

Epigenetic information carriers in human sperm cells

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Darja Elzer

geb. Am 23.06.1994 in Karaganda, Kasachstan

Mainz, 2023

Dean: Prof. Dr. Eckhard Thines

First rapporteur: Prof. Dr. Hans Zischler

Second rapporteur: Prof. Dr. Miguel Andrade

Date of defence: 21.12.2023

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich diese Dissertation selbständig angefertigt und alle verwendeten Hilfsmittel in der Arbeit angegeben habe. Ich habe die Dissertation, oder Teile davon, an keiner anderen Fakultät, bzw. einem anderen Fachbereich eingereicht. Ich habe keinen anderen Promotionsversuch unternommen.

Mainz, Oktober 2023

Darja Elzer

Acknowledgements

Abstract

The idea of inheritance of acquired characteristics is associated with early evolutionary theory being superseded by the modern synthesis combining Darwinian theories and Mendelian genetics. Whereas there is consensus on the possible consequences of the environment triggering an epigenetic (health) status of an individual, much debate centers around a possible inheritance of acquired characters via epigenetic mechanisms. Generally accepted epigenetic mechanisms include DNA methylation, histone modifications, and non-coding RNAs. During embryogenesis, epigenetic marks are erased and chromatin is remodelled. Our knowledge on epigenetic information carriers or mechanisms, including small non-coding (snc)RNAs and residual histone and protamine (modifications) is patchy, especially in the context of its evolutionary potential. My working hypothesis is that an erasure of epigenetic marks during germline reprogramming might not be fully efficient, thus leading to the intergenerational inheritance of epialleles. Therefore, the complete non-coding (nc)RNA cargo in human sperm heads was described, mainly focussing on one group of sncRNA, namely Y RNAs and their fragments (YsRNA). I found that specific Y RNA fragments, namely Ys1RNA and Ys4RNA are selectively enriched in sperm heads and almost absent in oocytes, suggesting a potential intergenerational epigenetic transfer of YsRNA as paternal contribution to the zygote. In human mature spermatozoa, about 85 % of histones are replaced by protamines. Hence, a subset of genomic loci may escape complete removal of their histone packaging proteins, with epigenetic effects of the respective gene regulation. Moreover, not only residual histones but protamines themselves could have an impact in epigenetic regulatory processes. Therefore, I ChIP-analysed the sperm head chromatin for residual histones, represented by H2B, and protamine chromatin (PRM1 and PRM2). I aimed to correlate the ncRNA expression with the chromatin data and scrutinize the possible role of the chromatin components in the transgenerational inheritance of epigenetic information. I found that H2B, PRM1 and PRM2 peaks are preferably found in satellite regions, mainly in centromeric and telomeric regions, and that those peaks overlap to great parts between the three proteins. H2B and protamine peaks overlap with lncRNA genomic regions suggesting the at least partially possible transcription of lncRNAs in human sperm heads. Histone H2B and the protamines PRM1 and PRM2 contain intrinsically disordered regions, which bind preferably lncRNA in an unspecific manner. Human sperm heads thus contain specific RNA (Y RNA), in the sense of sequence-specific function and “unspecific” RNA cargos (lncRNA) that are transmitted as epigenetic information carriers into the zygote possessing potential adaptive value.

Zusammenfassung

Die Idee der Vererbung erworbener Merkmale steht im Zusammenhang mit der frühen Evolutionstheorie, die von der modernen Synthese aus Darwin'schen Theorien und Mendelscher Genetik abgelöst wurde. Während über die möglichen Auswirkungen der Umwelt auf den epigenetischen (Gesundheits-)Status eines Individuums Einigkeit besteht, dreht sich vieles um die mögliche Vererbung erworbener Merkmale über epigenetische Mechanismen. Zu den allgemein anerkannten epigenetischen Mechanismen gehören DNA-Methylierung, Histon-Modifikationen und nicht-kodierende RNAs. Während der Embryogenese werden epigenetische Markierungen gelöscht und das Chromatin wird umgestaltet. Unser Wissen über epigenetische Informationsträger oder -mechanismen, einschließlich kleiner nichtkodierender RNAs (snc)RNAs und restlicher Histon- und Protaminmodifikationen, ist lückenhaft, insbesondere im Zusammenhang mit ihrem evolutionären Potenzial. Meine Hypothese ist, dass die Löschung epigenetischer Markierungen während der Keimbahnprogrammierung möglicherweise nicht vollständig effizient ist, was zur intergenerationalen Vererbung von Epiallelen führt. Daher wurde die gesamte nicht-kodierende (nc)RNA-Fracht in menschlichen Spermienköpfen beschrieben, wobei der Schwerpunkt auf einer Gruppe von sncRNAs lag, nämlich den Y RNAs und ihren Fragmenten (YsRNA). Ich fand heraus, dass bestimmte Y RNA-Fragmente, nämlich Ys1RNA und Ys4RNA, selektiv in Spermienköpfen angereichert sind und in Oozyten fast gar nicht vorkommen, was auf einen potenziellen intergenerationalen epigenetischen Transfer von YsRNA als väterlichen Beitrag zur Zygote hindeutet. In reifen menschlichen Spermien werden etwa 85 % der Histone durch Protamine ersetzt. Daher ist es möglich, dass ein Teil der genomischen Loci der vollständigen Entfernung ihrer Histonproteine entgehen, was epigenetische Auswirkungen auf die jeweilige Genregulation hat. Darüber hinaus könnten nicht nur die verbleibenden Histone, sondern auch die Protamine selbst einen Einfluss auf die epigenetischen Regulationsprozesse haben. Daher habe ich das Kopfchromatin der Spermien auf residuale Histone, repräsentiert durch H2B, und Protaminchromatin (PRM1 und PRM2) mittels ChIP-Analyse untersucht. Mein Ziel war es, die ncRNA-Expression mit den Chromatindaten zu korrelieren und die mögliche Rolle der Chromatinkomponenten bei der transgenerationalen Vererbung epigenetischer Informationen zu untersuchen. Ich fand heraus, dass H2B-, PRM1- und PRM2-Peaks vorzugsweise in Satellitenregionen zu finden sind, hauptsächlich in Zentromer- und Telomerregionen, und dass sich diese Peaks zum Großteil zwischen den drei Proteinen überschneiden. H2B- und Protamin-Peaks überschneiden sich mit lncRNA-Genomregionen,

was auf eine zumindest teilweise mögliche Transkription von lncRNAs in menschlichen Spermienköpfen hindeutet. Histon H2B und die Protamine PRM1 und PRM2 enthalten intrinsisch ungeordnete Regionen, die vorzugsweise lncRNA auf unspezifische Weise binden. Menschliche Spermienköpfe enthalten also spezifische RNA (Y RNA) im Sinne einer sequenzspezifischen Funktion und "unspezifische" RNA-Ladungen (lncRNA), die als epigenetische Informationsträger in die Zygote übertragen werden und einen potenziellen adaptiven Wert besitzen.

Table of content

1 INTRODUCTION	1
AIM OF THIS STUDY	6
2 RESULTS	11
2.1 HUMAN SPERM HEADS HARBOUR MODIFIED YsRNA AS TRANSGENERATIONALLY INHERITED NON-CODING RNAs	11
2.1.1 REVIEWERS COMMENTS	42
2.2 EPIGENETIC ROLE OF SPERM HEAD CHROMATIN AND RELEVANCE FOR INHERITANCE OF EPIGENETIC INFORMATION CARRIERS	43
2.2.1 <i>Material & Methods</i>	51
2.2.2 <i>Results</i>	60
2.2.3 <i>Discussion</i>	85
3 GENERAL DISCUSSION	95
4 APPENDIX	113
4.1 FIGURES	113
4.2 TABLES	116

List of figures

Introduction

1 Graphical abstract about the spermatogenesis

Aim of this study

1 Microscopic image of Sperm cells before and after differential lysis

2 PCR results of sperm heads, oral mucosa and urothelial cells using mtDNA specific primer

3 Illustration of the human Y RNA structures and their functional sites

Results

Chapter 1

1 BLASTN results querying smallRNA with Y RNA as database

2 Number of YsRNA hits in sperm heads, testis and oocytes before and after oxidation

3 Number of YsRNA hits in sperm heads without treatment, treatment with PNK, RppH and both enzymes

4 YsRNA hits shown for sperm heads and exosomes in reads per million mappable reads for the four Y RNA homologs divided in a 4' half and a 3' half

5 Profile of YsRNA in epididymis divided in caput, corpus and cauda region in reads per million alignable reads and depicted for the different Y RNA homologs

6 Illustration of the total sperm transcriptome-lincRNA profile in red and the Ro60-bound fraction in blue in reads per million alignable reads

Chapter 2

1 UNITAS annotation of testis and sperm head smallRNA

2 Illustration of the 100 most abundant genomic features in testis and sperm heads

- 3 Gene Ontology annotation of the top 100 transcripts in testis and sperm heads
- 4 Amount of PRM1 and PRM2 mRNA in different cell types during spermatogenesis
- 5 Dot blot results using the antibodies against PRM1 and PRM2
- 6 Western blot result using the antibody against H2B
- 7 Idiogram of Individual 1 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38
- 8 Idiogram of Individual 2 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38
- 9 Idiogram of Individual 3 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38
- 10 Idiogram of Individual 4 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38
- 11 HOMER annotation results for H2B, PRM1 and PRM2 peaks of Individual 3
- 12 HOMER peak annotation for H2B for all four donors
- 13 HOMER peak annotation for PRM1 for all four donors
- 14 HOMER peak annotation for PRM2 for all four donors
- 15 Venn diagrams of lncRNA overlaps between all four donors for H2B, PRM1 and PRM2
- 16 Venn diagram of the total lncRNA overlaps between H2B, PRM1 and PRM2

General discussion

- 1 Graphical abstract about the intergenerational epigenetic transfer of Ys1RNA as paternal contribution to the zygote

List of tables

Results

Chapter 2

1 Number of PRM1 and PRM2 peaks and overlapping peaks for all four donors

List of abbreviations

μg	Microgram
μl	Microliter
3' UTR	3' untranslated region
5' UTR	5' untranslated region
ALR	alpha satellite, α-satellite
APS	Ammonium persulfate
bp	base pair
BR	Broad Range
BSA	Bovine serum albumin
BSR	beta satellite, β-satellite
CCDS	Consensus CDS protein set
CENP-A	Centromere protein A
cenRNA	Centromeric RNA
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
CSPD	Disodium 3-(4-methoxy Spiro {1,2-dioxetane-3,2'-(5'-chloro)tricyclo [3.3.1.1 ^{3,7}]decan }-4-yl)phenyl phosphate
CUT&RUN	Cleavage Under Targets & Release Using Nuclease
ddH ₂ O	double-distilled water
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate

dUTP	Deoxyuridine, triphosphate
e.g.	exempli gratia
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FriP	Fraction of Reads in Peaks
G	G-force
Gb	Gigabase
GO	Gene Ontology
GSATII	gamma satellite II
gtf	Gene transfer format
H2B	Histone H2B
H ₂ O ₂	Hydrogen peroxide
H ₃ BO ₃	Boric acid
HCl	Hydrochloric acid
HENMT1	HEN Methyltransferase 1
HOMER	Hypergeometric Optimization of Motif EnRichment
HPLC	High Performance Liquid Chromatography
HSATI	Human satellite 1
i.e.	id est
ICSI	Intracytoplasmic sperm injection
IgG	Immunoglobulin G
kb	Kilobase
KCl	Potassium chloride
kDa	Kilodalton
KH ₂ PO ₄	Monopotassium phosphate

LiCl	Lithium chloride
lincRNA	long intergenic non-coding RNA
LINE	Long interspersed nuclear element
LNA	locked nucleic acids
lncRNA	long non-coding RNA
LTR	Long terminal repeat
M	Molar
Mb	Megabase
MERVL	Murine endogenous retrovirus-L
miRNA	microRNA
miscRNA	miscellaneous RNA
mM	Millimolar
mRNA	messenger RNA
mtDNA	mitochondrial DNA
$\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$	Sodium borate
Na_2HPO_4	Disodium phosphate
NaAc	Sodium acetate
NaCl	Sodium chloride
NaHCO_3	Sodium bicarbonate
NaIO_4	Sodium periodate
NaIO_4	Sodium periodate
NaOH	Sodium hydroxide
ncRNA	non-coding RNA
NGS	Next generation sequencing
nt	Nucleotide

PBS	Phosphate-buffered saline
PBST	PBS with Tween-20
PCR	Polymerase chain reaction
PE	Paired end
piRNA	PIWI-interacting RNA
PNK	T4-Polynucleotide Kinase
Pol I	Polymerase I
Pol II	Polymerase II
Pol III	Polymerase III
pri-miRNA	primary miRNA
PRM1	Protamine 1
PRM2	Protamine 2
pseudo	pseudo gene
PVDF	Polyvinylidene fluoride
RC	Rolling Circle
rDNA	ribosomal DNA
RIP	RNA immunoprecipitation
RIPA	Radioimmunoprecipitation assay
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
rpm	Reads per million
RppH	RNA 5' Pyrophosphohydrolase
rRNA	ribosomal RNA
RTG	Research training group
scRNA	Small conditional RNA

SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SE	Single end
SEM	Standard error of the mean
SINE	Short interspersed nuclear element
SmallRNA-Seq	Small RNA Sequencing
sncRNA	small non-coding RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SRA	Sequence Read Archive
SRP	Signal recognition particle
SSC	Spermatogonial stem cell
TAR1	Telomere-Associated Repeat
TEMED	Tetramethylethylenediamine
TEN	TRIS, EDTA, NaCl
TNP1	Transition Protein 1
tRF	tRNA-derived fragment
TRIS	Tris(hydroxymethyl)aminomethane
TRIzol	brand name of <u>guanidinium thiocyanate</u>
tRNA	transfer RNA
tsRNAs	tRNA fragments
TSS	Transcription start site
TTS	Transcription termination site
USER	Uracil-Specific Excision Reagent
ZGA	Zygotic genome activation

1 Introduction

The idea of inheritance of acquired characteristics is associated with early evolutionary theory, which was superseded by the modern synthesis combining Darwinian theories and genetics shortly after rediscovering Mendelian genetics. Darwin assumed that all cells of an organism emit small particles ("gemmules") that migrate into the germinal cells and are passed on from parents to offspring. These particles can be altered by the environment and thus influence the offspring (Liu, 2008). In this context, Lamarck had a similar idea of heredity. He was convinced that acquired characteristics are passed on to the descendants through use or disuse. Both were sure that the environment has a great influence on an organism and its evolution (Burkhardt, 2013). Today, this idea still holds truth, and it is generally accepted that the environment, and not only the DNA, has an influence on the phenotype and evolution of an organism. This concept is referred to as epigenetics.

Epigenetics describes alterations in gene expression without changing the DNA sequence itself (Weinhold, 2006). These alterations can be transmitted to the daughter cells and, when happening in the germline, even to their offspring. Many different environmental factors, including chemical pollutants, affect how much or little and which regions of our genome are expressed. It is known that changes in gene expression due to epigenetic changes can lead to cancer, autoimmune diseases, cardiovascular diseases, diseases associated with the reproductive system and many more illnesses (Weinhold, 2006; L. Zhang et al., 2020). Additionally, not only these serious diseases might be caused and influenced by epigenetic marks, but even more minor diseases like drug addiction are also known to be influenced by DNA methylation (Bali et al., 2011). This shows that epigenetics has a greater impact on the health and behaviour of organisms than previously thought and should be investigated more intensively.

Over the last decades, the molecular and mechanistic correlates underlying the idea of acquired and possible heritable characters have been intensively researched. Self-sustaining feedback loops, chromatin-based mechanisms including DNA methylation and both coding and non-coding RNA as well as structural templating (reviewed by Heard & Martienssen, 2014) are generally accepted molecular mechanisms that have the potential to render environmental influence into gene expression regulation. Whereas there is consensus on the possible consequences of the environment triggering the epigenetic as well as the health status of an individual, much debate centers around a possible inheritance of acquired characters via epigenetic mechanisms. To describe the phenomenon of epigenetic inheritance

more precisely the term “transgenerational” was introduced to set apart a non-sequence-based effect that is truly transmitted from one generation to the other from parental effects like in utero exposure of an embryo and its germline to environmental influences, primarily including dietary interventions, stress exposures, and toxin exposures. Effects caused by the latter cannot be found in generations that are not exposed to the triggering environmental signal. Initially formulated evolutionary theories that take account of two systems of inheritance, namely genetic and epigenetic inheritance, are impeded by the fact that the majority of environmentally induced epigenetic states tested to date do not persist over long timescales and are therefore considered as unlikely to play roles in macroevolutionary processes (Bošković & Rando, 2018). Though epigenetic inheritance is thought to be relatively common in plants and usually explained by a poorly defined plant germline (Heard & Martienssen, 2014) the widespread cytosine methylation changes observed in response to hyperosmotic stress in *Arabidopsis* or phosphate deprivation in rice are transient, being erased either immediately after returning plants to control conditions or in F1 offspring (summarized in Bošković & Rando, 2018). Compared to plants, epialleles in metazoan systems are even far less robustly inherited due to the process of germline reprogramming that removes and reestablishes epigenetic signatures. Consequently, a recently published summary (Perez & Lehner, 2019) lists only weak transgenerational effects detectable in only few Metazoan model species, including *Drosophila*, *C. elegans* and rodents.

Germline cells are crucial for reproduction and the transmission of genetic and epigenetic information from parent to offspring. In the female body those are the oocytes and in the male body the sperm cells (Cheng et al., 2022). August Weismann formulated the concept of germ plasm in the late 19th century. Germ plasm describes the sum of all hereditary particles and substances of an organism which are only found in the nuclei of germ cells. In contrast, all other cells only contain parts of these determinants (Hargitt, 1944). Therefore, germ cells are the only cells that can transmit heritable information and other cells do not take part in heredity. Weismann stated that somatic cells do not have an impact on the germ cells. This concept is called the “Weismann barrier” (Nilsson et al., 2020). With his idea of heredity, he rejected Lamarck’s theory regarding inheritance of acquired characteristics (Bline et al., 2020). Today we know that the Weismann barrier does not hold absolute truth and that somatic cells do have an impact on germ cells.

The process describing the maturation of undifferentiated diploid spermatogonia via meiotic processes first into tetraploid spermatids and then into highly differentiated haploid

spermatozoa is a highly complex one and called spermatogenesis (Hess & De Franca, 2008). Spermatogenesis is crucial for the continuity of the male germline by producing germ cells that have the potential to fertilise oocytes and eventually produce offspring (Kubota & Brinster, 2018). Functionally, spermatogenesis can be divided into three successive stages: spermatocytogenesis, meiosis and spermiogenesis (Johnson et al., 2000). Sperm production begins with spermatogonial stem cells (SSCs), which correspond to type A_{dark} spermatogonia. Type A_{dark} spermatogonia are seen as reserved stem cells that do not directly participate in producing sperm. They divide mitotically in a new type A_{dark} and a type A_{pale} spermatogonia. This type A_{pale} spermatogonia is considered the renewing stem cell and divides mitotically into a new type A_{pale} spermatogonia and a type B spermatogonia (Dym et al., 2009). After several rounds of mitosis, type B spermatogonia divides into pre-leptotene spermatocytes, which undergo the last mitotic division before the meiotic phase in the S phase (Mäkelä & Toppari, 2017). During the prophase of the first meiotic division, important processes like homologous chromosome alignment, meiotic recombination and crossover formation occur (Jan et al., 2017). The prophase is subdivided into five stages: leptotene, zygotene, pachytene, diplotene and diakinesis. In the leptotene stage the chromatin condensates and appears as long strands. During zygotene stage a synapsis between homologous chromosomes is formed. In the pachytene stage homologous recombination, including crossing over, takes place (Marques et al., 2018). The diplotene stage is characterised by chiasmata, the points of contact between chromatids of homologous chromosomes. The diakinesis is the last stage, representing the end of the prophase, and is characterised by the dissolution of the nuclear membrane (Page & Hawley, 2003). Following the prophase, the first meiotic division takes place. During this phase the primary spermatocyte is divided into two secondary spermatocytes, which contain all duplicated autosomal chromosomes and either a duplicated X or Y chromosome. After the first meiotic division, the second meiotic division starts. Here, each secondary spermatocyte divides into two spermatids, each with a haploid set of chromosomes (Kotaja, 2013). The round spermatids, containing spherical nuclei, differentiate into elongating spermatids and then into mature spermatozoa (Jan et al., 2017; Johnson et al., 2000). This maturation process is called spermiogenesis or spermiohistogenesis (Hofgärtner et al., 1979; Oliva & Castillo, 2011). The volume of the spermatid nuclei decreases, and they transform into a more elongated shape. Hereby, epigenetic changes, in particular chromatin remodelling, take place. Histones are replaced by protamines, resulting in highly compacted DNA. Additionally, most of the cytoplasm is removed and a flagellum and the acrosomal vesicle

are formed. On its way to the epididymis the spermatozoa undergo further maturation processes, where the chromatin gets even more compacted (Oliva & Castillo, 2011). In general, the process of spermatogenesis comprises a total duration of about 64 days (Heller & Clermont, 1963; Misell et al., 2006). Once in the epididymis, the spermatozoa undergo further maturation processes. While moving from the proximal (caput) to the distal (cauda) end of the epididymis spermatozoa become motile and able to fertilise oocytes (Cornwall & Von Horsten, 2007).

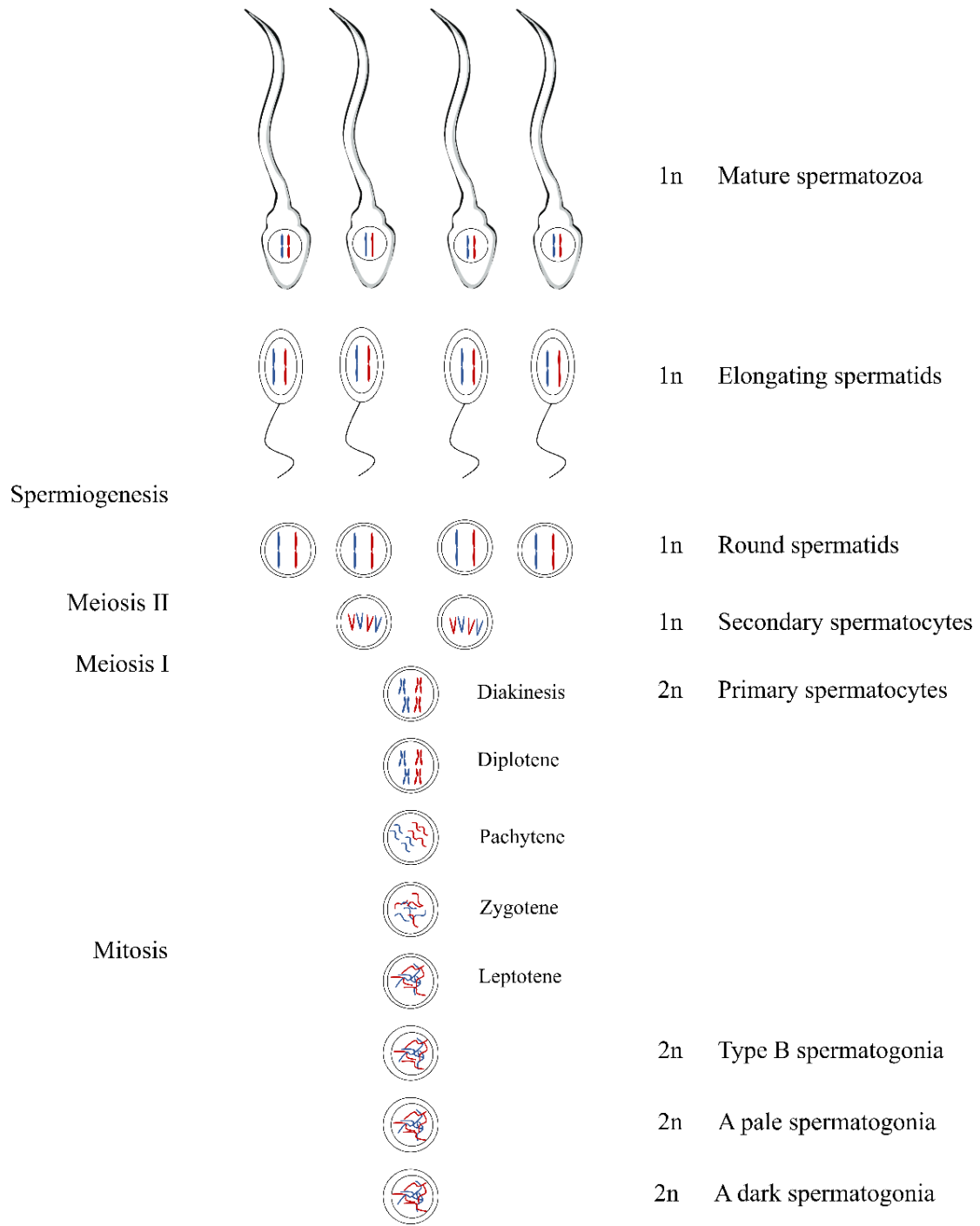


Figure 1: Graphical abstract about the spermatogenesis.

