, S. I., & Baba, T. (1999). A homologue of pancreatic trypsin is localized in the acrosome of mammalian sperm and is released during acrosome reaction. *Journal of Biological Chemistry*, *8; 274(41):29426-32.*

Oko, R., Aarabi, M., Mao, J., Balakier, H., Sutovsky, P., & Yanagimachi, R. (2017). Sperm-Specific WW-Domain-Binding Proteins. In C. J. De Jonge & C. L. R. Barratt (Eds.), *The Sperm Cell* (pp. 157–176). Cambridge University Press.

Ong, D. E., Newcomer, M. E., Lareyre, J. J., & Orgebin-Crist, M. C. (2000). Epididymal retinoic acid-binding protein. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, *1482*(1–2), 209–217.

Oulad-Abdelghani, M., Bouillet, P., Décimo, D., Gansmuller, A., Heyberger, S., Dollé, P., Bronner, S., Lutz, Y., & Chambon, P. (1996). Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene. *The Journal of Cell Biology*, *135*(2), 469.

Pavelka, N., Pelizzola, M., Vizzardelli, C., Capozzoli, M., Splendiani, A., Granucci, F., & Ricciardi-Castagnoli, P. (2004). A power law global error model for the identification of differentially expressed genes in microarray data. *BMC Bioinformatics 2004 5:1*, *5*(1), 1–12.

Pierre, V., Martinez, G., Coutton, C., Delaroche, J., Yassine, S., Novella, C., Pernet-Gallay, K., Hennebicq, S., Ray, P. F., & Arnoult, C. (2012). Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. *Development*, *139*(16), 2955–2965.

Plante, G., & Manjunath, P. (2015). Murine binder of sperm protein homolog 1: a new player in HDL-induced capacitation. *Reproduction*, *149*(4), 367–376.

Pujianto, D. A., Loanda, E., Sari, P., Midoen, Y. H., & Soeharso, P. (2013). Sperm-associated antigen 11A is expressed exclusively in the principal cells of the mouse caput epididymis in an androgen-dependent manner. *Reproductive Biology and Endocrinology 2013 11:1*, *11*(1), 1–12.

Pyttel, S., Nimptsch, A., Böttger, J., Zschörnig, K., Jakop, U., Wegener, J., Müller, K., Paasch, U., & Schiller, J. (2014). Changes of murine sperm phospholipid composition during epididymal maturation determined by MALDI-TOF mass spectrometry. *Theriogenology*, *82*(3), 396–402.

Reese, K. L., Aravindan, R. G., Griffiths, G. S., Shao, M., Wang, Y., Galileo, D. S., Atmuri, V., Triggs-Raine, B. L., & Martin-DeLeon, P. A. (2010). Acidic hyaluronidase activity is present in mouse sperm and is reduced in the absence of SPAM1: Evidence for a Role for Hyaluronidase 3 in mouse and human sperm. *Molecular Reproduction and Development*, *77*(9), 759.

Reilly, J. N., McLaughlin, E. A., Stanger, S. J., Anderson, A. L., Hutcheon, K., Church, K., Mihalas, B. P., Tyagi, S., Holt, J. E., Eamens, A. L., & Nixon, B. (2016). Characterisation of mouse epididymosomes reveals a complex profile of microRNAs and a potential mechanism for modification of the sperm epigenome. *Scientific Reports*, *6*.

Rikihisa, Y. (1985). Ultrastructural localization of carbonic anhydrase in lysosomes. *The Anatomical Record*, *211*(1), 1–8.

R Core Team (2020). R: A language and enviroment for statistical computing. R foudation for statistical computing, Vienna, Austria, URL https://R-project.org/.

Rohart, F., Gautier, B., Singh, A., & Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752.

Satouh, Y., Inoue, N., Ikawa, M., & Okabe, M. (2012). Visualization of the moment of mouse sperm-egg fusion and dynamic localization of IZUMO1. *Journal of Cell Science*, *125*(21), 4985–4990.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods 2012 9:7*, *9*(7), 676–682.

Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., Belleannee, C., Kucukural, A., Serra, R. W., Sun, F., Song, L., Carone, B. R., Ricci, E. P., Li, X. Z., Fauquier, L., Moore, M. J., Sullivan, R., Mello, C. C., Garber, M., & Rando, O. J. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science (New York, N.Y.)*, *351*(6271), 391.