Aim of this study

As this work is part of the RGT GenEvo-project, the main focus lies on transgenerational inheritance of epigenetic marks, and ultimately on how epigenetic mechanisms can become heritable and what role they play in evolution. Generally accepted epigenetic mechanisms include DNA methylation, histone modifications and non-coding RNA. The germline undergoes natural epigenetic reprogramming during embryonic development (Hajkova, 2011). During this process epigenetic marks are removed and reset, converting germ cells into stem cells. Without this reprogramming, germ cells would retain the parental epigenetic memory, which would then prevent the transmission of genetic information to the offspring (Sabour & Schöler, 2012). Moreover, during embryogenesis epigenetic marks are erased and chromatin is remodelled (Hajkova, 2011). My working hypothesis is that an erasure of epigenetic marks during germline reprogramming might not be fully efficient, thus leading to the intergenerational inheritance of epialleles.

Specifically, I aim to contribute to the questions on the presence of epigenetic information carriers in mature human sperm heads and thus the zygotic male pronucleus. The epigenetic states from nuclei of fully differentiated sperm are going to be determined to avoid confounding results caused by non-spermatid cells and extranuclear contaminants (such as theca proteins). After impregnation the first zygotic replication is carried out in separate paternal and maternal pronuclei; thus, it can be assumed that the epigenetic status of the paternal pronucleus is best reflected in sperm heads. Upon fertilization, e.g. unusual chromatin architectures are extensively remodelled and maternal and paternal genomes remain distinctively packaged for several cell divisions. To isolate sperm heads from ejaculates a differential lysis protocol was applied – controllable by the presence/absence of mitochondrial DNA (mtDNA) – to separate sperm heads from ejaculate cells. To give an impression of how the differential lysis protocol enriches sperm heads, a microscopic snapshot of human sperm before and after differential lysis is depicted below.



Figure 2: Microscopic image of sperm cells before (left) and after (right) differential lysis. Sperm head, acrosome and tail are clearly visible in sperm cells before differential lysis. After lysis sperm heads are visible almost exclusively.

The picture on the left shows the sperm cells before the differential lysis. The sperm heads, midpiece containing the mitochondria and the flagellum can be seen clearly. On the right site the sperm cells after the differential lysis are shown, with the flagella and midpieces almost perfectly removed leaving intact sperm heads. This differential lysis allows a removal of the mitochondria which can be assessed in its efficiency by simple mtDNA-PCRs. To compare different sources of DNA sperm heads, oral mucosa and urothelial cells were obtained from one individual and PCR checked with mtDNA-specific primers:

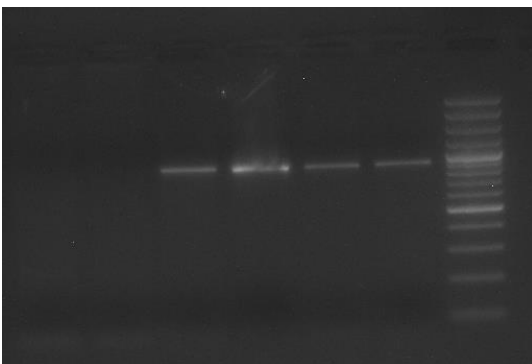


Figure 3: PCR results of sperm heads (lane 1 and 2), oral mucosa (lane 3) and urothelial cells (lanes 4, 5 and 6) using mtDNA specific primer. Lanes 3 to 6 show lines while the lanes 1 and 2 show no lines. A 100 bp Plus marker was used as ladder.

The agarose gel shows the result of the PCR using mtDNA-specific primer. The first two lanes from left to right were loaded with sperm head DNA, the third lane was loaded with oral mucosa DNA and the remaining three lanes were loaded with urothelial cell DNA. The used ladder was a 100 bp Plus marker. Lanes 1 and 2 show now lines, while the other four lines show clear lines. This result shows that there is no mitochondrial DNA in the sperm head samples, which means that the differential lysis is a suitable method to purify sperm heads from all other components.

In metazoans, epigenetic erasure and resetting in gametogenesis and early embryogenesis is probably best studied for mammalian imprinting (allele-specific DNA methylation), our knowledge on “reprogramming” other epigenetic information carriers or mechanisms, including small non-coding RNAs (sncRNAs) and residual histone and protamine (modifications), lags behind especially in the context of its evolutionary potential. Therefore, the complete non-coding RNA cargo in human sperm heads will be described, mainly focussing on one group of sncRNA in Chapter 2.1, namely Y RNAs and their fragments (YsRNA).

Y RNA was first described in 1981 when this group of RNA was found in patients affected by systematic lupus erythematosus (SLE) and Sjögren’s syndrome (Hendrick, 1981; Lerner et al., 1981). Because they were found in the cytoplasm of cells, they were given the prefix ‘Y’ (Lerner et al., 1981). There are four Y RNAs in humans, hY1 with a size of 112 nt, hY3 with a size of 101 nt, hY4 with 93 nt in size and hY5 with a size of 83 nt (Gulia et al., 2020) which are clustered on one single locus on chromosome 7q36 (R. Maraia et al., 1996; R. J. Maraia et al., 1994). The 5’ end and 3’ end typically bind together to form a double-stranded stem domain divided in a lower and upper stem domain (Kowalski & Krude, 2015). The upper stem domain is important for chromosomal DNA replication and the lower stem domain has a Ro60 binding site and is, when bound to Ro60, involved in RNA stability and stress response (Kowalski & Krude, 2015). Y RNA, like many other groups of RNA including tRNA, 5S rRNA and Alu (Nikitina & Tishchenko, 2005), are transcribed by RNA polymerase III (Kowalski & Krude, 2015). The transportation from Y RNAs from the nucleus to the cytoplasm uses the same proteins as it does for miRNAs, namely Exportin 5 and Ran GTPase (Bartel, 2004). Y RNAs do not only exist in their full length but also in fragments between 25 nt and 35 nt (Röther & Meister, 2011). Y RNAs and Y RNA fragments are reported to be possible tumor biomarkers and play a significant role in several tumor types (Guglas et al., 2020). The fragmentation of Y RNAs, which is performed by RNase L (Donovan et al., 2017), is increasing in apoptotic cells and upon activation of the innate immune system (Nicolas et al., 2012; Rutjes et al., 1999) but was also found in non-apoptotic proliferating cells (Nicolas et al., 2012) and within extracellular vesicles (Driedonks & Nolte-T’Hoen, 2019). Y RNA fragments in apoptotic cells are a result of Y RNA degradation performed by caspase-3 (Guglas et al., 2020).

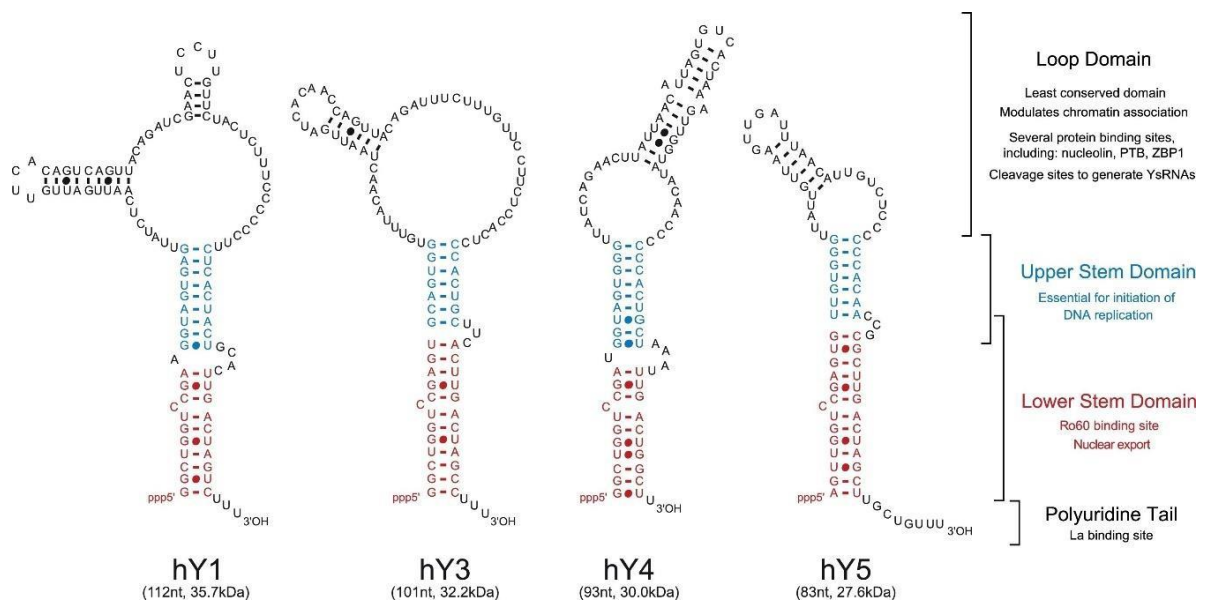


Figure 4: Illustration of the human Y RNA structures and their functional sites (Kowalski & Krude, 2015).

Somatic cells and gametes are characterized by a pronounced contrasting chromatin packaging. Histone modifications also exhibit germline-specific patterns (Bao & Bedford, 2016). Chromatin architecture in oocytes differs substantially from that in somatic cells and is characterized by e.g., the presence of histone variants and specific histone modifications in mammals. Even more pronounced differences are realized in sperm cells, though there is great variety across plants and metazoans. During mammalian spermatogenesis the presence of highly specialized testes histones is required. Most histones are replaced first by transition proteins and thereafter by two protamines, PRM1 and PRM2. In human sperm, protamines are phosphorylated at specific regions (PRM1S9, PRM1S11, and PRM2S59) but the purpose behind this process remains yet unknown. Protamines are basic, arginine-rich proteins that differ from histones in their enrichment in lysine and cysteine residues. The high arginine content results in a partial positive charge that promotes DNA binding and neutralises the negative charge along the phosphodiester backbone of DNA, allowing neighbouring DNA molecules to be packed more tightly, resulting in a 10 times more compacted genome in the male nucleoprotamine complex. This change causes the typical hydrodynamic shape and likely protects the genome from physical and chemical damage. Furthermore, a classical interpretation states that protamine chromatin is transcriptionally silenced. In mature spermatozoa, about 99 % of histones are replaced by protamines in mice, and about 85 % in humans (Brunner et al., 2014). Hence, a subset of genomic loci may escape complete removal of their histone packaging proteins, with epigenetic effects on the respective gene

regulation. Moreover, not only residual histones but protamines themselves could have an impact in epigenetic processes. My second overarching goal is therefore to ChIP-analyse the sperm head chromatin for residual histones, represented by H2B and protamine chromatin (PRM1 and PRM2). I aim to correlate the non-coding RNA expression with the chromatin data and scrutinize the possible role of the chromatin components in the transgenerational inheritance of epigenetic information.

2 Results

The results presented in Chapter 2.1 were submitted as an Original Research paper to Frontiers in genetics – RNA.

2.1 Human sperm heads harbour modified YsRNA as transgenerationally inherited non-coding RNAs

My contributions to this publication include the conceptualisation and design of the experiments, the (complete) laboratory work, the bioinformatic analysis, interpretation and graphical presentation of the data as well as the writing of the manuscript.

The planning of the experiments and the interpretation of data as well as the drafting of the manuscript were realised together with Prof. Dr. H. Zischler. The project was supervised by Prof. Dr. H. Zischler.

The goal of this work is to elucidate the biogenesis of YsRNA starting from Y RNA source genes and during sperm development proceeding to mature sperm heads. The main question is if and if so how YsRNAs might constitute a paternal contribution to the zygote delivered via the male pronucleus and whether this epigenetic information is inherited intergenerationally or even transgenerationally with possibly adaptive value.

Human sperm heads harbour modified YsRNA as transgenerationally inherited non-coding RNAs

Darja Elzer, Michelle Bremser & Hans Zischler^{1*}

¹Division of Anthropology, Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University Mainz, Mainz, Germany

*** Correspondence:**

Corresponding Authors

Hans Zischler, Darja Elzer

zischler@uni-mainz.de, delzer@uni-mainz.de

Keywords: human sperm heads, non-coding RNA payload, 2'-O-methylation, transgenerational epigenetic inheritance, Y RNA, small Y RNA, piRNA.

Abstract

Most epigenetic information is reprogrammed during gametogenesis and early development. However, some epigenetic information persists and can be inherited, a phenomenon that is common in plants. On the other hand, there are increasing examples of epigenetic inheritance in metazoans, especially for small non-coding RNAs. The presence of regulatory important RNAs in oocytes is undisputed, whereas the corresponding RNA payload in spermatozoa and its regulatory influence in the zygote and early embryogenesis is largely enigmatic. For humans, we herein describe small Y RNA fragments as paternal contribution to the zygote. First, we trace the biogenesis of these YsRNAs from source Y RNAs with respect to the 5' and 3' modifications. Both the length and modifications make these YsRNAs reminiscent of canonical piRNAs that are not derived from piRNA clusters. Second, from the early stages of spermatogenesis to maturation in the epididymis, we observe distinct YsRNA dynamics in the male germline. We track YsRNAs exclusively in mature sperm heads, the correlate of the male pronucleus in the zygote, suggesting an important role of the epididymis as a hub for the transfer and modification of epigenetic information related to YsRNA between soma

and germline in humans. Because this YsRNA-based epigenetic mechanism is effective across generations, we questioned whether the function of Y RNA binding to Ro60 might have adaptive value and described the profiles of non-coding RNAs bound to Ro60 in the human sperm head. Because specific binding profiles of RNA to Ro60 were detected, but no Y RNA bound to Ro60, we hypothesize that the Ro60 system is functional and efficient in sperm heads. An adaptive phenotype mediated by the presence of a large amount of YsRNA in the sperm head, and thus as a paternal contribution in the zygote, might be related to an association of YsRNA with Y RNA that prevents the adoption of a Y RNA secondary structure capable of binding Ro60, both from the oocyte and the sperm. We hypothesize that preventing Y RNAs from acting as Ro60-associated gatekeepers for misfolded RNAs in the zygote and early development may enhance RNA chaperoning and thus represent the adaptive molecular phenotype.

1 Introduction

Spermatogenesis is a dynamic developmental process starting from stem cell proliferation and differentiation, meiotic cell divisions and finally an almost complete replacement of the canonical histones by protamines in the sperm head leading to a stronger genome compaction. The main task during this process is to realize different genome safeguarding mechanisms as well as to undergo a differentiation program characterized by an extreme dynamic of the stage specific transcriptomes leading to a highly specialized sperm cell with markedly reduced transcriptional activity. The extreme complexity of the transcriptome during spermatogenesis is characterized by the fact, that nearly the whole genome is expressed in testes, more than in any other cell (Soumillon et al., 2013). This phenomenon is speculated to be linked to an evolutionary important “transcriptional scanning” of the genome exploiting the mechanism of the transcription-coupled repair machinery (Xia et al., 2020). In mammals, mRNA is massively eliminated during late spermiogenesis (Gou et al., 2014) and small RNAs are eliminated during post-testicular maturation of sperm. Thus, piRNAs are almost completely absent from ejaculated sperm, in which a small amount of sncRNAs comprised primarily of tRNA cleavage fragments together with a smaller population of microRNAs (Peng et al., 2012) can be traced. Hence, the characteristic feature of the germline epigenome erasure/reestablishment also affects the ncRNA-transcriptome in the male germline. Summarizing current research in metazoans, it is widely accepted that the oocyte carries the majority of relevant regulator RNAs that function in the early embryo

of most species. However, there is substantial and increasing evidence that sperm also carry a functional RNA payload. Theoretically, small RNAs in the male germline can affect subsequent generations either indirectly by directing chromatin or DNA modifications during gametogenesis or directly via delivery to the zygote. Since the sperm heads are precursors of the male pronucleus, sperm head-located ncRNAs could influence the transcriptomes of early zygotes and before the zygotic genome activation (ZGA) becomes efficient.

It is therefore of utmost importance to precisely describe the complete RNA cargo of mature male gametes, its biogenesis with respect to possible modifications and thus functionality as well as the RNA profiles during development from testis and early stages to ejaculated sperm. Furthermore - and especially for the human situation - it is important to get experimental evidence related to the possible functions of small RNAs in the early embryo (Bošković & Rando 2018).

An annotation of human sperm head smallRNA by UNITAS (Gebert et al., 2017) revealed that a major fraction of miscRNA and RNA, that cannot be annotated, are derived from Y RNAs. There are four Y RNA loci in humans, hY1 with a size of 112 nt, hY3 with a size of 101 nt, hY4 with 93 nt in size and hY5 with a size of 83 nt (Gulia et al., 2020) which are clustered on chromosome 7q36 (Maraia et al., 1994, Maraia et al., 1996). The 5' end, having a triphosphate, and 3' end typically hybridize to form a double-stranded stem domain divided in a lower and upper stem domain (Kowalski & Krude, 2016). Y RNAs are Pol III transcribed and the La binding site is located at the 3' polyuridine tail. Binding of La protein to this tail protects the Y RNA from 3' and 5' exonucleolytic degradation and promotes its retention in the nucleus (Wolin & Cedervall, 2002, Simons et al., 1996). The upper stem domain is important for chromosomal DNA replication and the lower stem domain has a Ro60 binding site and is, when bound to Ro60, involved in RNA stability and stress response (Kowalski & Krude, 2015). Y RNAs do not only exist in their full length but also in fragments between 25 nt and 35 nt (Röther & Meister, 2011) which are referred to as YsRNA. To detect the smallRNA profiles we used smallRNA sequencing, mapping tools and local BLAST-analyses with the Y RNA-homologs as database. To elucidate the YsRNA profiles during spermatogenesis, we NGS-sequenced one human testis small RNA and sperm heads from six individuals. In addition, we downloaded SRA-deposited smallRNA datasets from different early stages of human spermatogenesis, human oocytes, semen vesicles, and epididymis, the latter to take a reproductive support tissue into account. To detect smallRNA

modifications that were introduced during the biogenesis of YsRNAs, we carried out PNK-, RppH- and before/after oxidation experiments with sperm head and testis smallRNA by comparatively quantifying the smallRNA-seq readouts with respect to the various YsRNAs.

2 Material & Methods

2.1 Sperm head preparation

Semen samples were collected from six volunteers between the age of 24 and 62. Donors were asked to abstain from sexual activity for two to three days. The entire ejaculates gained by masturbation were collected in a sterile 50 ml tube. The samples were used within one hour after ejaculation. Semen samples were evenly distributed to 2 ml reaction tubes and centrifuged at room temperature at 16,000 x g for 5 minutes. The supernatants were discarded and the pellets were resuspended in 700 µl TEN (20 mM TRIS, 20 mM EDTA, 200 mM NaCl, [pH 8.0]), 400 µl HPLC Gradient Grade water and 300 µl 10 % SDS and centrifuged at room temperature at 16,000 x g for 5 minutes. The supernatants were removed and the washing steps were repeated a second time. After the final wash the pellets were dissolved in 48 µl ddH₂O and 2 µl 1 M dithiothreitol (DTT) or stored at -80 °C until further usage.

2.2 RNA isolation

Total RNA was isolated using the Quick-RNA™ MiniPrep kit from Zymo Research according to the manufacturer's instructions whereby the samples were mixed with 600 µl of RNA Lysis Buffer in the first step. The isolated RNA was dissolved in 20 µl RNase-free water and stored at -80 °C until further usage.

2.3 small RNA Sequencing

The isolated sperm heads and testis RNA samples were sent to BGI for library preparation and small RNA sequencing. RNA from same donor sperm heads were pooled from several preparations and combined to meet the company's requirements of a total amount exceeding 1 µg RNA. Size fractions with a cut-off at 50 nt were used to enrich small RNA for library preparation and the sequencing strategy was SE 50.

2.4 Oxidation and β -elimination of sperm head and testis RNA

Samples from one individual were used to find out if sperm head YsRNAs are protected at their 3' end due to 2'-O-methylation. To this end, periodate treatment coupled with subsequent β -elimination were performed. Unprotected RNAs with vicinal 2' and 3' diol groups react with sodium periodate (NaIO_4). During subsequent β -elimination these RNAs are shortened by 1 nt, leaving them with a 3' monophosphate. This 3' monophosphate prevents adaptor ligation during library preparation. 2 μg of human testis RNA (BioChain®; #R1234260-50) and total RNA isolated from sperm heads were dried in a SpeedVac and resuspended in 17.5 μl borate buffer (4.38 mM $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ and 50 mM H_3BO_3 [pH 8.6]). 7.5 μl 100 mM NaIO_4 was added to reach a final concentration of 28.6 mM and the samples were incubated 10 minutes in the dark at 24 °C. 3 μl 50 % glycerol and 2 μl HPLC water were added to the samples to reach a final concentration of 5 % glycerol. The samples were incubated for another 10 minutes in the dark at 24 °C. Afterwards, the samples were concentrated to 5 μl in a SpeedVac. For the following β -elimination, 50 μl borate buffer (33.75 mM $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ and 50 mM H_3BO_3 [pH 9.5, adjusted by NaOH]) was added to the samples. The samples were incubated for 90 minutes at 45 °C. Afterwards, the samples were ethanol precipitated. For that, 150 μl 100 % ethanol was added to the samples together with 2 μl glycogen as carrier. The samples were precipitated overnight at -20 °C. After that, the samples were centrifuged at 16,000 x g at 4 °C for 30 minutes. The supernatants were carefully discarded. 100 μl 75 % ethanol was added to the pellets and the samples were centrifuged at 16,000 x g at 4 °C for 10 minutes. The supernatants were removed and the pellets were left to dry by evaporation. After that, the pellets were resuspended in 20 μl HPLC water.

For the control samples 2 μg of testis and sperm head RNA were mixed with HPLC water to reach a volume of 20 μl . SmallRNA-Seq was done by BGI implementing the above-mentioned strategies with respect to size fractionation and sequencing strategy.

2.5 PNK and RppH-treatment

The Y RNA source genes of the YsRNAs are triphosphorylated at the 5' terminus. To find out whether sperm head YsRNA harbour 5' and 3' modifications that prevent them from proper sequencing or if they can be enriched after modifying their ends a 5'- and 3'-termini

modification approach was performed essentially as described by Shi and coworkers (2021). T4 Polynucleotide Kinase (PNK) and RNA 5' Pyrophosphohydrolase (RppH) were used individually and combined according to the manufacturer's instructions. T4 Polynucleotide Kinase catalyzes the transfer of phosphate from ATP to the 5' -hydroxyl terminus of polynucleotides and catalyzes 5'-P dephosphorylation and exchange reactions as well (Eun, 1996). It also catalyzes the removal of 3'-phosphoryl groups. RNA 5' Pyrophosphohydrolase removes pyrophosphate from 5' end of triphosphorylated RNA leaving a 5' monophosphate RNA. For this purpose, samples from one individual were collected and sperm heads were isolated and pooled. RNA was isolated, dissolved and divided into 4 equal fractions. One fraction was kept untreated as the control sample, one sample was treated with PNK (New England BioLabs), one sample was treated with RppH (New England BioLabs) and the remaining fraction was treated with PNK and RppH respectively. The untreated sample contained 1.203 μg RNA, the PNK, RppH and PNK+RppH samples contained 1.508 μg RNA as starting material.

For the treatment with PNK the RNA sample was mixed with 2 μl T4 PNK buffer, 2 μl PNK and HPLC water to reach a final volume of 20 μl . The sample was incubated at 37 $^{\circ}\text{C}$ for 30 minutes. For the RppH treatment the sample was mixed with 2 μl NEB buffer, 2 μl RppH and HPLC water to reach a final concentration of 20 μl . The sample was incubated at 37 $^{\circ}\text{C}$ for 30 minutes. For the combined treatment sample PNK treatment was done followed directly by RppH treatment as mentioned above. After the incubation time all samples were ethanol precipitated. For that 60 μl 100 % ethanol and 0.5 μl glycogen were added to each sample. The samples were vortexed and incubated for 30 minutes at -20 $^{\circ}\text{C}$. After that, the samples were centrifuged for 30 minutes at 16,000 x g at 4 $^{\circ}\text{C}$. The supernatant was carefully discarded, 100 μl 70 % ethanol was added to the pellets and the samples were centrifuged at 16,000 x g at 4 $^{\circ}\text{C}$ for 10 minutes. The supernatants were removed carefully and the pellets were left to dry. After that, the pellets were resuspended in 15 μl HPLC water and sent for sequencing. SmallRNA-Seq was performed by BGI as outlined above.

2.6 RNA Immunoprecipitation

To investigate further on the already known function of Y RNAs we concentrated on their binding capability to Ro60 and the possible role of YsRNA in this context. To do so, one sample of isolated sperm heads was resuspended in 1 ml PBS, cross-linked using formaldehyde at a final concentration of 1 % and incubated for 10 minutes at room

temperate. 265 μ l ice-cold 1 M glycine was added, the sample was incubated on a shaker for 5 minutes at room temperate and centrifuged at 13,000 x g and 4 °C for 5 minutes. The pellet was washed three times with 1.5 ml ice-cold PBS and centrifuged at 13,000 x g for 5 minutes at 4 °C. After the final wash the pellet was dissolved in 500 μ l cold RIPA buffer (50 mM TRIS-HCl, 150 mM NaCl, 2 mM EDTA [pH 8], 1 % Triton X-100, 0.5 % sodium deoxycholate, 0.1 % SDS) and incubated for 10 minutes on ice. To open the sperm heads the sample was sonicated using the Covaris E220 Focused-Ultrasonicator. The Peak Incident Power was set to 140, Duty Factor to 10 %, Cycle of Burst to 200, temperature to 20 °C and duration to 120 seconds. After the sonication immunoprecipitation was performed. To do so, 20 μ l Protein-A/G beads were washed twice with 1 ml 1x PBS + 0.1 % Triton X-100. The supernatant was discarded, the beads were resuspended in 1 ml blocking solution (1 mM EDTA, 10 mM TRIS-HCl [pH 8], 1 % BSA, 1 % polyvinylpyrrolidone) and incubated for 15 minutes on a rotor at room temperature. The beads were placed in a magnetic rack and the supernatant was discarded. 5 μ l anti-TROVE2-antibody (antibodies-online; ABIN7188011) were dissolved in 1 ml 1x PBS + 0.1 % Triton X-100, mixed with the beads and incubated on a rotor for 30 minutes at room temperature. The tube was placed in a magnetic rack and the supernatant was discarded. The beads were washed twice with 1 ml 1x PBS + 0.1 % Triton X-100 and afterwards dissolved in 100 μ l 1x PBS + 0.1 % Triton X-100. The sample was mixed with the beads and incubated on a rotor overnight at 4 °C, placed in a magnetic rack to remove the supernatant. The beads were washed twice with low salt buffer (2 mM EDTA, 20 mM TRIS-HCl [pH 8], 150 mM NaCl, 1 % Triton X-100, 0.1 % SDS), followed by two wash steps with high salt buffer (2 mM EDTA, 20 mM TRIS-HCl [pH 8], 500 mM NaCl, 1 % Triton X-100, 0.1 % SDS) and a final wash with LiCl buffer (1 mM EDTA, 10 mM TRIS-HCl [pH 8], 0.25 M LiCl, 1 % Triton X-100, 1 % sodium doexycholate). The supernatant was discarded. To remove cross-links, the sample was mixed with 120 μ l elution buffer (100 mM NaHCO₃, 1 % SDS), incubated for 15 minutes at 30 °C and centrifuged at 2,000 x g for 1 minute at room temperature. The supernatant was put in a fresh reaction tube and 4.8 μ l 5 M NaCl and 2 μ l Proteinase K (20 mg/ml) were added. The sample was incubated for 30 minutes at 65 °C, 400 μ l TRIzol was added and the sample was incubated for 5 minutes at room temperature. After adding 80 μ l chloroform and incubation for 3 minutes at room temperature with intermitted vortexing, the sample was centrifuged at 16,000 x g for 5 minutes at 4 °C. The aqueous phase was transferred into a fresh reaction tube and precipitated by adding 750 μ l ice-cold ethanol, 1 μ l glycogen and 25 μ l NaAc and was incubated at -20 °C overnight. After centrifugation at 7,500 x g for 1 hour at 4 °C, the

pellet was shortly air-dried and dissolved in 40 μ l RNase-free water and incubated for 15 minutes at 60 °C. Custom library preparation and sequencing were done by Novogene, to enrich for lincRNA rRNA removal was done and library prepared without size selection.

2.7 Bioinformatic analysis

2.7.1 UNITAS annotation

To analyse the data provided by BGI, UNITAS 1.8.0 was used to get a classification into different small RNA classes. Sequences annotated to either miscRNA or sequences that could not be annotated were retrieved as collapsed FASTA files, aligned with SEAVIEW and used to query the non-redundant database from NCBI.

2.7.2 Local BLAST of transcriptomic data with Y RNA sequences

FASTQ files of small RNA transcriptomes or transcriptomes as retrieved from the SRA (Gong et al., 2022; Vojtech et al., 2014; Yang et al., 2019) were adapter-trimmed and mapped to the human genome (hg38) using TRIM GALORE and HISAT2. Next, we ran local BLAST analyses of the transcriptome data with Y RNA sequences as database. Sequences of the four Y RNA homologs were retrieved from NCBI. To take the Y RNA secondary structure into account and minimize redundant hits due to the intrinsic stem part reverse complementarity of Y RNAs the sequences were splitted into two halves (5' and 3'). BLASTN-settings included “-perc_identity 100” to retrieve exclusively 100 % identical hits. Hits were corrected for the overall alignment rate in the respective datasets as determined by HISAT2 and given in reads per million of alignable sequences.

2.7.2. Bioinformatic evaluation of RIP data

FASTQ files as obtained from Novogene were HISAT2-mapped to hg38 and the resulting bam-files were used to extract the RNA-seq reads defined by both the hg38 transcriptome ENSEMBL gtf-file and a gtf as obtained from the LNCipedia database version 5.2 (Volders et al., 2019). Finally, the counting of mapped reads was carried out applying the FEATURECOUNTS routine from the SUBREAD package (Liao et al., 2014). The counts

were corrected for the alignable reads in the dataset as obtained by HISAT2 and is given in reads per million alignable sequences.

3 Results

3.1 Small non-coding RNA-Seq and UNITAS annotation uncover fragments of Y RNA as a major RNA payload in sperm

Sperm heads were obtained from six individual ejaculates, RNA isolated and smallRNA-sequenced. A range of 25 to 46 million total reads were obtained from each individual and the quality of the datasets was determined by HISAT2 mapping to hg38, which resulted in a mean percentage of alignable reads of 80.49 %. A complete annotation of the sperm head smallRNA payload was done by UNITAS. As a result, the no annotation fraction-values ranged from 40 % to 60 % and mean values for rRNA, tRNA, miRNA and lincRNA were 19.06 %, 9.76 %, 5.15 % and 2.53 %, respectively. Other fractions mapped to piRNA clusters (5.49 %) and miscRNA (4.07 %), respectively (for the respective individual donut charts as obtained by UNITAS see Supplemental figure 1). To get an idea about the not annotated and miscRNA-sequences, the respective collapsed FASTA-files provided by UNITAS were retrieved and aligned by SEAVIEW. BLAST searches of consensus queries revealed that the main fraction of these sequences were similar to human Y RNA.

To more precisely describe these short Y RNA sequences with respect to their Y RNA homologs and to narrow down the regions of similarity, local BLAST searches were carried out. To this end, the four human homologs of Y RNA (hY1, hY3, hY4 and hY5) - each around 100 nt in length - were divided in roughly two halves and used as database, to avoid a confounding of our results due to the cryptic complementarity of the Y RNA-stem regions.

3.2 The 5' region of hY1 is enriched in human sperm heads and almost absent in oocytes

Our sperm heads smallRNA data was blasted against the Y RNA-database. Only hits with a percentage identity of 100 % were used to calculate mean values for Y RNA hits. The SRA deposited oocyte datasets (BioProject PRJNA376426) were quality checked with TRIM GALORE and clean sequences retrieved. Subsequently, HISAT2 was carried out to determine the overall alignment rate. The clean sequences were used for the local BLAST

analyses. The comparison of perfect hit numbers specific for the 5' and 3' part of the Y RNA sequences and separated for the four homologous Y RNA loci is depicted for both sperm heads and oocytes in Fig. 1.

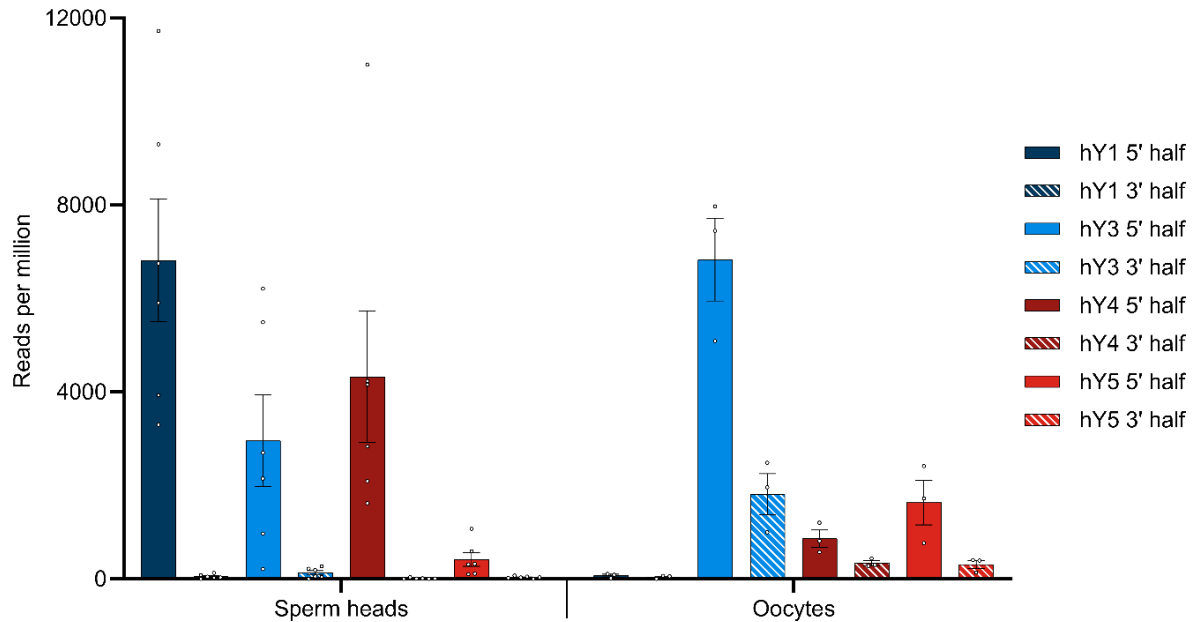


Fig. 1.: BLASTN results querying smallRNA with Y RNA as database. Hits were corrected for the overall alignment rate in the respective datasets as determined by HISAT2 and given in reads per million of alignable sequences. Dots indicate individual hit numbers and bars indicate the standard error of mean (SEM).

In general, the quantitative distribution of YsRNA uncovers that the 5' regions of the Y RNAs can be detected in high abundance in all datasets, whereas the respective 3' parts of the Y RNAs are virtually absent in sperm heads. This 5' part to 3' part imbalance is less pronounced in oocytes however with a clear tendency of preponderance of 5' Y RNA regions, too. Interestingly, the 5' part of hY1 is enriched in human sperm heads and almost absent in oocytes, for that the 5' fragments derived from hY3 show the highest abundance.

3.3 The sperm head hY1 5' regions are predominantly 30 nt and 31 nt in length

Next, we asked if the YsRNAs we detected exhibit a continuous size profile or if discrete size classes could be uncovered. To this end, we counted the number of perfect hits with a certain length (see supplemental figure 2).

The length distributions of Ys1RNA (5') and Ys3RNA (5') peak at 30 nt and 31 nt, respectively, starting from the first and second base of the source Y RNA. This length falls

into the size range of piRNAs. Ys1RNA of these sizes are restricted to the sperm heads, whereas these Y RNA fragments are virtually absent in oocytes. The most abundant YsRNA length in oocytes is 31 nt and 32 nt (data not shown). 5' Ys1RNA in the sperm head as precursor of the male pronucleus thus constitute a paternal contribution to the zygote.

3.4 5' & 3' modifications of YsRNA point towards Y RNA as source genes for piRNAs in human sperm head

To investigate further on the similarity of YsRNA and piRNA we checked sperm head RNA for possible modifications that are typical for piRNAs by setting up an oxidation/ β -elimination experiment. To check with the situation in early stages of spermatogenesis and to have an internal control we included bulk testis RNA in our assay. Samples from the same donor were used to find out if YsRNA are protected at their 3' end due to 2'-O-methylation. To this end, periodate treatment and β -elimination were performed respectively. SmallRNA-Seq was performed and the respective data was blasted against the Y RNA database. Early stages of spermatogenesis were represented by treating human testis RNA as described. In addition, oocyte data from before/after oxidation set ups were included in our analysis to directly compare the situation in early stages of spermatogenesis with sperm heads and the maternal site. Fig. 2 summarizes the results of the β -elimination experiments.

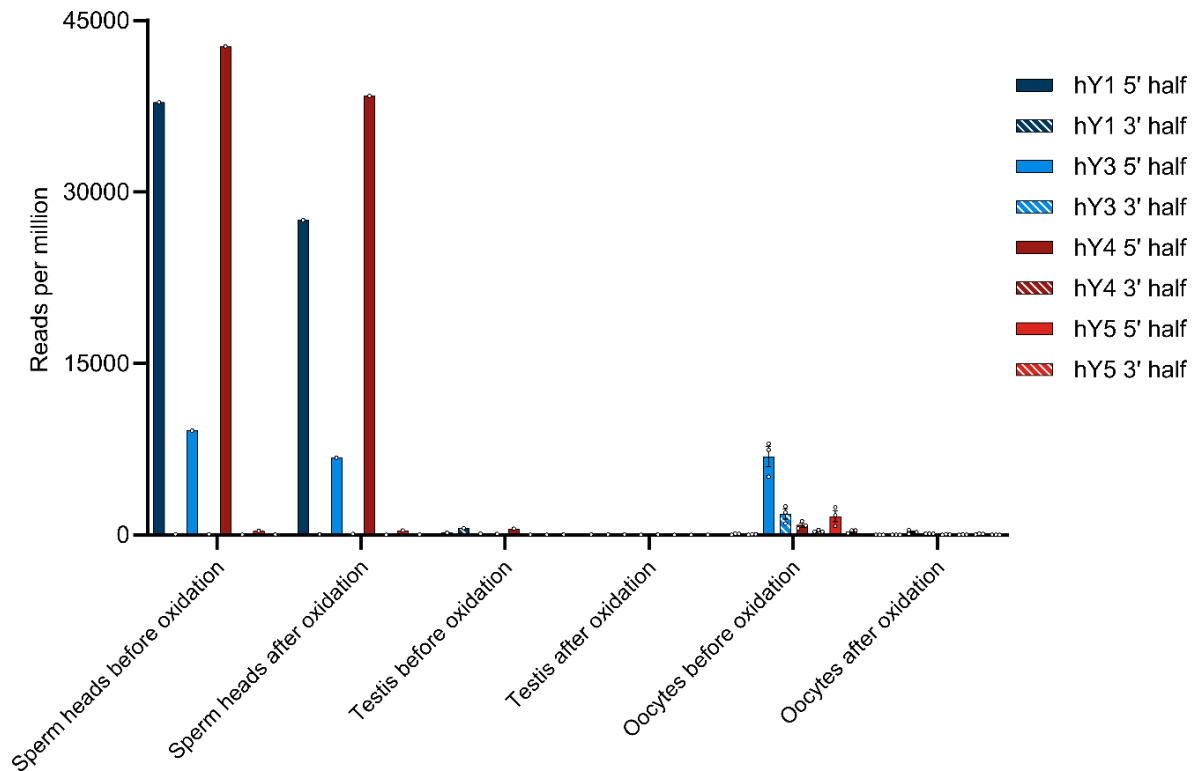


Fig. 2: Number of YsRNA hits in sperm heads, testis and oocytes before and after oxidation given in reads per million alignable reads. Individual hits are shown as dots with the error bars (SEM).

In the sperm head the Ys1RNA patterns point to an extensive protection of these RNAs - more precisely the 31 nt fraction (see supplemental figure 3). Both fragment size and 3' modification make these fragments reminiscent of canonical piRNA, albeit not derived from piRNA clusters. Sperm heads in addition contain considerable amounts of Ys4RNA that are protected at their 3' end due to 2'-O-methylation. In contrast, these piRNA-like YsRNAs are virtually absent from early developmental stages of spermatozoa, an observation we confirmed by analyzing bulk RNA from testis, containing spermatogonia, spermatocytes, spermatids and spermatozoa of different developmental stages. This makes the sperm head YsRNA/piRNAs almost entirely an exclusive RNA payload and paternal contribution to the zygote transmitted via the male pronucleus.

The source genes of the YsRNA are 5' triphosphorylated, a modification that would prevent a representation in NGS libraries. To test if these 5' modification is present in YsRNA and to elucidate further modifications, we set up a scheme of tests, similar to the PANDORA-Seq-strategy (Shi et al., 2021). Thus, RppH and PNK-treatment was carried out in a separate and combined setting. NGS readouts were HISAT2-mapped to hg38 to see if the different treatments result in a significantly less alignment rate, indicating the introduction of

inadvertent chemical modifications preventing adaptor ligation. Focusing on the canonical 5' triphosphate of Y RNA source gene products, we could observe an RppH- effect in 5' Ys4RNA and 5' Ys5RNA, representing an internal positive control and indicating the presence of 5' triphosphates in these molecules. For Ys1RNA and Ys3RNA the untreated sample results in the highest abundance of the fragments, which is indicative of these fragments to possess canonical 5' phosphates that can be ligated and sequenced.

Upon treating the samples with PNK alone or in combination with RppH, the hit number for the 5' regions of both hY1 and hY4 are markedly reduced.

An overview of the results is depicted in Fig. 3.

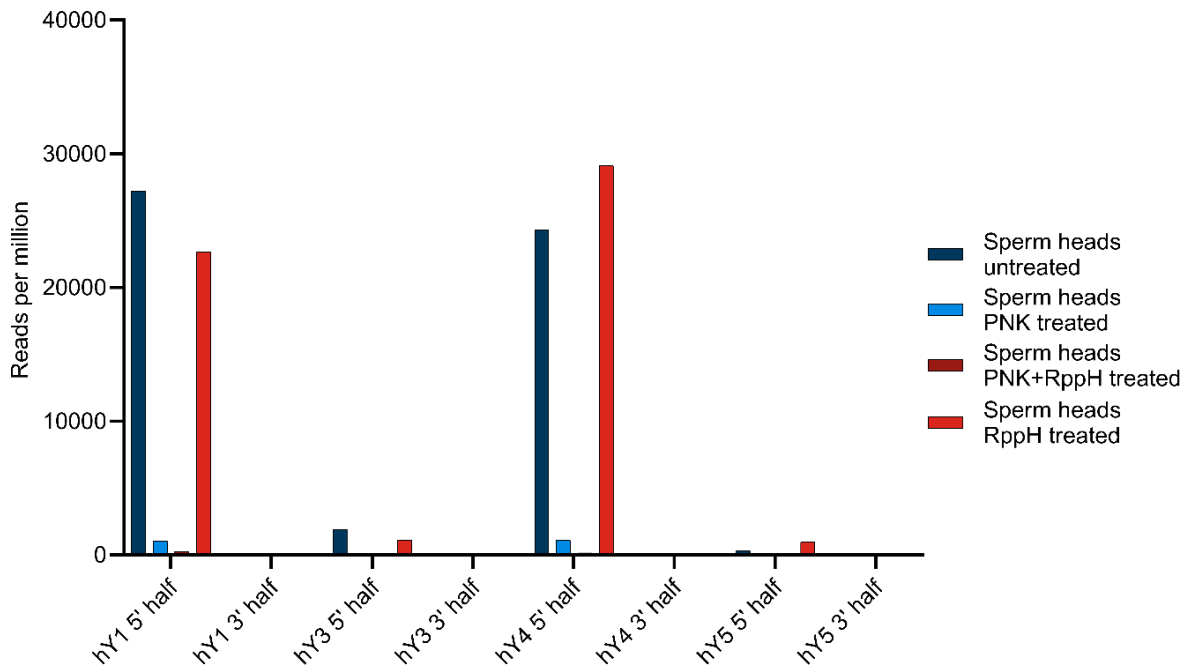


Fig. 3: Number of YsRNA hits in sperm heads without treatment, treatment with PNK, RppH and both enzymes, respectively in reads per million alignable reads shown for the four Y RNAs, divided in a 5' half and a 3' half.

3.5 YsRNA are soma-germline transmitted presumably via epididymal vesicles

We could not observe a biogenesis of the YsRNA/piRNAs profiles along with the developmental stages of sperm cells, instead a sudden appearance in mature sperm. Interestingly, Y RNAs and YsRNAs were described to be highly abundant in seminal plasma exosomes, more precisely in two size ranges from 20-40 nt and 40-100 nt, (Vojtech et al., 2014). We downloaded the respective SRA accessions (BioProject PRJNA242348) and analyzed the data with respect to abundances of the different YsRNA. The same BLAST-

strategies as above were applied and a comparison of sperm heads and exosomes with respect to hits per million of mappable reads is given in Fig. 4.

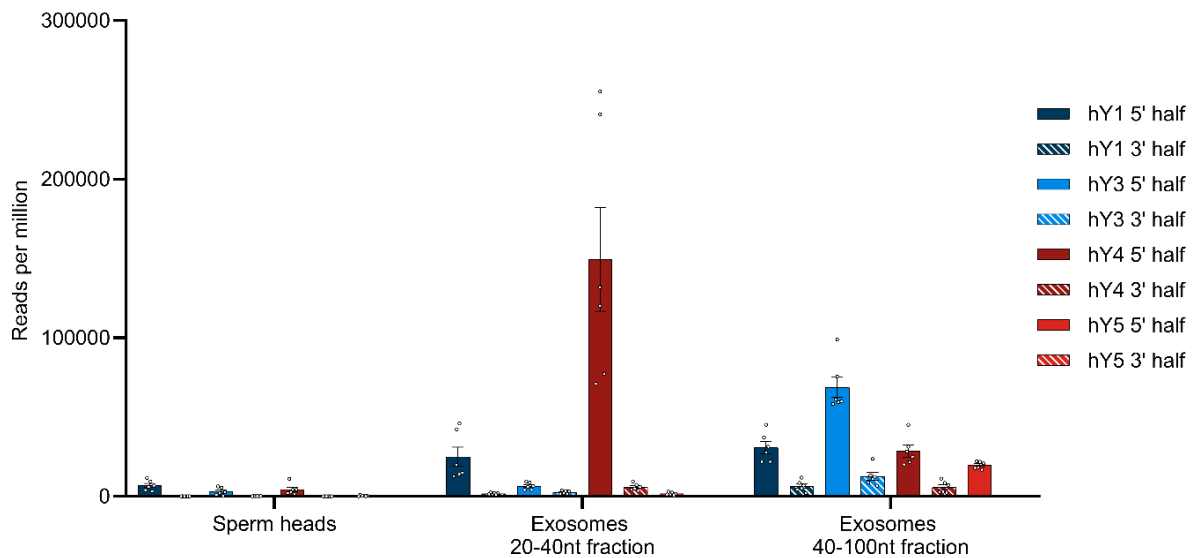


Fig. 4: YsRNA hits shown for sperm heads and exosomes in reads per million mappable reads for the four Y RNA homologs divided in a 5' half and a 3' half. Bars are shown with error bars (SEM) and individual hits are shown as dots.