Sharma, U., Sun, F., Conine, C. C., Reichholf, B., Kukreja, S., Herzog, V. A., Ameres, S. L., & Rando, O. J. (2018). Small RNAs are trafficked from the epididymis to developing mammalian sperm. *Developmental Cell*, *46*(4), 481.

Skerget, S., Rosenow, M. A., Petritis, K., & Karr, T. L. (2015). Sperm Proteome Maturation in the Mouse Epididymis. *PLoS ONE*, *10*(11).

Srivastava, P. N., & Ninjoor, V. (1982). Isolation of rabbit testicular cathepsin D and its role in the activation of proacrosin. *Biochemical and Biophysical Research Communications*, *109*(1), 63–69.

Stopkova, R., Klempt, P., Kuntova, B., & Stopka, P. (2017). On the tear proteome of the house mouse (Mus musculus musculus) in relation to chemical signalling. *PeerJ*, *5*(7).

Strobl, B., Wechselberger, C., Beier, D. R., & Lepperdinger, G. (1998). Structural Organization and Chromosomal Localization ofHyal2,a Gene Encoding a Lysosomal Hyaluronidase. *Genomics*, *53*(2), 214–219.

Sutovsky, P., Manandhar, G., McCauley, T. C., Caamaño, J. N., Sutovsky, M., Thompson, W. E., & Day, B. N. (2004). Proteasomal Interference Prevents Zona Pellucida Penetration and Fertilization in Mammals. *Biology of Reproduction*, *71*(5), 1625–1637.

Talbot, P., & Dicarlantonio, G. (1985). Cytochemical localization of dipeptidyl peptidase II (DPP-II) in mature guinea pig sperm. *Journal of Histochemistry & Cytochemistry*, *33*(11), 1169–1172.

Tardif, S., Wilson, M. D., Wagner, R., Hunt, P., Gertsenstein, M., Nagy, A., Lobe, C., Koop, B. F., & Hardy, D. M. (2010). Zonadhesin is essential for species specificity of sperm adhesion to the egg zona pellucida. *Journal of Biological Chemistry*, *285*(32), 24863–24870.

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., …

Teodoro, D. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489.

Toshimori, K., Saxena, D. K., Tanii, I., & Yoshinaga, K. (1998). An MN9 Antigenic Molecule, Equatorin, Is Required for Successful Sperm-Oocyte Fusion in Mice1. *Biology of Reproduction*, *59*(1), 22–29.

Trigg, N. A., Eamens, A. E. L., & Nixon, B. (2019). The contribution of epididymosomes to the sperm small RNA profile. *Reproduction (Cambridge, England)*, *157*(6), R209–R223.

Wang, H., Wan, H., Li, X., Liu, W., Chen, Q., Wang, Y., Yang, L., Tang, H., Zhang, X., Duan, E., Zhao, X., Gao, F., & Li, W. (2014). Atg7 is required for acrosome biogenesis during spermatogenesis in mice. *Cell Research 2014 24:7*, *24*(7), 852–869.

Westbrook-Case, V. A., Winfrey, V. P., & Olson, G. E. (1995). Sorting of the domain-specific acrosomal matrix protein AM50 during spermiogenesis in the guinea pig. *Developmental Biology*, *167*(1), 338–349.

Xiao, N., Kam, C., Shen, C., Jin, W., Wang, J., Kwong, M. L., Jiang, L., & Xia, J. (2009). PICK1 deficiency causes male infertility in mice by disrupting acrosome formation. *Journal of Clinical Investigation*, *119*(4), 802–812.

Xu, J., Yao, G., Ru, Y., & Xie, S. (2017). Expression of tamoxifen-inducible CRE recombinase in Lcn5-CreERT2 transgenic mouse caput epididymis. *Molecular Reproduction and Development*, *84*(3), 257–264.

Yao, R., Ito, C., Natsume, Y., Sugitani, Y., Yamanaka, H., Kuretake, S., Yanagida, K., Sato, A., Toshimori, K., & Noda, T. (2002). Lack of acrosome formation in mice lacking a Golgi protein, GOPC. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11211–11216.