Seminal plasma exosomes do contain a considerable fraction of the YsRNAs that can be found in human sperm heads, albeit the profiles exhibit quantitative differences mainly in the Ys1RNA to Ys3RNA ratio.

Next, we asked what role the reproductive support tissues possess in that respect. Spermatozoa mature during spermiogenesis and acquire functional competence. It is generally accepted that over a broad range of metazoans the RNA payload in sperm is also controlled by the soma (Conine & Rando, 2022). Different reproductive support tissues are described for a taxonomically broad metazoan sample, for the mammalian site, a key player in this process is represented by the epididymis. To scrutinize if the YsRNAs are present in epididymis, we downloaded RNA-Seq data obtained from Gong and coworkers (BioProject PRJNA821911). These data were divided in the epididymal regions of caput, corpus and cauda. BLAST routines were carried out as above and the results in hits per million hg38-alignable reads are depicted for the three epididymis regions and the Y RNA sequences subdivided in the 5' and 3' regions.

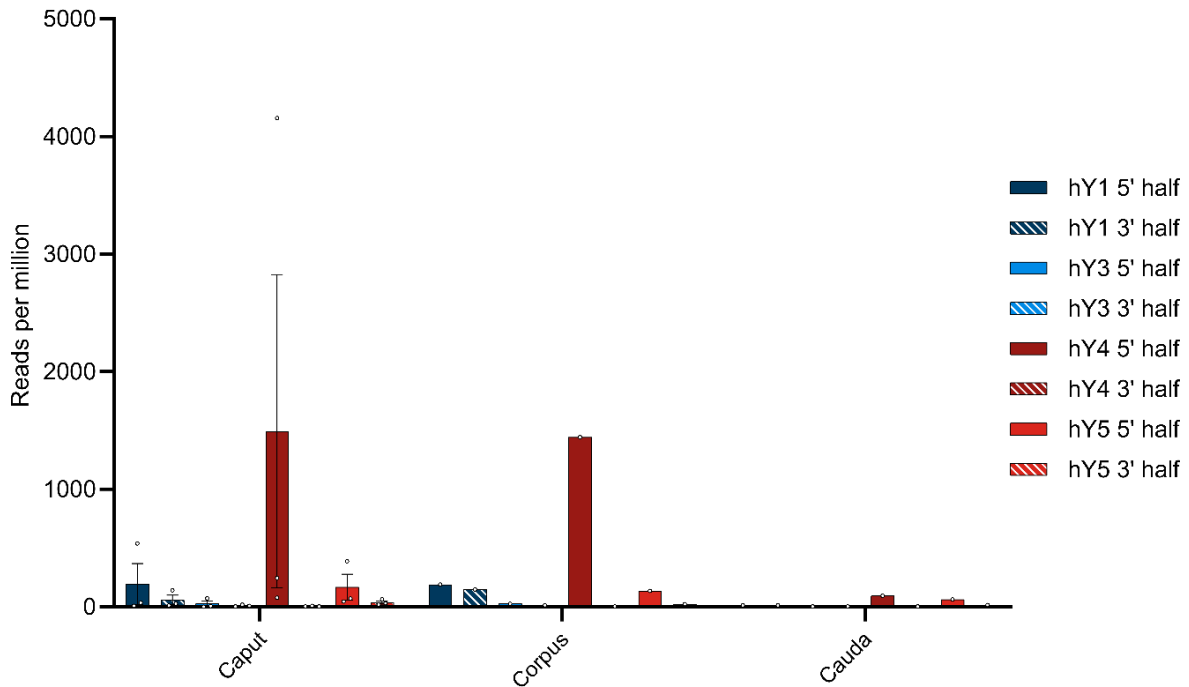


Fig. 5: Profile of YsRNA in epididymis divided in caput, corpus and cauda region in reads per million alignable reads and depicted for the different Y RNA homologs. Dots show the individual hits and bars are depicted with error bars (SEM).

The preponderance of Ys4RNA reflects both the situation in semen exosomes (see above) as well as in the epididymal samples. From this distribution of the abundances of individual YsRNA homologs, we conclude that the epididymis is a main determinant of the semen exosome profiles with respect to their YsRNA payload. The hits decrease with the passage through the epididymis ranging from high values in caput and corpus to smaller values in the caudal portion of the epididymis. This pattern largely recapitulates the observation of the soma-to-germline transfer via vesicles as described by Conine and Rando (2022) in mice, albeit with the YsRNA as a typical payload in humans. Interestingly, HENMT1 message, an enzyme responsible for the 3' 2'-O-Methylation, can be traced in the epididymis datasets, highlighting a possible role of the epididymis in the 3' protection of YsRNAs that are transferred from soma to germline.

3.6 Sperm head Ro60 is not associated with Y RNAs

Y RNAs are bound by Ro60, a protein that also binds to misfolded non-coding RNAs, including pre-5S rRNA (Kowalski & Krude, 2015). Beside other possible functions, Ro60 binds to endogenous Alu retroelements and regulates their expression. To test the

functionality and substrate specificity of the Ro60 system in sperm heads, the profiles of Ro60 bound RNA were compared to the sperm head transcriptome (BioProject PRJNA890147), more precisely the non-coding RNA transcriptome including both small non-coding and long non-coding RNA. For this purpose, RNA Immunoprecipitation was performed and the bound RNAs were recorded in a transcriptome-wide readout. LincRNA transcriptome gtf-files were obtained from the LNCipedia-database (hg38) and used as input files for READCOUNT following a mapping with HISAT2. The overall alignment rate was used to correct the total number of reads for the number of alignable reads in our dataset, and both transcriptomes and immunoprecipitation-obtained sequences were analyzed comparatively.

Fig. 6 shows the two profiles for the total lincRNA-transcriptome in red and the Ro60 bound fraction in blue, respectively.

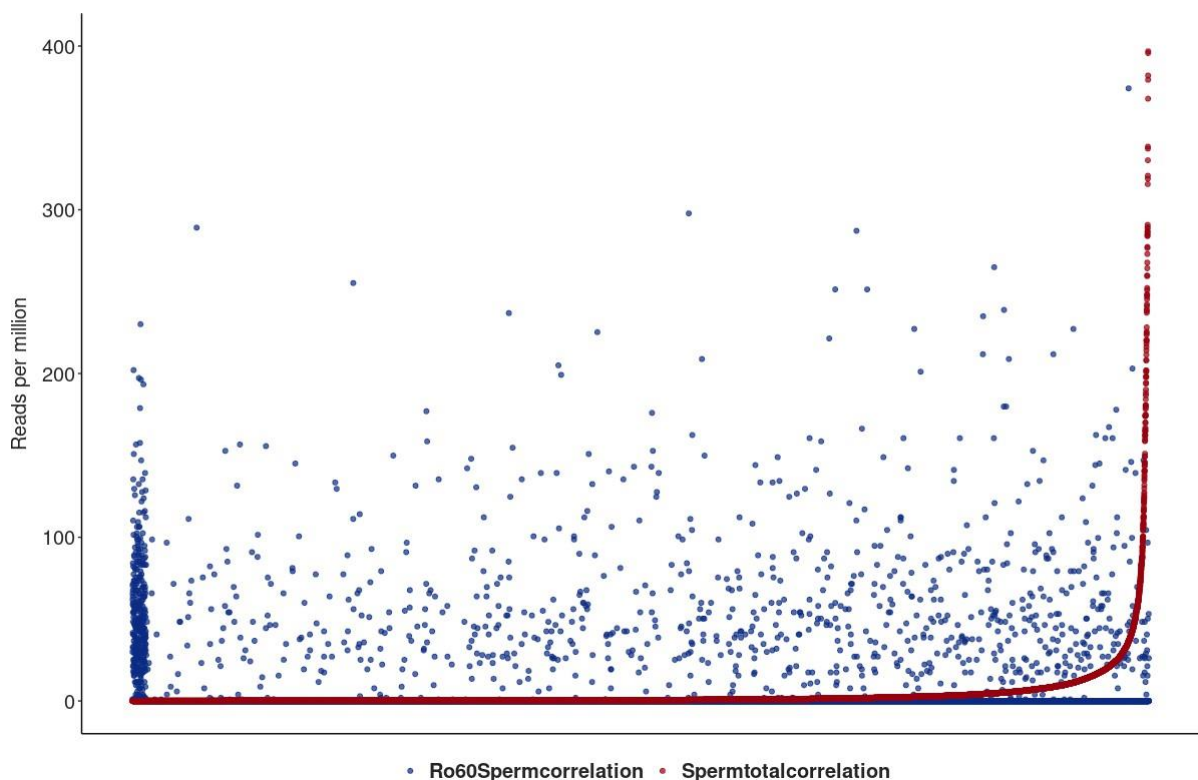


Fig. 6: Illustration of the total sperm transcriptome-lincRNA profile in red and the Ro60 bound fraction in blue in reads per million alignable reads. On the x-axis the lincRNA loci are listed. lincRNAs are sorted by number of reads per million reads from smallest to largest for the total sperm transcriptome. For better graphical readability, the y-axis was cut at 400 reads per million reads (many lincRNAs exceed this hit number in the total transcriptome).

Figure 6 shows the lincRNA profile as obtained from the total sperm transcriptome (red) together with the profile of the Ro60-bound lincRNA (blue) in reads per million alignable reads. Obviously, the profiles are markedly different and we conclude, that Ro60 binding to lincRNA - as prerequisite for chaperoning - is specific and apparently functional in sperm heads. Interestingly, and though we observed merely a small number of Y RNA BLAST-hits in the total transcriptome, Ro60 was not associated with Y RNAs of the different homologous loci as could be seen after querying the Ro60-RNA-ChIP-data with Y RNA in a BLAST routine.

4 Discussion

In metazoans, the processes of gametogenesis, fertilization, and early embryonic development are characterized by an erasure and restoration of epigenetic marks. However, unlike in plants, where epigenetic inheritance is a widespread phenomenon, it is increasingly believed that at least some epigenetic information in metazoans can also be transmitted between generations. We focused on small non-coding RNAs as carriers of epigenetic information.

Animal oocytes are much larger and carry more RNA than spermatozoa, and the maternal load of long and small RNAs is essential for the initiation of embryogenesis. In contrast, the contribution of paternal RNAs to the zygote and their regulatory effects, especially prior to zygotic genome activation, are not fully understood. Our work initially focused on sperm heads to reflect the situation at the onset of male pronucleus formation and first performed an unbiased analysis of total smallRNA cargo by UNITAS annotation. Although the relative amounts of smallRNAs annotated by UNITAS vary between individuals and biological replicates, clear trends can be observed, such as a relative lack of piRNAs in mature spermatocytes compared to the situation in testis. In addition to mass abundance profiles of

broad classes of small RNAs (between testis and spermatozoa), we uncovered small Y RNAs as a component of the smallRNA profile in the sperm head.

Y RNAs are short, non-coding RNAs that are transcribed from Pol III. In humans, there are four homologs encoded in a cluster on Chr. 7. The evolutionary history of this cluster includes gene losses and duplications during vertebrate evolution (Mosig et al., 2007). In addition, a substantial number of pseudogenes have been described for hg38 that coalesce relatively deeply. The transcripts of all Y RNA homologs are approximately 100 nt long and are characterized by a 5' triphosphorylation and a 3' polyuridine tail that forms the La-binding site of the Pol III transcript. The secondary structures (Kowalski & Krude 2015) show different stem-loop regions in all homologs, which can be assigned different functional roles. In particular, the lower stem domain is crucial for Ro60 binding and thus for the formation of RoRNPs. The Y RNAs of the different homologs regulate the subcellular localization of Ro60 by binding to the outer surface of Ro60 and mainly regulate the entry of misfolded or defective non-coding RNAs into the circular Ro molecule. Smaller fragments from the Y RNA source genes are referred to as Y RNA-derived small RNAs (YsRNA), which have been described for apoptotic cells, for example (Guglas et al., 2020). Because the sequencing strategy applied for small RNA includes size enrichment with cut-off at 50 nt and SE50 – NGS sequencing, we observed YsRNA in sperm heads with UNITAS, grouping them in miscRNA and to a lower extent in non-annotated sncRNA, respectively.

To further discriminate between the different source Y RNA homologs, we constructed a database using the homolog-specific sequences as obtained from NCBI to search for Y RNA and YsRNA. To this end, the Y RNA sequences were each divided in two halves, mainly to avoid a confounding of our search results by the complementarity of the 5' and 3' ends. Local BLAST routines were carried out and only identical sequences counted as hits. From that YsRNA profiles were obtained, which are generally characterized by the sheer absence of hits for the 3' YsRNA halves and a preponderance for Ys1RNA followed by Ys4RNA and Ys3RNA in sperm heads. The least hit number was obtained for Ys5RNA. Next, we checked the YsRNA profiles in oocytes. To this end, we downloaded an SRA deposited human oocyte single cell dataset consisting of three biological replicates that were analyzed without further treatment and after an oxidation/ β -elimination experiment. (BioProject PRJNA376426). This oocyte dataset was analyzed to unravel sncRNA-patterns with unprecedented resolution and sensitivity (Yang et al., 2019). The respective oocyte profiles were compared to the sperm head situation, which generated similar results with respect to the observation that 3'

halves of the YsRNA were underrepresented as well. However, the main observation was that Ys1RNA prevails in sperm heads, whereas it is virtually absent in oocytes. For the latter, we could detect Ys3RNAs as the major constituent of the YsRNA fractions.

To get a more precise picture of the YsRNAs, we separately counted the different length of the perfect hits with the result that sperm head Ys1RNA peaks at 30 nt and 31 nt, falling into the size range of canonical piRNAs. These 30 nt and 31 nt Ys1RNA are absent in human oocytes, thus representing a paternal contribution to the zygote introduced upon fertilization and upon forming the male pronucleus.

Since the Ys1RNA fall in the size class of piRNAs and because the likely source Y RNA is triphosphorylated at the 5' end, we asked if we do find specific YsRNA modifications, with some of them having the potential to exclude the YsRNAs from successful adaptor ligation and/or reverse transcription and thus excluding them from being represented in the NGS datasets. To check the 2'-O-methylation status at the 3' terminus of YsRNA - a hallmark of piRNAs - before/after oxidation assays were performed in sperm heads. For this purpose, NaIO₄ oxidation treatment and β -elimination protocols were carried out for the sperm heads RNA and the respective oxidation data from single cell oocyte-smallRNA seq (Yang et al., 2019) were compared. In addition, we smallRNA-sequenced human testis RNA in a before/after oxidation-comparison to check for early stages of sperm development in a bulk testis RNA-analysis. The most striking change inferred from UNITAS analysis of RNA is the enrichment of piRNA in sperm heads and concomitant reduction of most ncRNA classes as annotated by UNITAS (with an exception of 5' tRFs and lincRNA). Among YsRNAs in sperm heads, we observe a high abundance of 5' halves of both Ys1RNA and Ys4RNA, both slightly decreasing after oxidation. In contrast, the much lower amounts of the respective YsRNA in oocytes completely disappear after oxidation/ β -elimination. We conclude that Ys1RNA and Ys4RNA – beside possessing the canonical length – also exhibit the 2'-O-methylation at the 3' terminus characteristic of canonical piRNAs and contrasting the situation in oocytes. We therefore propose that the sperm head YsRNA is a functional piRNA that is not encoded on piRNA-clusters and thus not annotated as piRNA by UNITAS. Querying the RNACentral database (rnacentral.org) with the “search by sequence option”, the 30 nt and 31 nt sized Ys1RNA fragment yielded piRNA hits in several species – not humans - including a new world monkey (*Callithrix jacchus*). Interestingly, this hit was obtained for an analysis of *Callithrix* testis, whereas our human testis smallRNA analysis exhibited a complete lack of YsRNAs in testis, both in control and oxidation experiments.

Considering the 5' triphosphate modifications of the source Y RNA genes and possibly resolving cyclic 2'-3' phosphates by the phosphatase activity of polynucleotide kinase (PNK), we set up an enzymatic treatment scheme with RppH- and PNK-treatment of total RNA obtained from sperm heads. Considering the RppH- assays we could not observe great change in the abundance of the YsRNA, suggesting that in contrast to the triphosphorylated 5' terminus of canonical full-length Y RNAs, the respective fragments exhibited adaptor-ligatable ends to be fully represented in our NGS output. In the PNK-assay and more broadly, as determined by UNITAS annotation, we found significant differences between the smallRNA profiles before and after treatment. Most strikingly, PNK-treatment increases the abundance of UNITAS annotated rRNAs, suggesting a significant proportion of rRNA-annotated smallRNAs with 2',3' -cyclic phosphate-containing 3' -ends. This observation partially confirms interpretations from Shigematsu et al. (2019) serving as an internal control of the PNK-effect for the herein presented data on YsRNA. With respect to YsRNA, we did observe a prominent decrease in the abundance of YsRNAs after PNK treatment. Because RppH experiments indicate a monophosphate at the YsRNA 5' terminus and the 2'-O-methylation at the 3' terminus prevents the formation of 2',3' -cyclic phosphate at the 3' -ends, we conclude that the 5' phosphatase activity of PNK (Eun, H-M. 1996) catalyzes the removal of the canonical 5' phosphates of the YsRNAs, thus hindering efficient ligation and representation in NGS libraries.

Since we covered sperm cells from all stages of spermatogenesis in our testis RNA analysis and did not find YsRNAs, we conclude that YsRNAs are not present in early stages of sperm development. It is known that spermatozoa gain full function only after leaving the testis and migrating through the epididymis, where they acquire motility and the ability to fertilize oocytes (Cornwall & von Horsten, 2007). Which role YsRNAs take in that process and how and when small RNAs in general are gained or lost during this post-testicular maturation is a question we attempted to answer.

Recent studies in mice suggest that sperm carry RNAs, that are synthesized in epididymal somatic cells (Sharma et al., 2018). The data presented demonstrate that soma-germline RNA transfer occurs in male mammals, most likely via vesicular transport from the epididymis to maturing sperm. To elucidate if this is similarly realized for YsRNAs in humans too, we queried NGS datasets of human semen exosomes and human epididymis (BioProjects PRJNA242348 and PRJNA821911), the latter supposed to be a hub of soma-germline-

transfer via vesicles in mice. Interestingly, we obtained rather similar YsRNA profiles upon comparing semen exosomes and epididymis, subdivided in different structures, including caput, corpus, and cauda. The relative proportion of YsRNA in the RNA preparations was very high as compared to sperm heads and with a preponderance of the Ys4RNA (5' part) exceeding the amount of Ys1RNA (5' part). The similarity between epididymal YsRNA profiles and semen exosome patterns suggests that these exosomes originated from epididymis epithelium, most prominently from the caput and corpus regions. Our data extend the observations of a high dynamic of the smallRNA payload during spermatogenesis in the murine model system (Sharma et al., 2018) by the formation of YsRNA and a soma to germline transfer during spermiogenesis in the epididymis in humans. Though mechanistically several questions are still open, upon checking the presence of HENMT1 – the factor that is responsible for catalyzing the 2'-O-methylation at the 3' end of piRNAs – we could trace the respective message in the epididymal transcriptomes. The epididymis is thus a highly likely tissue to carry out essential steps of YsRNA biogenesis in humans, rendering the source molecule of Y RNAs into a canonical piRNA that is later transferred between soma and germline. This epigenetic inheritance in humans represents a paternal contribution to the zygote and more importantly is not restricted to intergenerational inheritance but represents the hallmarks of transgenerational inheritance in that the soma-germline transfer occurs in every generation. Since the soma-germline transfer of individual classes of sncRNA seems to vary between species, there is an obvious quest for the adaptive value of the paternal contribution of YsRNAs to the zygote upon fertilization in humans. Human sperm require epididymal passage to become a functional sperm cell. It is widely recognized, and has been impressively demonstrated by Liu and co-workers (2001), that sperm origin is critical for intracytoplasmic sperm injection (ICSI) - the most commonly used technique to aid reproduction in humans. Using epididymal sperm appears to adversely affect the morphologic grade and cleavage stage of the resulting embryos as compared to ejaculated sperm. Overall, this leads to a differential reproduction or – synonymously - selection phenomenon. Although it is difficult to link this general phenomenon to the mechanisms by which these YsRNAs may achieve regulation in progeny, we hypothesize that – in case of YsRNAs – this may be related to the function of the source Y RNAs, which act as a gate keeper for the RNA chaperoning Ro60 system.

At first, we detected Ro60 bound RNA in the sperm head with a binding profile that we interpret as functional and specific when compared to the total RNA payload in human sperm

heads. Ro60 protein is thus present in sperm heads and was recently shown to be present in human oocytes as determined by mass spectrometry (Dang et al., 2023). Upon fertilization and before the zygotic genome activation is realized, a process that requires Pol III transcripts to ensure proper translation, functional and 2'-O-methyl-protected YsRNA is delivered into the zygote via the male pronucleus. Upon Pol III transcription, we hypothesize that Y RNAs are synthesized abundantly which could - upon binding to Ro60 - significantly reduce RNA chaperoning activity of Ro60. With this in mind, we speculate that the presence of paternally contributed YsRNA in a molar ratio exceeding the oocyte and early embryo Y RNA might interfere with the folding of Y RNA stem domain that is binding Ro60, resulting in a more efficient RNA chaperoning activity of Ro60 in early embryogenesis.

Because our hypothesis is mainly correlative, further experiments are warranted to understand the largely enigmatic role of the sperm RNA payload in early embryo and the associated regulatory phenomena. On one side, our results contribute to knowledge about the importance of epigenetic inheritance in recent human evolution and its possible adaptive value. Secondly, the knowledge on the biogenesis and profile of small RNAs as epigenetic information carriers in human mature sperm could be of medical importance, especially for reproduction assisting techniques, too. Mammalian model systems might offer the opportunity to interfere with the soma-germline transfer, the herein involved vesicles and their small RNA cargo, to directly observe the quantitative effects on reproduction. With the use of small extracellular vesicles as “delivery vesicles” in oncology in mind, it is reasonable to assume that assisted reproduction techniques could benefit from the findings on small RNA as epigenetic information carriers in human sperm transmitted between soma and germline.

5 Author Contributions

Conceptualization DE and HZ; methodology DE and HZ; bioinformatic analysis DE and MB; data interpretation DE and HZ; figures, tables, graphics, DE; manuscript writing DE and HZ; manuscript review and editing DE, MB, and HZ. All authors read and approved the final manuscript.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Funding

“This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2526/1 – Projectnr. 407023052.”

8 Acknowledgments

This publication is based on data collected as part of DE’s dissertation project at the Department of Biology, Johannes Gutenberg University, Mainz, Germany. We thank members of the DFG-RTG GenEvo (GRK2526/1) and the Thesis Advisory Committee for helpful discussions and advice. Our thanks go to the IMB (Mainz) Genomics Core Facility for help with the Covaris treatment and G. Schamber for pilot experiments on the interaction of Ro60 and Y RNA. We are indebted to all donors of sample material.

9 References

- Bošković, A., & Rando, O. J. (2018). Transgenerational epigenetic inheritance. In *Annual Review of Genetics* (Vol. 52). <https://doi.org/10.1146/annurev-genet-120417-031404>
- Conine, C. C., & Rando, O. J. (2022). Soma-to-germline RNA communication. In *Nature Reviews Genetics* (Vol. 23, Issue 2). <https://doi.org/10.1038/s41576-021-00412-1>
- Cornwall, G. A., & von Horsten, H. H. (2007). Sperm Maturation in the Epididymis. In *The Genetics of Male Infertility*. https://doi.org/10.1007/978-1-59745-176-5_13
- Dang, Y., Zhu, L., Yuan, P., Liu, Q., Guo, Q., Chen, X., Gao, S., Liu, X., Ji, S., Yuan, Y., Lian, Y., Li, R., Yan, L., Wong, C. C. L., & Qiao, J. (2023). Functional profiling of stage-specific proteome and translational transition across human pre-implantation embryo development at a single-cell resolution. In *Cell Discovery* (Vol. 9, Issue 1). <https://doi.org/10.1038/s41421-022-00491-2>
- Eun, H.-M. (1996). Enzymology Primer for Recombinant DNA Technology. In *Enzymology Primer for Recombinant DNA Technology*. <https://doi.org/10.1016/b978-0-12-243740-3.x5000-5>
- Gebert, D., Hewel, C., & Rosenkranz, D. (2017). Unitas: The universal tool for annotation of small RNAs. *BMC Genomics*, 18(1). <https://doi.org/10.1186/s12864-017-4031-9>
- Gong, J., Wang, P., Liu, J. C., Li, J., Zeng, Q. X., Yang, C., Li, Y., Yu, D., Cao, D., & Duan, Y. G. (2022). Integrative Analysis of Small RNA and mRNA Expression Profiles Identifies Signatures Associated With Chronic Epididymitis. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.883803>
- Gou, L. T., Dai, P., Yang, J. H., Xue, Y., Hu, Y. P., Zhou, Y., Kang, J. Y., Wang, X., Li, H., Hua, M. M., Zhao, S., Hu, S. Da, Wu, L. G., Shi, H. J., Li, Y., Fu, X. D., Qu, L. H., Wang, E. D., & Liu, M. F. (2014). Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Research*, 24(6). <https://doi.org/10.1038/cr.2014.41>
- Guglas, K., Kołodziejczak, I., Kolenda, T., Kopczyńska, M., Teresiak, A., Sobocińska, J., Bliźniak, R., & Lamperska, K. (2020). Y RNAs and Y RNA-derived fragments as new players in cancer research and their potential role in diagnostics. In *International Journal of Molecular Sciences* (Vol. 21, Issue 16). <https://doi.org/10.3390/ijms21165682>

- Gulia, C., Signore, F., Gaffi, M., Gigli, S., Votino, R., Nucciotti, R., Bertacca, L., Zaami, S., Baffa, A., Santini, E., Porrello, A., & Piergentili, R. (2020). Y RNA: An overview of their role as potential biomarkers and molecular targets in human cancers. In *Cancers* (Vol. 12, Issue 5). <https://doi.org/10.3390/cancers12051238>
- Kowalski, M. P., & Krude, T. (2015). Functional roles of non-coding Y RNAs. In *International Journal of Biochemistry and Cell Biology* (Vol. 66). <https://doi.org/10.1016/j.biocel.2015.07.003>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7). <https://doi.org/10.1093/bioinformatics/btt656>
- Liu, Z., Fluker, M. R., Nigro, M., Elliott, S., & Yuzpe, A. A. (2001). Fertilization Rates, Embryo Development and Pregnancy Rates after Intracytoplasmic Sperm Injection (ICSI) Using Epididymal and Testicular Sperm. *Journal SOGC*, *23*(7). [https://doi.org/10.1016/s0849-5831\(16\)31317-9](https://doi.org/10.1016/s0849-5831(16)31317-9)
- Maraia, R. J., Sasaki-tozawa, N., Driscoll, C. T., Green, E. D., & Darlington, G. J. (1994). The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. *Nucleic Acids Research*, *22*(15). <https://doi.org/10.1093/nar/22.15.3045>
- Maraia, R., Sakulich, A. L., Brinkmann, E., & Green, E. D. (1996). Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic Acids Research*, *24*(18). <https://doi.org/10.1093/nar/24.18.3552>
- Mosig, A., Guofeng, M., Stadler, B. M. R., & Stadler, P. F. (2007). Evolution of the vertebrate Y RNA cluster. *Theory in Biosciences*, *126*(1). <https://doi.org/10.1007/s12064-007-0003-y>
- Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., Lei, L., Han, C., Ning, L., Cao, Y., Zhou, Q., Chen, Q., & Duan, E. (2012). A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. In *Cell Research* (Vol. 22, Issue 11). <https://doi.org/10.1038/cr.2012.141>
- Röther, S., & Meister, G. (2011). Small RNAs derived from longer non-coding RNAs. In *Biochimie* (Vol. 93, Issue 11). <https://doi.org/10.1016/j.biochi.2011.07.032>

- Scheuren, M., Möhner, J., & Zischler, H. (2023). R-loop landscape in mature human sperm: Regulatory and evolutionary implications. *Frontiers in Genetics, 14*.
<https://doi.org/10.3389/fgene.2023.1069871>
- 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).
<https://doi.org/10.1016/j.devcel.2018.06.023>
- Shi, J., Zhang, Y., Tan, D., Zhang, X., Yan, M., Zhang, Y., Franklin, R., Shahbazi, M., Mackinlay, K., Liu, S., Kuhle, B., James, E. R., Zhang, L., Qu, Y., Zhai, Q., Zhao, W., Zhao, L., Zhou, C., Gu, W., ... Chen, Q. (2021). PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nature Cell Biology, 23*(4).
<https://doi.org/10.1038/s41556-021-00652-7>
- Shigematsu, M., Morichika, K., Kawamura, T., Honda, S., & Kirino, Y. (2019). Genome-wide identification of short 20,30-cyclic phosphate-containing RNAs and their regulation in aging. *PLoS Genetics, 15*(11). <https://doi.org/10.1371/journal.pgen.1008469>
- Simons, F. H. M., Rutjes, S. A., Van Venrooij, W. J., & Pruijn, G. J. M. (1996). The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. *RNA, 2*(3).
- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., deMassy, B., Mikkelsen, T. S., & Kaessmann, H. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Reports, 3*(6).
<https://doi.org/10.1016/j.celrep.2013.05.031>
- Vojtech, L., Woo, S., Hughes, S., Levy, C., Ballweber, L., Sauteraud, R. P., Strobl, J., Westerberg, K., Gottardo, R., Tewari, M., & Hladik, F. (2014). Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Research, 42*(11). <https://doi.org/10.1093/nar/gku347>
- Volders, P. J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., & Vandesompele, J. (2019). Lncipedia 5: Towards a reference set of human long non-coding rnas. *Nucleic Acids Research, 47*(D1). <https://doi.org/10.1093/nar/gky1031>
- Wolin, S. L., & Cedervall, T. (2002). The La protein. In *Annual Review of Biochemistry* (Vol. 71). <https://doi.org/10.1146/annurev.biochem.71.090501.150003>

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S. Y., Keefe, D. L., Alukal, J. P., Boeke, J. D., & Yanai, I. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*, *180*(2).
<https://doi.org/10.1016/j.cell.2019.12.015>

Yang, Q., Li, R., Lyu, Q., Hou, L., Liu, Z., Sun, Q., Liu, M., Shi, H., Xu, B., Yin, M., Yan, Z., Huang, Y., Liu, M., Li, Y., & Wu, L. (2019). Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. *Nature Communications*, *10*(1).
<https://doi.org/10.1038/s41467-019-11312-8>

10 Supplementary Material

10.1 Figures

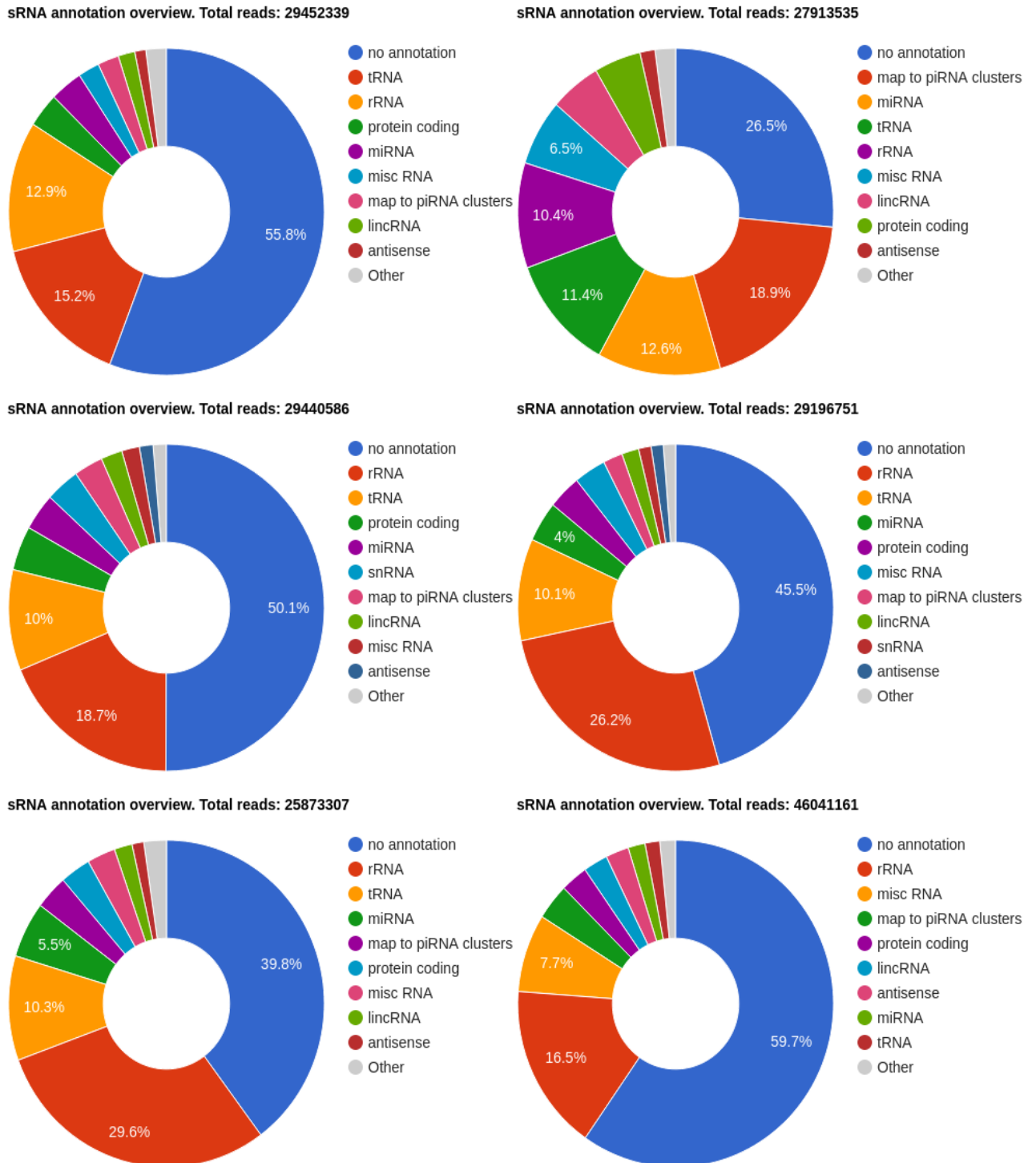


Figure 1: Donut charts of the UNITAS annotation for the six sperm head samples

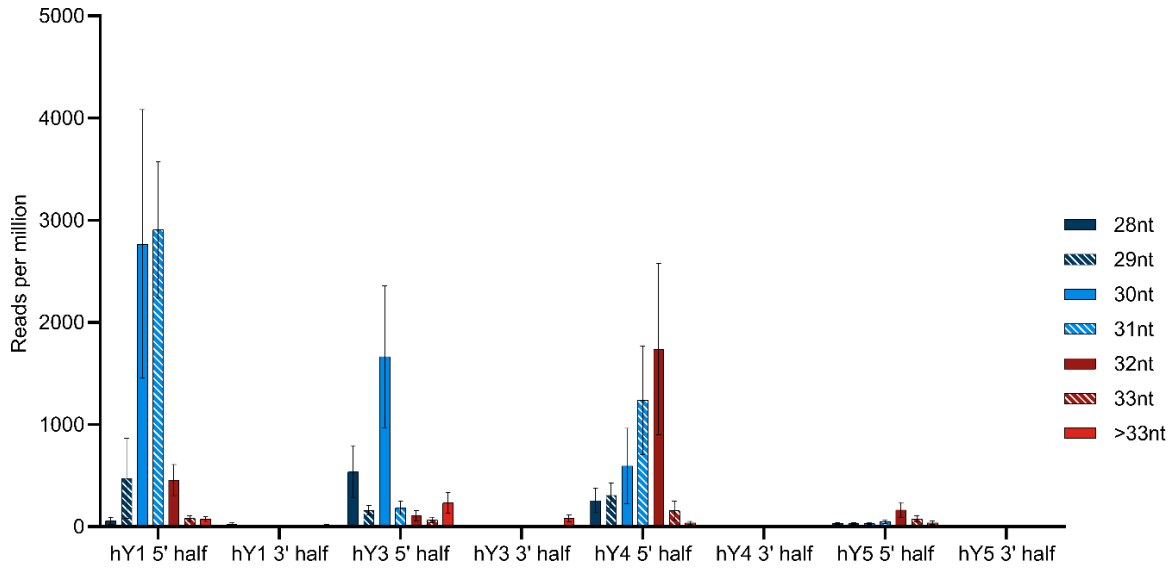


Figure 2: Hit numbers for the 5' and 3' parts of the four Y RNA homologs are shown in mean values and given in reads per million alignable reads. Lengths of the YsRNA are given in different colors as indicated on the right. Dots indicate individual hit numbers and bars show the standard error of mean.

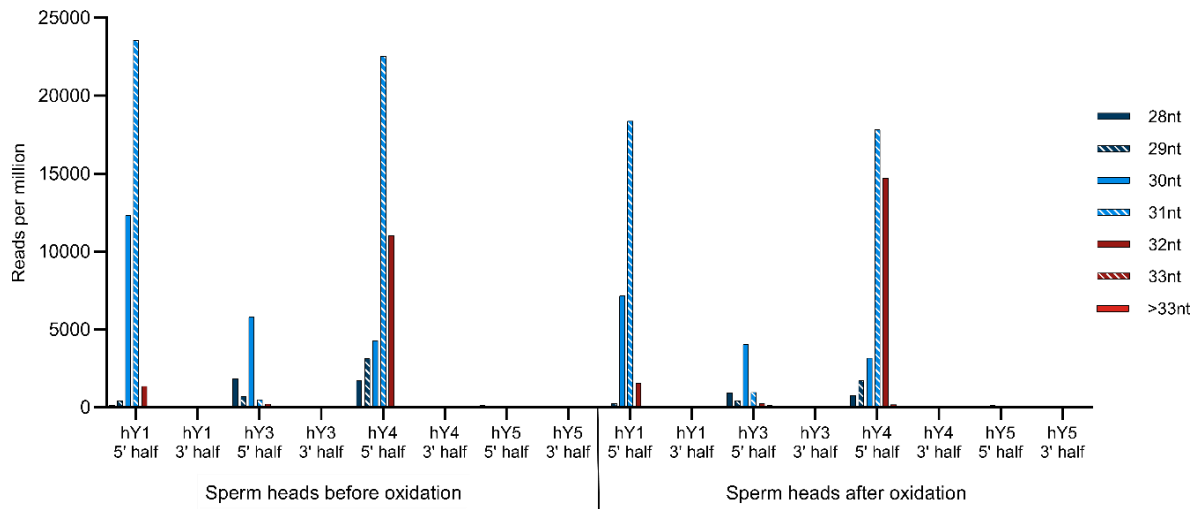


Figure 3: Length distribution of YsRNAs in sperm heads before and after oxidation in reads per million alignable reads.

10.2 Tables

Table 1: Information about the used SRA data

BioProject	SRA accession numbers	Description	Author/lab
PRJNA376426	SRR8329699, SRR8329700, SRR8329701	small RNA-seq of human oocyte before NaIO ₄ oxidization	Yang et al., 2019
PRJNA376426	SRR8329702, SRR8329703, SRR8329704	small RNA-seq of human oocyte after NaIO ₄ oxidization	Yang et al., 2019
PRJNA242348	SRR1200712, SRR1200711, SRR1200710, SRR1200709, SRR1200708, SRR1200707	20-40 nucleotide fraction	Vojtech et al., 2014
PRJNA242348	SRR1200706, SRR1200705, SRR1200704, SRR1200703, SRR1200702, SRR1200701	40-100 nucleotide fraction	Vojtech et al., 2014
PRJNA821911	SRR18575306	Cauda, single, normal	Gong et al., 2022
PRJNA821911	SRR18575307, SRR18575308, SRR18575311	Caput, single, normal	Gong et al., 2022
PRJNA821911	SRR18575309	Corpus, single, normal	Gong et al., 2022
PRJNA890147	SRR21902579	transcriptomic	Scheuren et al., 2023

2.1.1 Reviewers comments

This paper is still in revision, but I received first reviewer's comments. A short summary of the reviewer's comments that will be answered/implemented in a revised version of the manuscript can be formulated as follows:

Addressing different factors that might have an impact on the YsRNA distribution the age range as well as the ethnicity will be specified in more detail. As Y RNAs are not widely researched on, their structure and known functionalities will be specified further to enable a wider appreciation for my research. The validation of my findings in terms of their potential epigenetic inheritance will be discussed in more detail.

2.2 Epigenetic role of sperm head chromatin and relevance for inheritance of epigenetic information carriers

During spermatogenesis the transcriptomes attributable to the different developmental stages of sperm are characterized by extreme dynamics. Starting from a virtually complete expression of the whole genome (Soumillon et al., 2013), mRNA is massively eliminated during late spermiogenesis (Gou et al., 2014). Apparently, these dynamics also hold for smallRNAs in that their relative abundances are specifically changed during e.g. post-testicular maturation of sperm. A prominent example are piRNAs, which are almost completely absent from ejaculated sperm. Instead, a small amount of sncRNAs comprised primarily of tRNA cleavage fragments together with a smaller population of microRNAs (Peng et al., 2012) can be traced.

My own results, as obtained for mature sperm heads and presented in Chapter 2.1 (see supplemental Figure 1), point towards an interindividual biological variation in the composition of smallRNA as uncovered by UNITAS annotation. However, major trends can be observed in the relative abundances of different smallRNA classes in sperm heads. To get an idea about the dynamics of the smallRNA pool from the beginning of spermatogenesis to the mature sperm head, I analysed the smallRNA pool from human testis and compared the UNITAS annotations to another proband's sperm head profile. The following figures depict the respective results for human sperm heads (left) and testis (right).

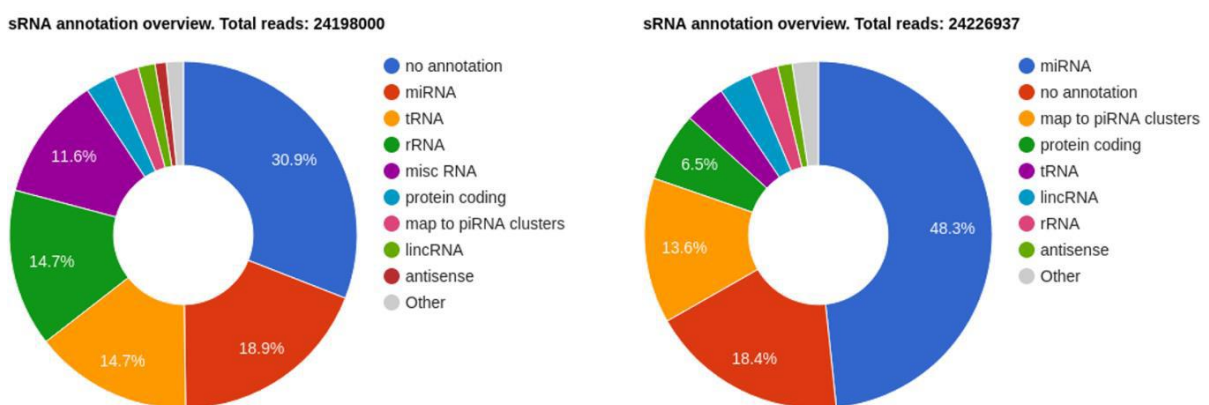


Figure 5: UNITAS annotation of testis (left) and sperm head (right) smallRNA. The donut charts show the annotation results of testis and sperm head smallRNA from Chapter 2.1.

UNITAS annotation routine screens different databases with smallRNA NGS data. These databases specify the different classes, which can be shortly characterized in the following:

Micro RNA

MicroRNAs (miRNAs) are a class of single-stranded non-coding RNAs with a length around 22 nt. They are considered to regulate the cleavage of target mRNAs post-transcriptionally or repress their translation (Bartel, 2004). miRNAs are transcribed by polymerases II and III (Pol II and Pol III) through cleavage of primary miRNA (pri-miRNA) and bind to proteins to form the RNA-induced silencing complex (RISC). miRNAs function here as a guide by base-pairing with the target mRNA to negatively regulate their expression by either cleavage of the target mRNA followed by degradation or by inhibiting their translation (MacFarlane & R. Murphy, 2010).

Transfer RNA

Transfer RNAs (tRNAs) are another class of non-coding RNAs with a size ranging from 70 nt to 90 nt (Lyons et al., 2018). They are transcribed by Pol III (Arimbasseri & Maraia, 2016) and build a cloverleaf structure that consists of a 3' acceptor site, a 5' terminal phosphate, a D arm, an anticodon arm and a T arm. Their main function is to translate messenger RNA (mRNA) into proteins by carrying amino acids on their 3' acceptor site to a ribosome complex (Raina & Ibba, 2014). Besides their role in protein synthesis, they take part in various cellular processes, including metabolism and cell death. tRNA fragments (tsRNAs) are generated from precursor or mature tRNAs by endonuclease cleavage (Zhu et al., 2018). This cleavage leads to two types of tsRNAs: tRNA halves and tRNA-derived fragments (tRFs). tRNA halves are generated by cleavage in the anticodon loop of mature tRNA and can be divided into a 5'-tRNA half and a 3'-tRNA half. The typical length of these tRNA halves is 31 – 40 nt. tRFs are shorter fragments with a size of 14 – 30 nt and can be divided into three subclasses: tRF5, tRF3 and tRF1 (Shen et al., 2018). tRF5s are generated by a cleavage in the D-loop or stem region between the D-loop and anticodon loop and can be further divided into three subclasses depending on their size. tRF3s are generated from the 3' end of mature tRNAs through cleavage in the T-loop. They can be divided into two subclasses. tRF1s are generated from the 3' end of primary tRNA transcripts (Kumar et al., 2016).

Ribosomal RNA

Another class of RNA that is also involved in protein synthesis is ribosomal RNA (rRNA). rRNAs form together with proteins ribosomes, which play a crucial role in protein synthesis. In eukaryotes four rRNAs are found: 28S rRNA, 18S rRNA, 5.8S rRNA and 5S rRNA. 28S rRNA, 18S rRNA and 5.8S rRNA are transcribed by Pol I and are the processing products of the 47S precursor transcript, while 5S rRNA is transcribed by Pol III (Stults et al., 2008; Tafforeau et al., 2013). The 28S rRNA, 5.8S rRNA and 5S rRNA form the 60S subunit of the ribosome, while the 18S rRNA is part of the 40S subunit. Together, they form the ribosome. rRNAs are extensively modified during transcription and maturation, which stabilises the structure of the ribosome and facilitates the efficient and accurate synthesis of proteins. The most abundant modifications here are 2'-O-methylation and isomerisation of uridine to pseudouridine (Sloan et al., 2017).

PIWI-interacting RNA

PIWI-interacting RNAs (piRNAs) are a class of small non-coding RNAs with a typical length of 24-31 nt that bind to PIWI proteins. They are found in germline cells and are thought to be involved in germline maintenance and transposon silencing (Iwasaki et al., 2015). piRNAs typically possess a 2'-O-methylation at their 3' end. The majority of piRNAs derive from large genomic loci called piRNA cluster (Brennecke et al., 2007). piRNAs are generated by two pathways: the primary processing pathway, producing primary piRNAs, and the ping-pong pathway, generating secondary piRNAs (Iwasaki et al., 2015).

Long intergenic non-coding RNA

Long intergenic non-coding RNAs (lincRNAs) are a class of non-coding RNAs with a size of more than 200 nt, that are transcribed exclusively from intergenic regions, differentiating them from other long non-coding RNAs (lncRNAs) (Ransohoff et al., 2018). Most lincRNAs are cell-type specific and found in the nucleus (Cabili et al., 2015). lincRNAs are associated with many different functions such as chromatin remodelling, genome architecture, RNA stability and transcription regulation (Ransohoff et al., 2018).

The ncRNA profile in human sperm heads differ greatly from that in testis, aligning to the findings of Peng et al. e.g. regarding piRNA dynamics in mice. Summarising the UNITAS annotation, miRNAs and piRNAs are massively eliminated during spermiogenesis, while tRNAs and rRNAs can be traced in much higher frequency in mature sperm cells.

Interestingly, the fraction of protein coding small RNAs does not show such dramatic differences. Therefore, a more detailed comparative analysis was carried out using FEATURECOUNTS and the hg38 gtf file obtained from HISAT2. FEATURECOUNTS counts mapped reads for genomic features depending on the gtf (gene transfer format) input and outputs numbers of reads assigned to those features. Here, the sperm head and testis bam files generated by HISAT2 analysis were mapped to the human reference genome hg38. The results were corrected for the overall alignment rate and given as reads per million alignable reads. The following graphs show the comparison between sperm heads and testis for the top 100 hits sorted for maximum abundance in testis (upper graph) and sperm heads (lower graph).