Yi, Y.-J., Zimmerman, S. W., Manandhar, G., Odhiambo, J. F., Kennedy, C., Joná, V., Postlerová, P., Sutovsky, M., Park, C.-S., & Sutovsky, P. (2011). Ubiquitin-activating enzyme (UBA1) is required for sperm capacitation, acrosomal exocytosis and sperm-egg coat penetration during porcine fertilization. *International Journal of Andrology*, *35*, 196–210.

Yoshida, K., Ito, C., Yamatoya, K., Maekawa, M., Toyama, Y., Suzuki-Toyota, F., & Toshimori, K. (2010). A model of the acrosome reaction progression via the acrosomal membrane-anchored protein equatorin. *Reproduction*, *139*(3), 533–544.

Zakrzewski, P., Redowicz, M. J., Buss, F., & Lenartowska, M. (2020). Loss of myosin VI expression affects acrosome/acroplaxome complex morphology during mouse spermiogenesis. *Biology of Reproduction*, *103*(3), 521–533.

Zhao, Y., Diao, H., Ni, Z., Hu, S., Yu, H., & Zhang, Y. (2011). The epididymis-specific antimicrobial peptide β-defensin 15 is required for sperm motility and male fertility in the rat (Rattus norvegicus). *Cellular and Molecular Life Sciences*, *68*(4), 697–708.

Zhou, Q., Nie, R., Li, Y., Friel, P., Mitchell, D., Hess, R. A., Small, C., & Griswold, M. D. (2008). Expression of stimulated by retinoic acid gene 8 (Stra8) in Spermatogenic Cells Induced by retinoic acid: An in vivo study in vitamin A-sufficient postnatal murine testes. *Biology of Reproduction*, *79*(1), 35–42.

Zhou, Y., Zheng, M., Shi, Q., Zhang, L., Zhen, W., Chen, W., & Zhang, Y. (2008). An Epididymis-Specific Secretory Protein HongrES1 Critically Regulates Sperm Capacitation and Male Fertility. *PLoS ONE*, *3*(12).

Zigo, M., Manaskova-Postlerova, P., Jonakova, V., Kerns, K., & Sutovsky, P. (2019). Compartmentalization of the proteasome-interacting proteins during sperm capacitation. *Scientific Reports*, *9*(1), 1–18.

8 Supplementary material

Gene symbol (UniprotKB)	Protein name	Biological process	Additional information
Adss (P46664)	Adenylosuccinate synthethase	AMP biosynthetic process, purine metabolism	
Ppp2r1a (Q76MZ3)	Protein phosphatase 2, regulatory subunit A, alpha	Chromosome and chromatid segregation	
Pgk2 (P09041)	Phosphoglycerate kinase 2	Glycolytic process, sperm motility	Essential for sperm fertility and sperm motility
Ptgds (O09114)	Prostaglandin-H2 D-isomerase	Prostaglandin biosynthetic process, negative regulation of male germ cell proliferation, mast cell degranulation	Catalyse the conversion of Prostaglandin H_2 to Prostaglandin D_2
Psmb7 (P70195)	Proteasome subunit, beta type 7	Proteolysis, proteasomal protein catabolic process	Component of the 20S core proteasome complex involved in the proteolytic degradation
Cul3 (Q9JLV5)	Cullin 3	Ubiquitin dependent protein catabolic process, vesicle-mediated transport, positive regulation of mitotic metaphase/anaphase transition	Core component of multiple cullin-RING-based E3 ubiquitin-protein ligase complexes. Together with ARIH1 collaborate in tandem to mediate ubiquitination of target proteins.
Ca2 (P00920)	Carbonic anhydrase 2	Regulation of intracellular pH	
Prep (Q9QUR6)	Prolyl endopeptidase	Proteolysis	
Ltf (P08071)	Lactotransferrin	Antimicrobial response, Iron ion transport	
Pebp4 (Q9D9G2)	Phosphatidylethanolamine binding protein 4	Any available information	
Lcn5 (A2AJB7)	Lipocalin 5	Retinoic acid metabolic process	Binds all trans and 13-cis retinoic acid. Might act as a retinoid carrier protein which is required for epididymal function and/or sperm maturation.
Hspa4 (Q61316)	Heat shock protein 4	Stress response: Chaperone-mediated protein complex assembly	.
Park7/ /DJ-1 (Q99LX0)	Parkinson disease 7	Response to oxidative stress, Fertilization	Has an important role in cell in protection against oxidative stress and normal male fertility.
Hk1 (P17710)	Hexokinase 1	Glycolytic process, inflammatory response	Catalyses the phosphorylation of various hexoses.
Otub1 (Q7TQI3)	Ubiquitin thioesterase OTUB1	Protein deubiquitination, DNA repair	