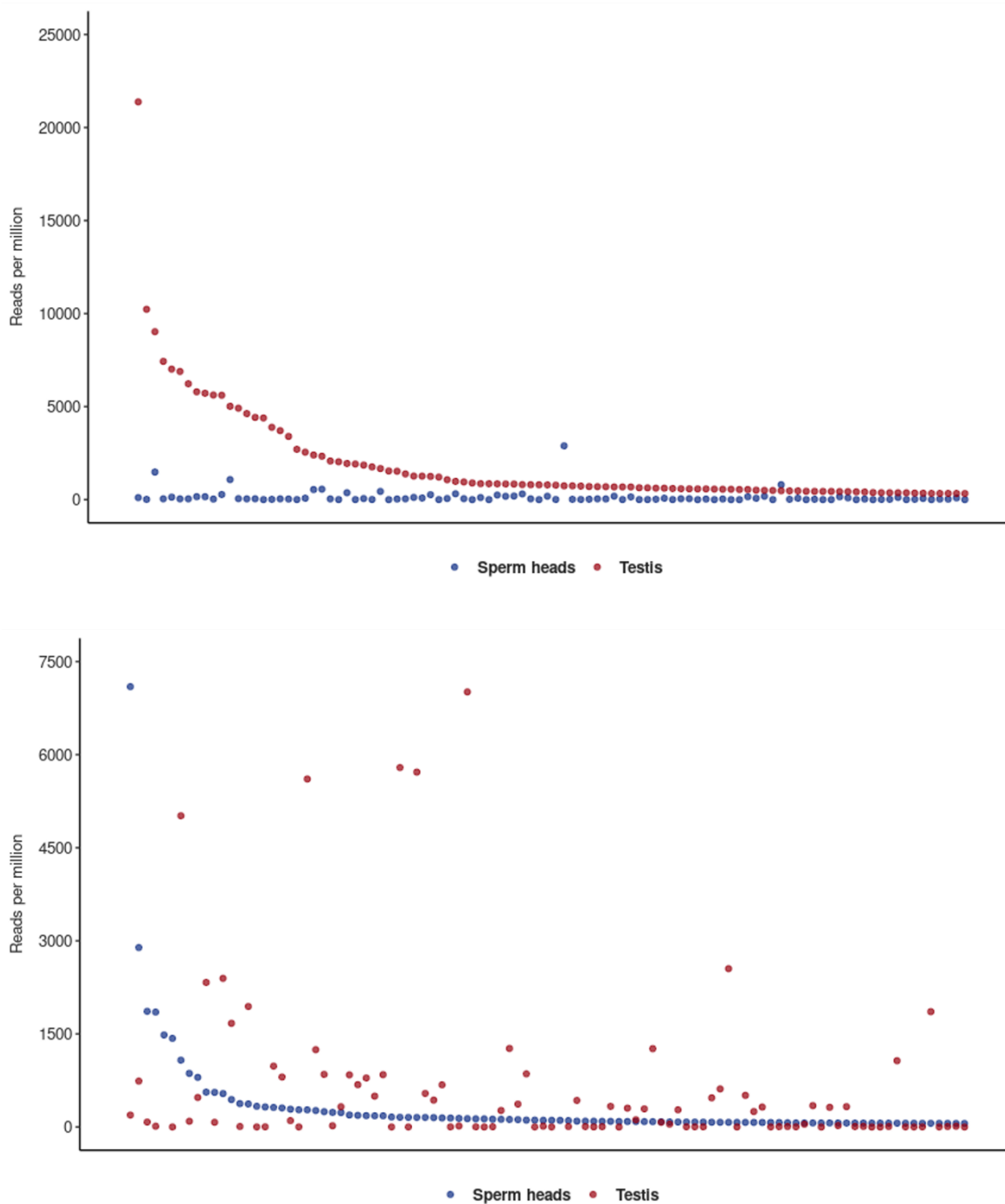


Figure 6: Illustration of the 100 most abundant genomic features in testis (upper graph) and sperm heads (lower graph). The upper graph shows the 100 most abundant genomic features in testis sorted from highest to lowest abundance and the corresponding amount of those features in sperm heads. The lower graph shows the 100 most abundant genomic features in sperm heads sorted from highest abundance to lowest abundance together with the corresponding abundance of those transcripts in testis. Sperm head transcripts are shown in blue and testis transcripts are shown in red.

The upper graph shows the 100 most abundant genomic features in testis sorted from the highest to lowest abundance (red dots) together with the abundance of the corresponding genomic features in sperm heads (blue dots). The lower graph shows the 100 most abundant genomic features in sperm heads sorted from the highest to lowest abundance as blue dots and the abundance of the corresponding genomic features in testis as red dots. Looking at the two graphs, it is evident that the transcripts of sperm heads differ greatly from that of testis. The upper graph shows that the abundance of the 100 most abundant testis transcriptomes in sperm heads is smaller, with the exception of two transcripts (microRNA 375 and microRNA 149). The graph sorted for the most abundant transcripts in sperm heads shows a completely different pattern. Around half of the transcripts show a higher abundance in testis (46 out of 100), while the other half (54 out of 100) show a lesser abundance. It is evident that the general hypothesis of an overall reduced transcript abundance in sperm heads compared to testis is reproduced also for the smallRNA fractions. A more detailed focus on the 100 most abundant sperm head transcripts reveals Y RNA 1 (see Chapter 2.1 section 3.2 for details) as the most abundant RNA. Out of the 100 most abundant sperm head transcripts, 38 show an at least 10-fold higher abundance in sperm heads as compared to the testis smallRNA fractions. Y RNA 1 was found in sperm heads with an almost 40-times higher abundance than in testis. Interestingly, the transcript that shows the greatest difference between sperm heads and testis is protamine 1 (PRM1) with an abundance more than 4000 times higher in sperm heads than in testis. Protamine 2 (PRM2) and Transition Protein 1 (TNP1) were also found in the 100 most abundant transcripts in sperm heads and showed a 600-times and 2000-times higher abundance compared to the testis transcripts. In testis on the other hand, the majority of the 100 most abundant transcripts belong to the class of non-coding RNAs, i.e. miRNAs. More specifically, 19 miRNAs were found in the top 100 sperm head transcripts together with 13 lincRNAs and 5 lncRNAs. In the testis transcripts 59 miRNAs and 10 lincRNAs could be traced. These miRNA-related findings correlate with the UNITAS annotation and confirm the decrease of miRNA in sperm heads compared to testis. Moreover, in the UNITAS annotation, lincRNAs decrease slightly from 3.1 % in testis to 1.6 % in sperm heads, whereas the diversity of sperm head lncRNA is greater.

A high abundance of mitochondrial ribosomal RNA was detected in the top 100 sperm head transcripts, which was counterintuitive due to proper functionality of the sperm head preparation and the absence of an intact translation machinery in sperm heads. However, ribosomal RNAs are commonly found in sperm heads and Villegas et al. (2002) proposed a process of translocation of transcripts from the mitochondria into the nucleus of

spermatogenic cells and possibly associated with meiotic chromosomes. Thus, sperm head ribosomal RNA transcripts are likely leftovers from earlier stages of spermatogenesis.

To elucidate the possible functional consequences of the different top 100 transcript profiles, a pathway enrichment analysis was done applying the publicly available METASCAPE tool (Y. Zhou et al., 2019). The results are depicted below, with testis METASCAPE outputs above and the output for sperm heads below.

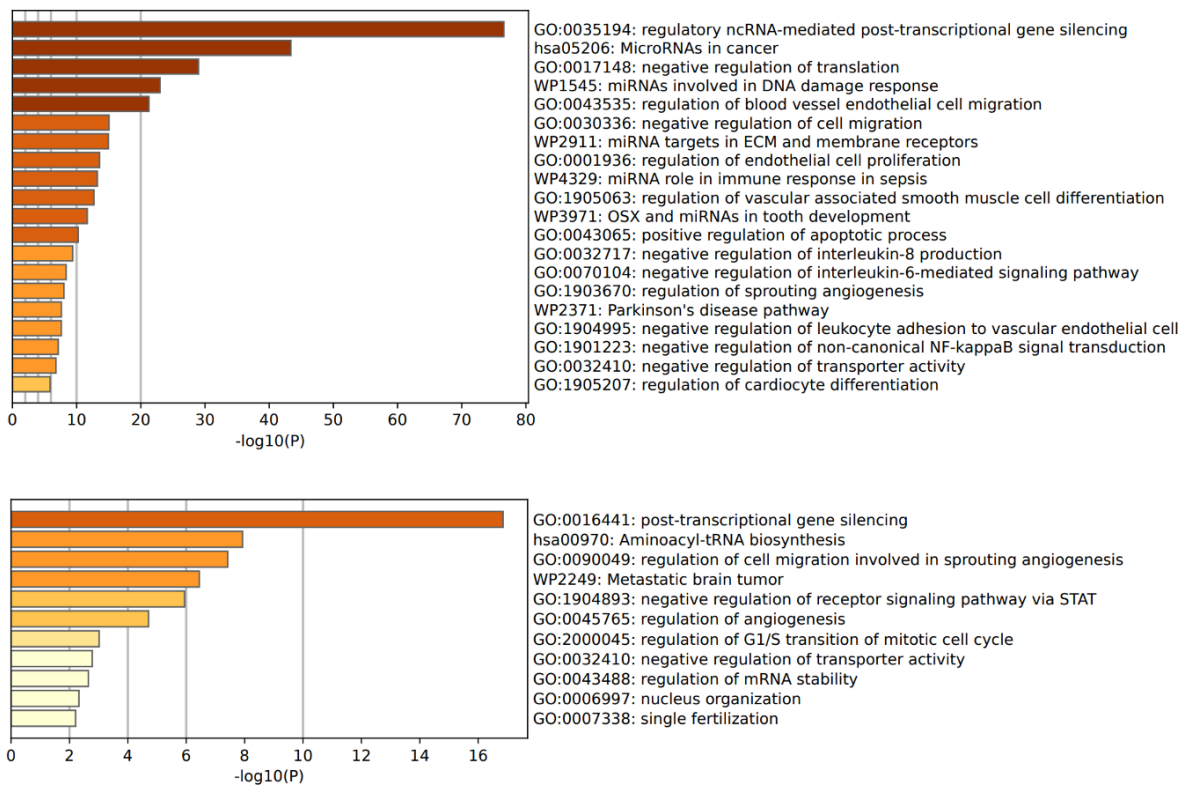


Figure 7: Gene Ontology annotation of the top 100 transcripts in testis (upper chart) and sperm heads (lower chart).

Because the Gene Ontology (GO) patterns differ remarkably between testis and sperm heads, it can be hypothesized that the transcript profiles of sperm heads cannot be fully explained by a passive reduction in transcript abundance compared with earlier stages of spermatogenesis. It is important to note at this point, that the findings in sperm cells were usually obtained from whole sperm cells, mostly isolated after applying swim-up protocols. This study concentrates on sperm heads as precursors of the male pronucleus in the zygote, thus the question arising from these findings is twofold. First, can RNA information potentially be delivered via the male pronucleus into the zygote? Secondly, to what extent

do the transcript dynamics result from elimination of RNAs or are these RNAs actively transcribed in sperm heads? Active transcription would require the accessibility of specific genomic regions.

During spermatogenesis most histones are replaced firstly by transition proteins and thereafter by protamines. These basic, arginine-rich proteins differ from histones in their enrichment in lysine and cysteine residues. The high arginine content results in a partial positive charge that promotes DNA binding and neutralises the negative charge along the phosphodiester backbone of DNA, allowing neighbouring DNA molecules to be packed more tightly, resulting in a 10 times more compacted genome. This condensation leads to an inactivation of the genomic regions. In humans, two nucleoprotamines are found, namely protamine 1 (PRM1) and protamine 2 (PRM2), which are mainly found in testis. In testis, about 50 % of the basic nuclear proteins are histones, while sperm only contains about 15 % of histones (Tanphaichitr et al., 1978).

The first aim was therefore to find out whether the observed ncRNAs in sperm heads are transcribed from genomic regions bound by histones or if they are associated with protamines, which would lead to the assumption that those RNAs are generated in earlier stages of spermatogenesis or that they are generated outside the sperm cells and integrated during late stages of spermiogenesis, like YsRNAs, Chromatin Immunoprecipitation was performed and the genomic regions binding to H2B, PRM1 and PRM2 were compared to regions transcribing different RNAs.

The second intention was to uncover possible non-coding RNA-Chromatin interactions in the sperm head. The possibly associated RNAs would constitute another fraction of RNAs that could serve as epigenetic information carriers that are delivered into the zygote via the male pronucleus. This “chromatin as carrier-effect” is likely a possible scenario, since protamines are intrinsically disordered proteins, with PRM2 being 100 % disordered and about 60 % of PRM1 being disordered. Histone H2B, on the other hand, has two intrinsically disordered regions summing up to 42 % of the complete protein. Intrinsically disordered proteins have the ability to bind RNA and thus function as carriers for RNAs. Therefore, RNA immunoprecipitation with H2B, PRM1 and PRM2 were performed to elucidate if the potentially bound RNAs do exhibit a certain profile which is different from the chromatin protein component.

2.2.1 Material & Methods

2.2.1.1 Local BLAST of sperm cell transcriptomic data with protamine sequences

Sperm head and testis protein-coding data obtained from UNITAS annotation (Chapter 2.1) was blasted against the nucleotide sequences of protamine 1 and protamine 2 to find out if these protein transcripts are expressed in early stages of spermatogenesis and/or mature sperm. To get an even more precise understanding about PRM1 and PRM2 mRNA abundance during spermatogenesis transcriptomic data of different developmental stages during spermatogenesis (Jan et al., 2017; BioProject PRJNA310976) was obtained from SRA. The SRA data was adapter-trimmed using TRIM GALORE and HISAT2-mapped to get the overall alignment rate. The clean sequences were used for the local BLAST analysis. Nucleotide sequences of PRM1 and PRM2 were obtained from the NCBI Consensus CDS protein set (CCDS) database (PRM1 CCDS ID: 10547.1, PRM2 CCDS ID: 42118.1) and used to build a local database. SRA data was blasted against this database and compared to the BLAST-results of the protein-coding-associated data of sperm heads and testis obtained from the UNITAS annotation.

2.2.1.2 Sperm heads preparation

Semen samples were collected from four volunteers. Donors were asked to abstain from sexual activity for two to three days. The entire ejaculates gained by masturbation were collected in a sterile 50 ml tube. The samples were used within one hour after ejaculation. Semen samples were evenly distributed to 2 ml reaction tubes and centrifuged at room temperature at 16,000 x g for 5 minutes. The supernatants were discarded and the pellets were resuspended in 700 μ l TEN (20 mM Tris(hydroxymethanol)aminomethane (TRIS), 20 mM Ethylenediaminetetraacetic acid (EDTA), 200 mM NaCl, [pH 8.0]), 400 μ l HPLC Gradient Grade water and 300 μ l 10 % SDS, incubated for 5 minutes at room temperature and afterwards centrifuged at room temperature at 16,000 x g for 5 minutes. The supernatants were removed and the washing steps were repeated two more times. After the final wash the pellets containing the intact sperm heads were stored at -80 °C until further usage.

2.2.1.3 Dot Blots

One sperm head sample from one person was resuspended in HPLC water. This suspension as well as a complete ejaculate sample from another individual were sonicated using the Covaris E220 Focused-Ultrasonicator. The Peak Incident Power was set to 140, the Duty Factor to 10 %, Cycle of Burst to 200, the temperature to 20 °C and the duration to 120 seconds. Afterwards, the protein concentrations of the samples were measured using the Qubit® 2.0 Fluorometer together with the Qubit Protein Broad Range (BR) Assay-Kit. For each sample and standard 199 µl Working Solution and 1 µl Qubit Protein Reaction Buffer were mixed together. 199 µl of this solution was mixed with 1 µl sample, vortexed and incubated for 15 minutes in the dark. To prepare the standards 190 µl of the working solution mix was mixed with 10 µl of each standard, vortexed and incubated for 15 minutes in the dark. To calibrate the Fluorometer the three standard samples were measured first in the correct order. Afterwards, the samples were measured. The same amount of protein was used for each sample. The samples were filled up with HPLC water to reach the same volume and concentration. The samples were pipetted on a dry neutral nylon membrane (Amersham #RPN303 N) and left to dry. As a control for the secondary antibody (Cell Signaling Technology; #7056S and #7054S) 1 µl of the primary antibody (PRM1: Abnova, #H00005619-B01P; PRM2: Cloud-Clone Corp., #PAH307Hu01) was pipetted on the membrane. To bind the proteins to the membrane the membrane was UV-crosslinked using the UV Stratalinker® 1800 from Stratagene and 1200 mJ UV light. Afterwards, the membrane was blocked overnight on a rotor at 4 °C using a 5 % blocking solution (1x PBS [137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄], 0.1 % Tween-20, 5 % nonfat dried milk). The blocking solution was discarded and 10 µl primary antibody was dissolved in 10 ml blocking solution and put on the membrane. The membrane was incubated for two hours on a rotor at room temperature. The antibody solution was discarded and the membrane was washed thrice with 1x PBST (1x PBS, 0.1 % Tween-20) for 10 minutes on a shaker. The solution was discarded, and the membrane was incubated in blocking solution for 30 minutes on a shaker at room temperature. Afterwards, the membrane was incubated in 10 ml blocking solution containing 10 µl of the corresponding secondary antibody for one hour at room temperature on a shaker. The membrane was washed twice with 1x PBST and the solution was discarded. 10 µl CSPD solution was dissolved in 990 µl detection buffer (0.1 M Tris-HCl, 1 M NaCl, pH 9.5) and put on the membrane. The membrane was wrapped in plastic foil and aluminium foil and incubated for 15 minutes at 37 °C. Afterwards, the signals were analysed using the ChemoCam Imager from INTAS.

2.2.1.4 Western blot

One sperm head sample was dissolved in 48 μ l HPLC water and 2 μ l DTT. For the western blotting an SDS-PAGE gel electrophoresis was performed. Firstly, a separation gel was created by mixing 4.42 ml HPLC water, 2.62 ml 1.5 M TRIS [pH 8.8], 105 μ l 10 % Sodium Dodecyl Sulfate (SDS), 3.28 ml Acrylamide: Bisacrylamide 19:1 (40 %), 52.5 μ l 10 % ammonium persulfate (APS) and 7.5 μ l tetramethylethylenediamine (TEMED). This gel was put in an electrophoresis chamber, covered with isopropanol and put in a 65 °C incubator to polymerise. The collection gel was prepared by mixing 2.38 ml HPLC water with 940 μ l 1.5 M TRIS [pH 8.8], 38 μ l 10 % SDS, 375 μ l Acrylamide-Bisacrylamide 19:1 (40 %), 18.8 μ l 10 % APS and 7.5 μ l TEMED. The isopropanol was removed, the collection gel was put on top of the separation gel and let polymerise at 65 °C. 1x SDS-running-buffer was created by dissolving 14.4 g Glycine, 3.02 g TRIS and 0.1 g SDS in one litre HPLC water. The gel electrophoresis chamber was filled up with the running buffer. 15 μ l of the sperm head sample was mixed with 15 μ l Laemmli buffer and placed in the pockets. 5 μ l Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standard (Bio-Rad #1610375) were used as ladder. The gel was run at 100 V until the sample reached the separation gel. After that the gel run at 200 V for 45 minutes. A Polyvinylidene fluoride (PVDF) membrane was cut to the size of the gel and one minute in methanol activated. The gel, membrane and the fleece pieces used for the transfer were incubated for 10 minutes in transfer buffer (3.03 g TRIS, 14.4 g Glycine, 2.5 ml 10 % SDS and 200 ml methanol dissolved in 800 ml HPLC water). One fleece piece was placed on the negative side of the transfer cassette, followed by a few pieces of filter paper, the PVDF membrane, the gel, more pieces of filter paper and the remaining fleece piece. The transfer cassette was closed and placed into the blotting apparatus. The blotting apparatus was filled up with the transfer buffer. The transfer was performed at 100 V for one hour. After the transfer the membrane was blocked in a 1 % nonfat dried milk-blocking solution (1x PBS [137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄], 0.1 % Tween-20, 1 % nonfat dried milk). 10 ml blocking solution was mixed with 10 μ l Monoclonal Antibody to Histone H2B (D2H6) (Cell Signaling Technology #12364). The blocking solution was discarded, and the membrane was incubated in the antibody solution over night at 4°C on a rotor. The antibody solution was discarded and the membrane was washed thrice with 1x PBST for 5 minutes on a shaker. The solution was discarded. Afterwards, the membrane was incubated in 15 ml blocking solution containing 5 μ l of the secondary Goat anti-Rabbit IgG antibody (Abcam #ab6721) for one

hour at room temperature on a shaker. The membrane was washed twice with 1x PBST for 5 minutes and the solution was discarded. Two chemiluminescence solutions (ECL) were prepared that react with one another when combining them. The first solution contained 200 μ l Luminol dissolved in Dimethyl sulfoxide (DMSO), 88 μ l p-Coumaric acid dissolved in DMSO (0.15 g p-Coumaric acid dissolved in 10 ml DMSO), 3 ml 1 M Tris-HCl [pH 8.5] and 17.2 ml HPLC water. The second solution contained 12 μ l 30 % Hydrogen peroxide (H_2O_2), 2 ml 1 M Tris-HCl [pH 8.5] and 18 ml HPLC water. The solutions were stored in the dark. The solutions were put on the membrane simultaneously and the membrane was incubated for two minutes in the dark. The signal detection was performed using the ChemiDoc MP (Bio-Rad) and the corresponding ImageLab 4.1 program.

2.2.1.5 Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed using the antibodies against H2B (Cell Signaling Technology; #12364), PRM1 (Abnova; #H00005619-B01P) and PRM2 (Cloud-Clone Corp.; #PAH307Hu01). Sperm head pellets from 4 individuals were pooled for each individual, dissolved in HPLC water and sonicated using the Covaris E220 Focused-Ultrasonicator. The Peak Incident Power was set to 140, the Duty Factor to 10 %, Cycle of Burst to 200, the temperature to 20 °C and the duration to 120 seconds. After the sonication the samples were put in 1.5 ml reaction tubes and quenched using a SpeedVac to reach a volume of ~50 μ l. The samples were incubated for 10 minutes at 95 °C and afterwards placed directly on ice. 150 μ l incubation buffer (100 μ l 5 M NaCl, 200 μ l 1 M TRIS [pH 8] and 10 μ l 0.5 M EDTA filled up with HPLC water to 10 ml) and 10 μ l of the chosen antibody were added to the samples and incubated overnight at 4 °C on a rotor. For each sample 100 μ l SureBeadsTM Protein G Magnetic beads (Bio-Rad; #161-4023) were washed three times with incubation buffer. After the final wash the magnetic beads were dissolved in the same volume of incubation buffer. 100 μ l magnetic beads were added to each sample and the samples were incubated on a rotor for two hours at 4 °C. The samples were spun down briefly to remove any liquid in the cap and placed on a magnetic rack. The supernatant was discarded. 400 μ l wash buffer A (500 μ l 1 M TRIS [pH 8], 200 μ l 0.5 M EDTA and 100 μ l 5 M NaCl filled up with HPLC water to a final volume of 10 ml) was used to resuspend the samples. The samples were placed on a magnetic rack and the supernatant was discarded. The beads were dissolved in the wash buffer B (500 μ l 1 M TRIS [pH 8], 200 μ l 0.5 M EDTA and 200 μ l 5 M NaCl filled up with HPLC water to 10 ml), placed on a magnetic rack and the supernatant

was discarded. Afterwards, the samples were washed two times with wash buffer C (500 μ l 1 M TRIS [pH 8], 200 μ l 0.5 M EDTA and 300 μ l 5 M NaCl filled up with HPLC water to 10 ml) the same way as before. To isolate the DNA the QIAamp® DNA Mini Kit (Qiagen; #51306) was used. The samples were mixed with 250 μ l AL buffer and placed on magnetic rack. The supernatant was transferred in a fresh 1.5 ml reaction tube and 25 μ l Proteinase K was added. The samples were incubated for 10 minutes at 56 °C and afterwards mixed with 250 μ l 100 % ice-cold ethanol and placed on a Qiagen column in a collection tube. The samples were centrifuged 1 minute at 10,000 x g at room temperature. The flow-through was discarded and 500 μ l AW1 wash buffer was put on the column. Again, the samples were centrifuged for 1 minute at 10,000 x g at room temperature and the flow-through was discarded. 500 μ l AW2 buffer was put on the column and the samples were centrifuged for 3 minutes at 18,000 x g at room temperature. The column was placed on a new collection tube. 500 μ l HPLC water was added to the column and the samples were incubated for 5 minutes at room temperature. Afterwards, the samples were centrifuged for 1 minute at 10,000 x g at room temperature. The eluate was transferred to a fresh 1.5 ml reaction tube. 100 μ l HPLC water was put on the column and incubated and centrifuged as before. The eluate was added to the 200 μ l from before. The samples were quenched in a SpeedVac to a volume of 25 μ l. Library preparation and ChIP sequencing were performed by Novogene. Library preparation was done by fragmenting, repairing and dA-tailing the DNA. The A-tailed DNA fragments were ligated to sequencing adaptors. The final DNA library was obtained by size selection and PCR amplification. Sequencing was done using the Illumina NovaSeq platform.

2.2.1.5.1 Bioinformatic analyses by Novogene

Raw reads were quality-checked and trimmed. The trimmed data was quality-checked again. Afterwards, the clean reads were mapped to the reference genome. RseQC software was used to predict the fragment sizes of the mapping results. These predicted fragment sizes were used for peak calling. Peak calling was performed by mapping to the reference genome to obtain information about the protein-DNA binding sites. The software MACS2 was used to calculate the number of peaks, the peak width, its distribution and to identify the peak related genes. Novogene provides an Excel sheet with the count of summits and a file with the narrow peaks.

2.2.1.6 RNA Immunoprecipitation

Intrinsically disordered proteins are able to bind RNA. To find out which type of RNA binds to H2B, PRM1 and PRM2 an RNA Immunoprecipitation Sequencing (RIP-Seq) was performed using the ChIP-IT Magnetic Chromatin Immunoprecipitation Kit from Active Motif (#53024). At first, the sperm head pellets were dissolved in HPLC water and pooled for each individual. Afterwards, the samples were sonicated using the Covaris E220 Focused-Ultrasonicator. The Peak Incident Power was set to 140, the Duty Factor to 10 %, Cycle of Burst to 200, the temperature to 20 °C and the duration to 120 seconds. Afterwards, the samples were put in a fresh 1.5 ml reaction tube and filled up with DEPC water to equalise the sample volume between individuals. The sonicated samples were treated with DNase I. Therefore, 10 µl DNA Digestion Buffer were added to the samples followed by 10 µl DNase I. The samples were incubated for 20 minutes at 37 °C. To stop the reaction 10 µl 0.5 M EDTA was added and mixed well by inverting the tube. The samples were then centrifuged for 10 minutes at 15,000 rpm at 4 °C and the supernatant was transferred to a fresh 1.5 ml reaction tube. Before starting with the immunoprecipitation, the magnetic beads were washed twice with DEPC water. To do so, 25 µl for each immunoprecipitation reaction were put in a fresh RNase-free reaction tube. The tube was placed on a magnetic rack and the supernatant was removed using a pipette to measure the removed volume. The beads were dissolved in 400 µl DEPC water, placed on a magnetic rack and the water was removed. This washing step was repeated a second time. The same volume of DEPC water as the one that was removed in the beginning was added to the beads. 25 µl pre-washed Protein G Magnetic Beads, 10 µl RNA IP Buffer, 0.1 µl RNase Inhibitors and 1 µl Protease Inhibitor Cocktail were added to each sample. 1 µg of the used antibody was added to the samples. The samples were incubated overnight at 4 °C on a rotor. Afterwards, the tubes were spinned down briefly to collect any liquid from the inside of the cap and placed on a magnetic rack. The supernatant was carefully discarded. To wash the beads Complete RNA-ChIP Wash Buffer 1 and Complete RNA-ChIP Wash Buffer 2 were prepared by adding 0.25 µl RNase Inhibitor per ml of wash buffer. The beads were wash 4 times with Complete RNA-ChIP Wash Buffer 1. Therefore, 200 µl wash buffer 1 was added to the beads and the beads were dissolved. The beads were placed on a magnetic rack and the supernatant was discarded. After the fourth washing step using Complete RNA-ChIP Wash Buffer 1 the beads were washed with Complete RNA-ChIP Wash Buffer 2 two times the same way. After the final wash, as much supernatant as possible was carefully removed without disturbing the beads. After that, the Complete RNA-ChIP Elution Buffer was prepared by adding 0.1 µl RNase

Inhibitors to 100 μ l RNA-ChIP Elution Buffer for each sample. 100 μ l Complete RNA-ChIP Elution Buffer was added to the beads and the beads were resuspended thoroughly. The samples were incubated for 15 minutes at room temperature on a rotor. Afterwards, the samples were briefly spun down to collect any liquid inside the cap, placed on a magnetic rack and the supernatant was carefully removed into a fresh reaction tube. 2 μ l 5 M NaCl and 2 μ l Proteinase K were added to each sample. The samples were incubated at 42 °C for 1 hour to digest the proteins. Then, the samples were incubated for 1.5 hours at 65 °C to reverse cross-links. Afterwards, RNA was purified using TRIzol. To do so, ~150 μ l DEPC water was added to the samples to reach a volume of 250 μ l. 750 μ l TRIzol (Invitrogen; #AM9738) was mixed with the samples by pipetting up and down several times. The samples were incubated 5 minutes at room temperature. 200 μ l chloroform was added and the samples were mixed vigorously by pipetting up and down for at least 15 seconds and incubated for 15 minutes at room temperature. The samples were centrifuged for 10 minutes at 12,000 x g at 4 °C. The upper aqueous phase containing the RNA was carefully transferred into a fresh reaction tube. The same volume isopropanol and 1 μ l glycogen were added to the samples. The samples were mixed well, incubated 15 minutes at room temperature and centrifuged for 10 minutes at 12,000 x g at 4 °C. The supernatant was carefully removed without disturbing the pellet. 1 ml ice-cold 75 % ethanol was added to the pellets. The samples were centrifuged for 5 minutes at 7,500 x g at 4°C. The supernatant was carefully removed without disturbing the pellets. The pellets were left to air dry briefly and then resuspended in 20 μ l DEPC water. After that, the samples were treated with DNase I. Therefore, the samples were incubated for 15 minutes at 37 °C. Afterwards, 2 μ l 10X DNase I Reaction Buffer and 2 μ l DNase I were added to the samples. The samples were incubated for 25 minutes at 22 °C. 2 μ l Stop Solution was added to the samples and incubated for 10 minutes at 65 °C to heat inactivate the DNase I. To purify the samples RNA purification using TRIzol was performed a second time. For that, 500 μ l TRIzol was added to the samples. The samples were mixed well and incubated for 5 minutes at room temperature. 100 μ l chloroform was added and the samples were mixed well and incubated for 2 minutes at room temperature. The samples were centrifuged for 15 minutes at 12,000 x g at 4 °C. The upper aqueous phase was carefully transferred into a new reaction tube. The same volume of isopropanol was added to the samples and mixed well. 0.5 μ l glycogen was added to the samples. The samples were mixed well and incubated for 15 minutes at room temperature. Afterwards, the samples were centrifuged for 10 minutes at 12,000 x g at 4 °C. The supernatant was carefully discarded, 500 μ l 75 % ice-cold ethanol was added to the

samples and they were centrifuged for 5 minutes at 7,500 x g at 4 °C. The supernatant was carefully discarded, and the washing step was repeated two more times. After the final wash the supernatant was carefully removed, and the samples were left to air dry briefly. Afterwards, the pellets were dissolved in 20 µl RNase-free water. Library preparation and lncRNA-Seq were performed by Novogene using PE150 as sequencing strategy. For the library construction ribosomal RNA was removed from the total RNA and the remaining RNA was ethanol precipitated. The RNA was fragmented, and the first strand cDNA was synthesised using random hexamer primer. During the second cDNA synthesis, dUTPs were replaced by dTTPs in the reaction buffer. The fragments were end-repaired and A-tailed. Afterwards, adapter ligation, size selection, USER (Uracil-Specific Excision Reagent) enzyme digestion, amplification and purification were performed. The library was checked with Qubit and real-time PCR for quantifications and the bioanalyzer was used to detect the size distribution. Afterwards, sequencing was performed on the Illumina platform. The data was quality checked and the raw data was filtered.

2.2.1.7 Bioinformatic analysis

2.2.1.7.1 ChIP-Seq analyses

Following the Peak calling protocol as carried out by Novogene and outlined above, the obtained data was analysed using the software Hypergeometric Optimization of Motif EnRichment (HOMER; (Duttke et al., 2019)). The obtained excel sheets were used to generate bed files containing the genome coordinates (chromosome, start position and end position). Those bed files were used to annotate the peaks to the human reference genome hg38 to find out in which genomic region most peaks are and where motifs are under-/overrepresented compared to the reference genome. To do so, the `annotatePeaks.pl` -routine was started to perform peak annotation with the human genome hg38 as reference. The annotation of peaks/regions is done by determining the distance to the nearest transcription start site (TSS) and by determining the genomic annotation of the region occupied by the center of the peak/region.

Furthermore, bedtools INTERSECT was used to compare the ChIP peak data to peaks of different genomic regions to find overlaps. The genomic regions data used for this analysis were lncRNA data from LNCipedia (Volders et al., 2019;; version 5.2, hg38), piRNA data from piRNAdb (Piuco & Galante, 2021), sno/miRNA and tRNA data from the UCSC Table

Browser (Karolchik et al., 2004; GRCh38/hg38) and rRNA data from dashr2 (Kuksa et al., 2019); v2.0, hg38; downloaded and filtered for rRNA). All obtained ChIP datasets were compared to the genomic regions of the above-mentioned RNA classes to find overlapping regions. Additional analyses using the INTERVENE (Khan & Mathelier, 2017) Venn diagram module were performed to compare data where lncRNA and ChIP data overlap between individuals to find compute and visualise unique and overlapping regions.

2.2.1.7.2 RIP-Seq analyses

The obtained fastq files from Novogene were mapped to the human genome hg38 using HISAT2 (Kim et al., 2019; Version 2.2.1) to calculate the alignment rate. The bam files generated by HISAT2 were mapped against the lncRNA gene gtf file of the human reference genome hg38 obtained from LNCipedia (Volders et al., 2019; version 5.2) using FEATURECOUNTS READCOUNT (Liao et al., 2014). Thus, sequences were aligned to the matching lncRNA genes and the hits were counted. The obtained list of lncRNA genes with the corresponding number of hits were compared between protein samples of all individuals and between all protein samples of each individual.

2.2.2 Results

2.2.2.1 Transcriptional dynamics of the sperm chromatin components PRM1 and PRM2

To compare the early stages of spermatogenesis with the mature human sperm heads, smallRNA from human testis total RNA preparations were isolated and smallRNA-sequenced. Since all developmental stages of sperm before entering the epididymis are present in testis, this preparation of bulk RNA was used as a proxy of early spermatogenesis. To obtain meaningful comparisons with the sperm head smallRNA-Seq results, the identical conditions with respect to size enrichment and sequencing strategy were applied (Chapter 2.1 Section 3.4) as depicted in Figure 3.

Obviously, the fraction of protein coding small RNAs decreased significantly during spermatogenesis from 6.5 % in testis to 2.7 % in mature sperm heads.

To get a more precise view on the abundance of PRM1 and PRM2 mRNA in different stages of spermatogenesis, SRA data of different cell types during spermatogenesis (BioProject PRJNA310976) was downloaded and analysed. The SRA datasets were quality checked with TRIM GALORE and clean sequences retrieved. Subsequently, HISAT2 was carried out using the trimmed data to determine the overall alignment rate. The clean sequences were used for BLAST analysis using the protamine sequences as database.

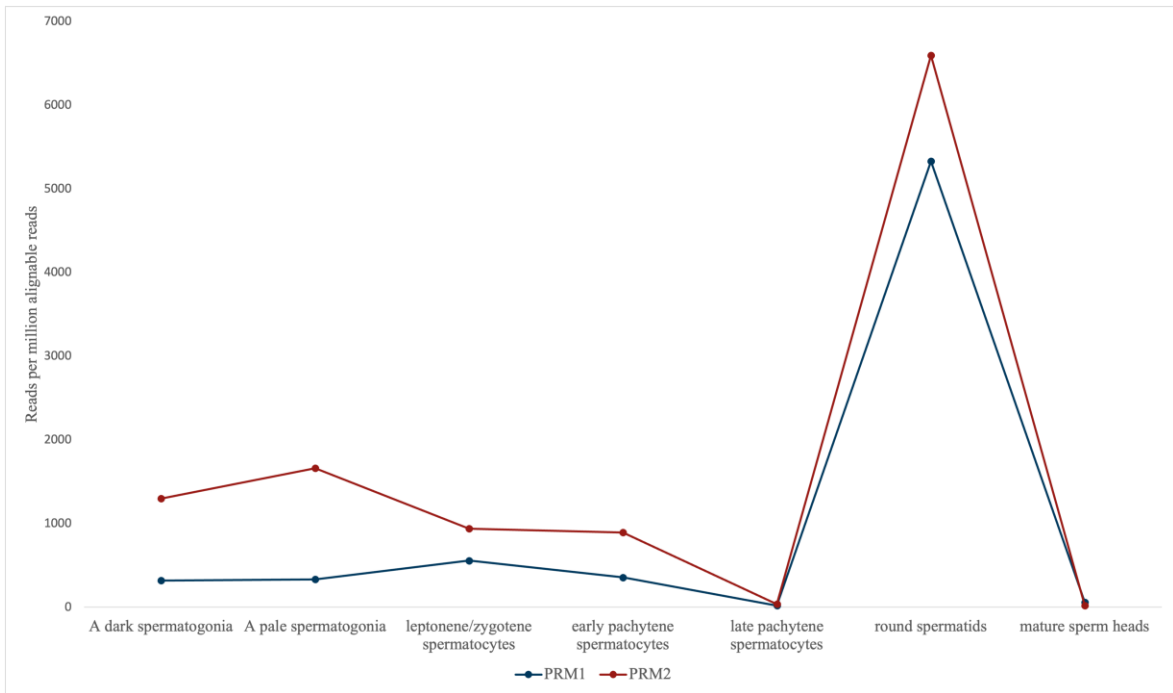


Figure 8: Amount of PRM1 and PRM2 mRNA in different cell types during spermatogenesis. The graph shows the amount of PRM1 and PRM2 mRNA in A_{dark} spermatogonia, A_{pale} spermatogonia, leptotene/zygotene spermatocytes, early pachytene spermatocytes, late pachytene spermatocytes, round spermatids and mature sperm heads in reads per million alignable reads. PRM1 mRNA is shown in blue and PRM2 mRNA is shown in red.

Figure 8 shows the amount of PRM1 and PRM2 mRNA in different cell types during spermatogenesis (BioProject PRJNA310976) and in sperm heads of mature sperm in reads per million alignable reads (rpm). The average amount of PRM1 mRNA in cell type A dark spermatogonia is 317.68 rpm, in cell type A pale spermatogonia 332.85 rpm, in leptotene/zygotene spermatocytes 556.08 rpm and in early pachytene 355.35 rpm. The amount of PRM2 mRNA in the respective cell types is 1299.09 rpm, 1661.42 rpm, 937.14 rpm and 891.23 rpm. In the late pachytene spermatocytes the amount of PRM1 and PRM2 mRNA transcripts is comparatively low with 18.21 rpm and 35.08 rpm, respectively. In round spermatids the amount of PRM1 and PRM2 mRNA transcripts increases sharply to 5328.07 rpm and 6592.15 rpm, respectively. In sperm heads of mature sperm cells, the amount of mRNA transcripts decreases drastically, with 57.69 rpm for the PRM1 mRNA and 20.34 rpm for the PRM2 mRNA. The amount of PRM2 mRNA transcripts is higher in cell types during spermatogenesis compared to PRM1 mRNA but slightly lower in mature sperm cells. The transcript levels during early stages of spermatogenesis are quite stable, decrease

drastically in late pachytene spermatocytes, then increasing sharply in round spermatocytes and decreasing again in mature sperm cells.

2.2.2.2 PRM1 and PRM2 proteins can be traced in human sperm heads

To validate the abundance of PRM1- and PRM2-proteins in sperm heads and total ejaculate and to ensure the proper functionality of the used antibodies dot blots using antibodies against PRM1 and PRM2 were carried out.

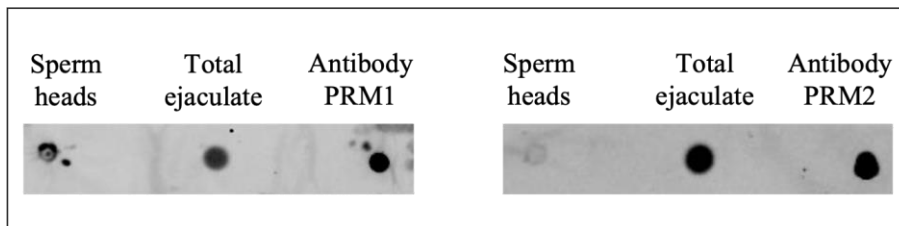


Figure 9: Dot blot results using the antibodies against PRM1 and PRM2. The left membrane was treated with PRM1 antibody and shows a sperm head sample (left), a total ejaculate sample (middle) and the primary antibody PRM1 (right). The right membrane was treated with PRM2 antibody and shows a sperm head sample (left), a total ejaculate sample (middle) and the primary antibody PRM2 (right).

The membrane on the left shows the sperm heads (left) and total ejaculate (middle) treated with PRM1 and the membrane on the right shows the sperm heads lysates (left) and total ejaculate (middle) treated with PRM2. The dots on the right side display the primary antibodies as positive control for signal detection. For every sample and used antibody clear signals could be detected. The dot for the sperm head sample treated with PRM2 antibody shows a weaker signal compared to the total ejaculate sample. The dots on the right containing the used antibody present a positive control for the functionality of the primary antibody.

To confirm the functionality of the antibody against H2B and the presence of H2B in sperm heads a western blot was carried out. The results are shown below.

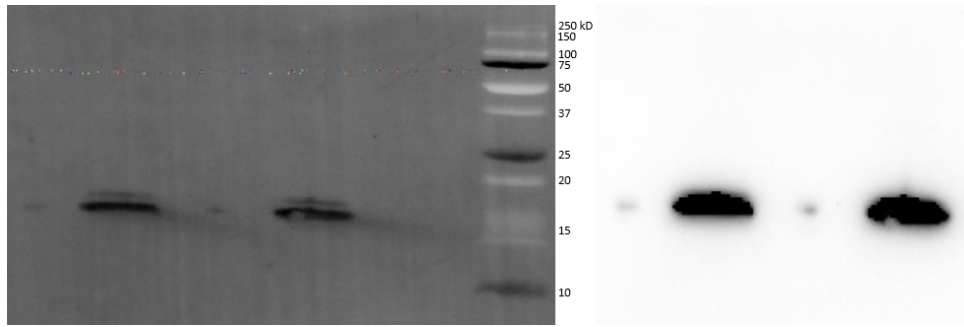


Figure 10: Western blot results using the antibody against H2B. The left side shows the PVDF membrane under light and the right side shows the membrane under chemiluminescence conditions. The bands show the sperm head samples.

Figure 10 shows the PVDF membrane under light (left) and under chemiluminescence conditions. Two bands can be seen clearly where the H2B antibody bound the corresponding proteins in the sperm head sample. The two bands are at a height of about 15 kDa.

2.2.2.3 ChIP-Seq with antibodies directed against PRM1, PRM2 and H2B uncover a small number of Peaks

Sperm heads from four individuals were immunoprecipitated using the antibodies against the sperm-specific nuclear basic proteins PRM1 and PRM2, as well as the core histone H2B. Peak calling was done by Novogene using MACS2 (Y. Zhang et al., 2008) routines and resulted in the respective peak coordinates. I could identify only a small number of peaks as compared to results presented in Jeon et al. (2020). ChIP-Seq analyses thus resulted in a total of 864, 401, 588 and 819 general peaks (count of summits) for the H2B samples, 474, 749, 424 and 758 general peaks for the PRM1 samples and 517, 826, 698 and 512 general peaks for the PRM2 samples of Individual 1, Individual 2, Individual 3 and Individual 4, respectively. The number of narrow peaks is slightly lower with 792, 362, 533 and 757 narrow peaks for the H2B samples, 438, 694, 385 and 687 narrow peaks for the PRM1 samples and 460, 747, 629 and 512 narrow peaks for the PRM2 samples of Individual 1, Individual 2, Individual 3 and Individual 4, respectively.

The differences between the general peaks and the narrow peaks can be explained by the fact that some peaks have the same coordinates but different p values and fold enrichment values. Those peaks are combined in the narrow peaks, resulting in a slightly smaller number of peaks. This fact however, does not impact the annotation results.

The Fraction of Reads in Peaks (FRiP) for the individual samples is 0.7910 %, 0.9818 % and 1.0379 % for the H2B, PRM1 and PRM2 samples of Individual 1, respectively. For Individual 2 the FRiP is 0.6562 % for the H2B sample, 0.9848 % for the PRM1 sample and 1.6417 % for the PRM2 sample. The FRiP for Individual 3 is 1.1252 %, 0.4854 % and 0.9004 % for the H2B, PRM1 and PRM2 sample, respectively. Individual 4 has FRiP values of 1.1321 % for the H2B sample, 0.6700 % for the PRM1 sample and 0.9238 % for the PRM2 sample.

2.2.2.4 PRM1 and PRM2 act in concert

To verify if PRM1 and PRM2 act in concert to constitute the nucleoprotamine packaging in mature human sperm heads BEDTOOLS (Quinlan & Hall, 2010) INTERSECT was used to compare the peak coordinates of PRM1 to PRM2 peak coordinates for the four individuals. The obtained Excel sheets from Novogene containing the peak information were used to create bed files with the peak coordinates (chromosome, start position and stop position). Those bed files were used for the INTERSECT analyses. The results are listed in the table below.

Table 1: Number of PRM1 and PRM2 peaks and overlapping peaks for all four donors.

	Number of PRM1 peaks	Number of PRM2 peaks	Number of overlapping peaks
Donor 1	474	517	498
Donor 2	749	826	604
Donor 3	424	689	383
Donor 4	758	567	458

2.2.2.5 The nucleoprotamine DNA interaction pattern as revealed by ChIP does not differ from that of nucleohistone H2B

To get a first impression about the peak distribution of H2B, PRM1 and PRM2 along the chromosomes, the NCBI Genome Decoration Page was used. The organism was specified to “Homo sapiens” and the representation of the chromosomes was set to “cytogenetic”,

“GRCh38.p12” and a resolution of 850 bands per haploid karyotype. The bed files of the H2B, PRM1 and PRM2 ChIP-results were uploaded for each individual. Also, the peak distribution for each protein was compared between individuals and is shown in appendix figures 1 – 3.