Arih1 (Q9Z1K5)	E3 ubiquitin-protein ligase ARIH1	Ubiquitin dependent protein catabolic process.	E3 ubiquitin-protein ligase which works together with cullin-RING ubiquitin ligase (CRL) complexes.
Ap1g1 (P22892)	Adaptor protein complex AP-1, subunit gamma-1	Vesicle-mediated transport	Involved in the trafficking of lysosomal enzyme cathepsin D between the trans-Golgi network (TGN) and endosomes.
Tkt (P40142)	Transketolase	Penthose-phosphate shunt	

Table 1: Table of annotated important acrosomal proteins according to the UniProt database. The list of 18 proteins were identified with Random Forest analysis were manually annotated using UniProt database. First two columns contain the gene and protein name with UniProt unique identifier. In the third column is described biological process are they involved in and relevant additional information is mentioned in fourth column.

Gene symbol (UniprotKB)	**Protein name**	**Biological process**	**Additional information**
Spink2 (Q8BMY7)	Serin protease inhibitor Kazal-type 2	Inhibition of trypsin proteases during acrosome assembly	Inhibitor of acrosine and other serin type endopeptidase activity, probably hinders premature activation of proacrosin.
Ube2v2 (Q9D2M8)	Ubiquitin conjugating enzyme E2 variant 2	Postreplication DNA repair, ubiquitin dependent protein catabolic process	Together with UBE2N catalyse Lys-63 chain, that does not lead to the proteasomal degradation
Tekt (Q6X6Z7)	Tektin	Sperm flagellum motility	Essential for sperm fertility and sperm motility
Ndufa2 (Q9CQ75)	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 2	Oxidative phosphorylation	
Cox7a2 (P488771)	Cytochrome C oxidase subunit 7A2, mitochondrial	Oxidative phosphorylation	
Lpp (Q8BFW7)	Lipoma-preferred partner homolog	Cell-cell adhesion	May play a structural role at sites of cell adhesion in maintaining cell shape and motility.
Tgm4 (Q8BZH1)	Protein-glutamine gamma-glutamyltransferase 4	Mating plug formation, peptide cross-linking	Has a role in the formation of the seminal coagulum through the cross-linking of specific proteins in the seminal plasma
C9orf135 (Q9CQC3)	Protein C9orf135 homolog	Any available information	Highly expressed in the testis.
Acrv1 (P50289)	Acrosomal protein SP-10	Serine protease	
Ribc1 (Q9D0B8)	RIB43a-like with coiled-coil proteins 1	Any available information	Expressed in testis. Belongs to the RIB43A family, with proposed function in the cilia basal bodies and protofilament ribbons.
Ndufas1 (Q91VD9)	NADH-ubiquinone oxidoreductase subunit, mitochondrial	Oxidative phosphorylation	
Dynll2 (QD0M5)	Dynein light chain 2	Cilium assembly	.

Gk2 (Q9WU65)	Glycerol kinase 2	Glycerol degradation	
Cox6c (Q9CPQ1)	Cytochrome C oxidase subunit 6C	Oxidative phosphorylation	
Izumo4 (D3Z690)	Izumo sperm-egg fusion protein	Fertilization	
Dld (O08749)	Dihydrolipoyl dehydrogenase, mitochondrial	Sperm capacitation	Involved in the hyperactivation of spermatozoa during capacitation

Table 2: Table of annotated important sperm cells proteins according to the UniProt database. The list of 16 proteins identified with Random Forest analysis were manually annotated using UniProt database. As in Fig. 3, the first two columns contain gene and protein names with UniProt unique identifier. In the third column is described biological process they are involved in and relevant additional relevant information is mentioned in the fourth column.