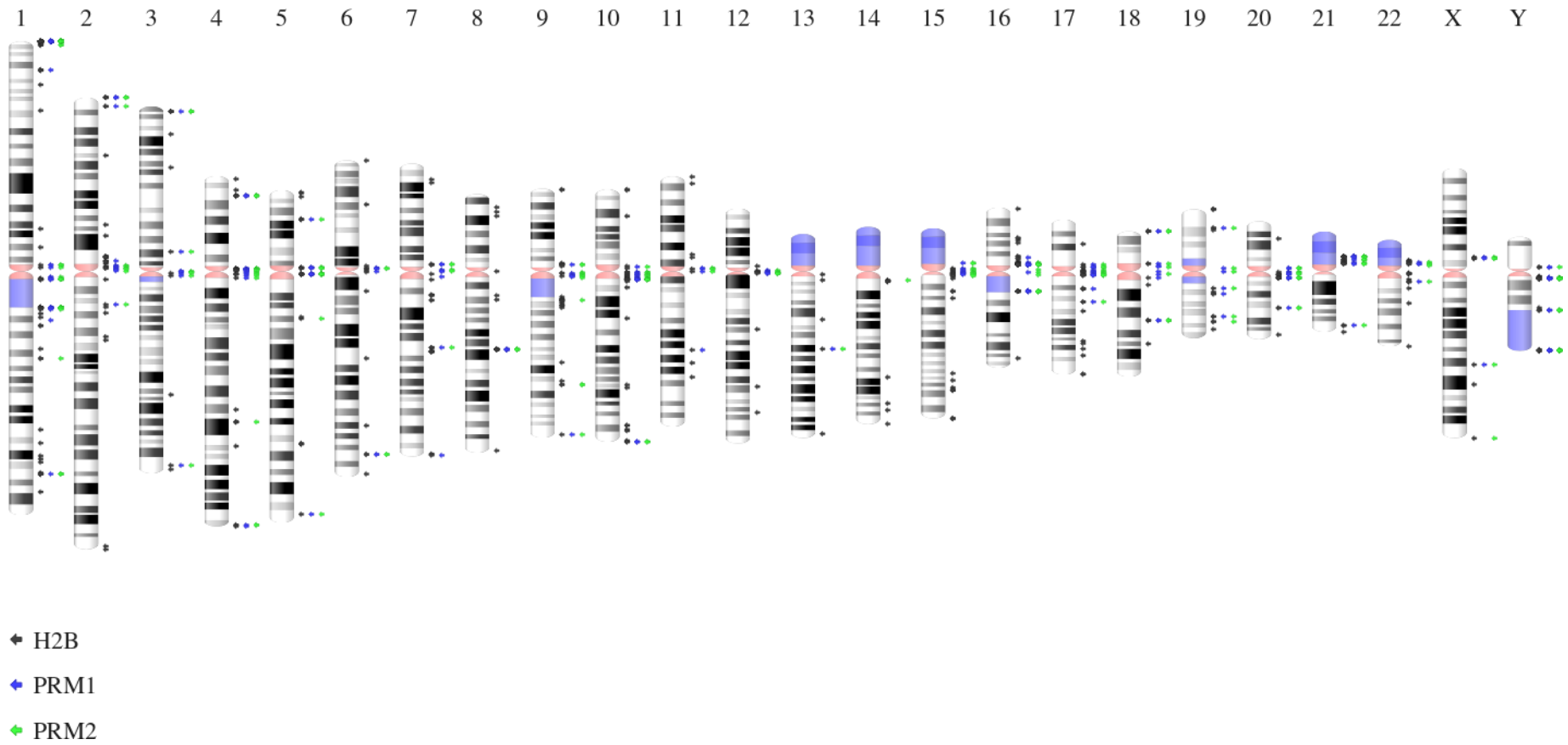


Figure 11: Idiogram of Individual 1 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38. H2B peaks are shown as black arrows, PRM1 peaks are shown as blue arrows and PRM2 peaks are shown as green arrows.

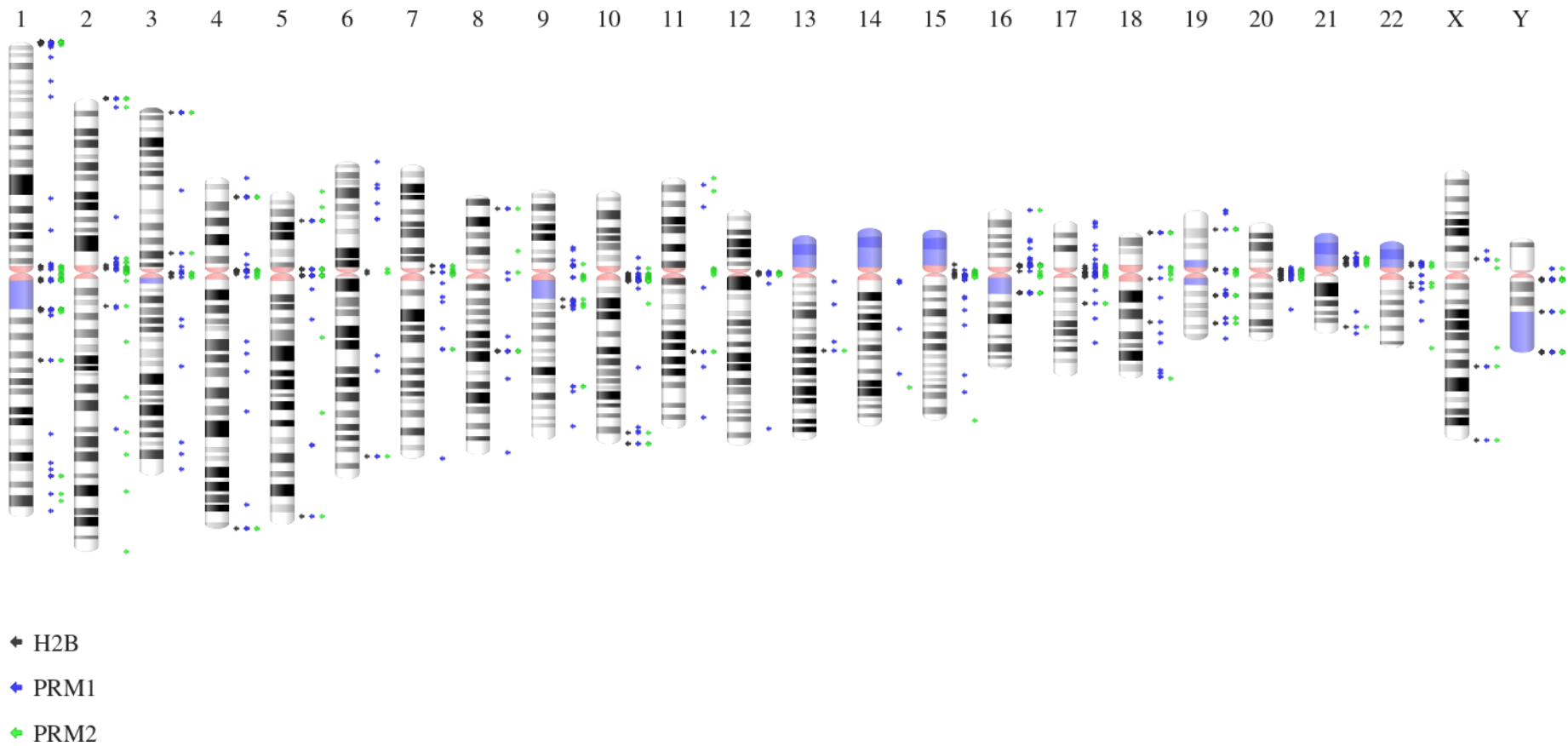


Figure 12: Idiogram of Individual 2 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38. H2B peaks are shown as black arrows, PRM1 peaks are shown as blue arrows and PRM2 peaks are shown as green arrows.

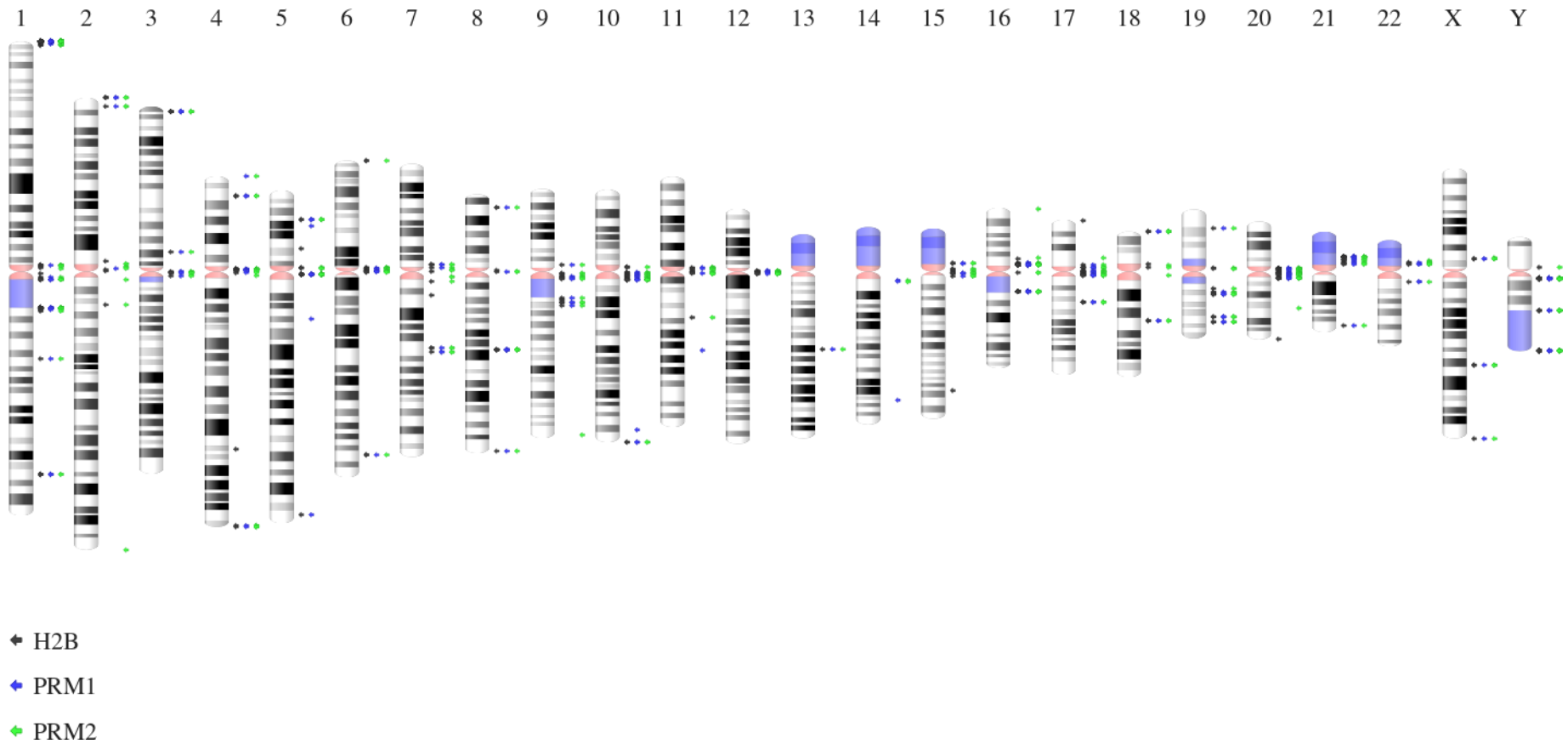


Figure 13: Idiogram of Individual 3 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38. H2B peaks are shown as black arrows, PRM1 peaks are shown as blue arrows and PRM2 peaks are shown as green arrows.

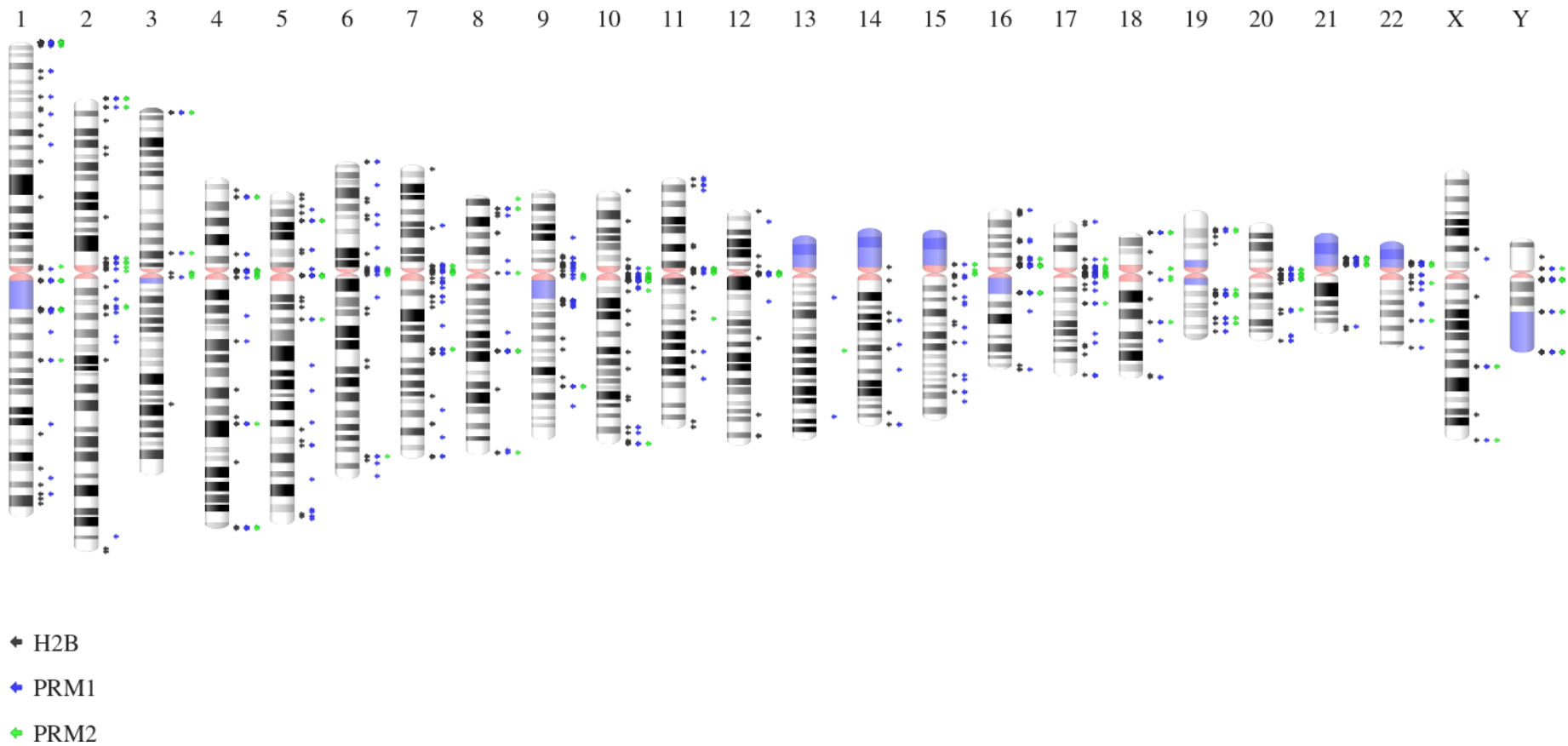


Figure 14: Idiogram of Individual 4 showing where H2B, PRM1 and PRM2 peaks are found across the human genome hg38. H2B peaks are shown as black arrows, PRM1 peaks are shown as blue arrows and PRM2 peaks are shown as green arrows.

The ideograms show where the peaks of the H2B, PRM1 and PRM2 samples are located along the chromosomes of hg38 for each individual (Individual 1 is shown in figure 11, Individual 2 is shown in figure 12, Individual 3 is shown in figure 13 and Individual 4 is shown in figure 14). The peaks of the H2B samples are shown as black arrows, the peaks of the PRM1 samples are shown as blue arrows and the PRM2 peaks are shown as green arrows. As can be seen, a lot of peaks are located around centromere and telomere regions and H2B, PRM1 and PRM2 peaks overlap mainly in those regions. The remaining peaks show partially great differences in terms of their localisation along the chromosomes between the individuals.

The previously defined peak regions generated by peak calling were annotated using HOMER to get an overview in which genomic regions ("genomic feature") the examined proteins bind. The annotations resulted in a similar pattern for the three proteins and individuals and are shown below. The genomic features are displayed on the x-axis for the different donors and the \log_2 fold-change(observed/expected) as compared to hg38 is shown on the y-axis.

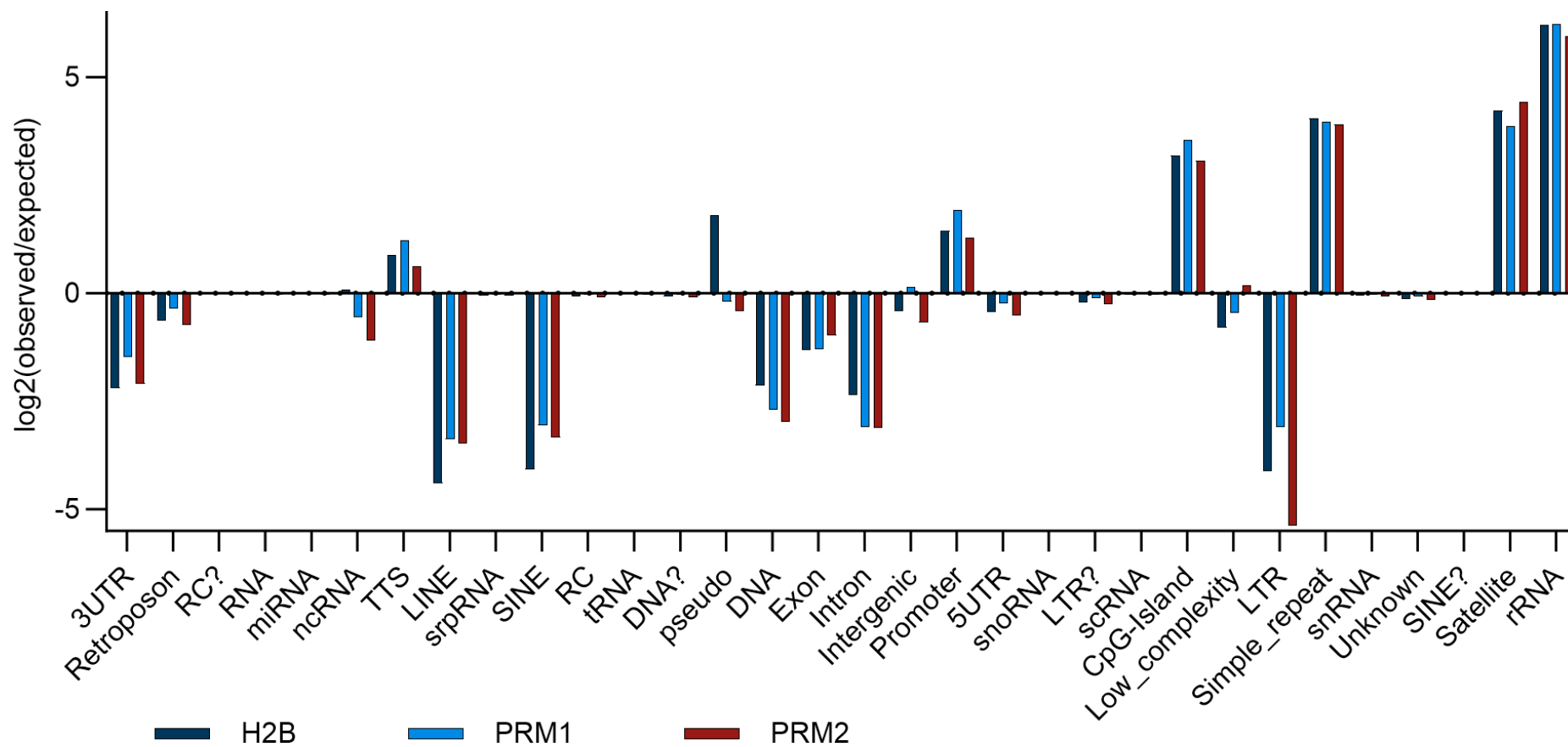


Figure 15: HOMER annotation results for H2B, PRM1 and PRM2 peaks of Individual 3. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed/expected})$ on the y-axis. Results for H2B are shown in dark blue, results for PRM1 are shown in light blue and PRM2 results are shown in dark red.

Figure 15 shows the HOMER annotation for the H2B, PRM1 and PRM2 peaks for one individual (the annotation results for the other three individuals are shown in appendix figures 4 – 6). The genomic features are displayed on the x-axis and the peak ratio is listed as $\log_2(\text{observed/expected})$ on the y-axis. For most genomic regions, the annotation results of H2B, PRM1 and PRM2 show the same pattern. More peaks than expected can be found in TTS (Transcription Termination Site), Promotor, CpG islands, Simple Repeat, Satellite and rRNA regions. Less peaks than expected are found in 3' UTR, LINE, SINE, DNA, Exon, Intron and LTR regions.

To compare the biological variability, the annotation results of the four individuals for each protein were compared and displayed below.

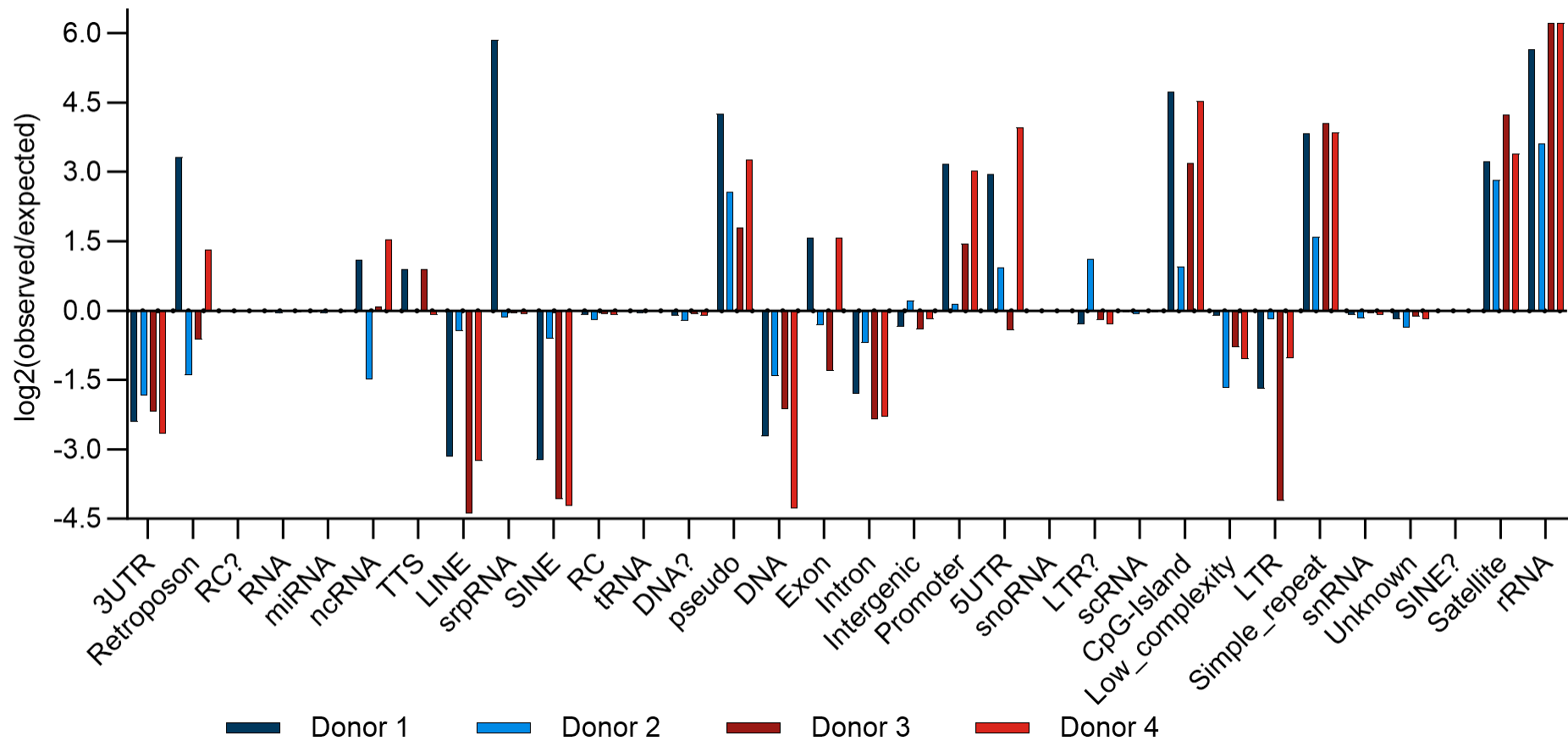


Figure 16: HOMER annotation results for H2B peaks of all four individuals. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed}/\text{expected})$ on the y-axis. Results for Donor 1 are shown in dark blue, results for Donor 2 are shown in light blue, results for Donor 3 are shown in dark red and results for Donor 4 are shown in red.

Most of the regions or genomic features show very similar patterns of representation compared to the reference genome hg38 for all donors. However, exceptions from this include Retroposon, ncRNA, Exon, Intergenic and 5' UTR regions that exhibit interindividual different patterns. The regions showing the biggest difference with respect to over/under-representation compared to the reference genome are the rRNA, Satellite and Simple Repeat regions with more peaks than the reference genome and LINE- and SINE-regions with less peaks compared to the reference genome.

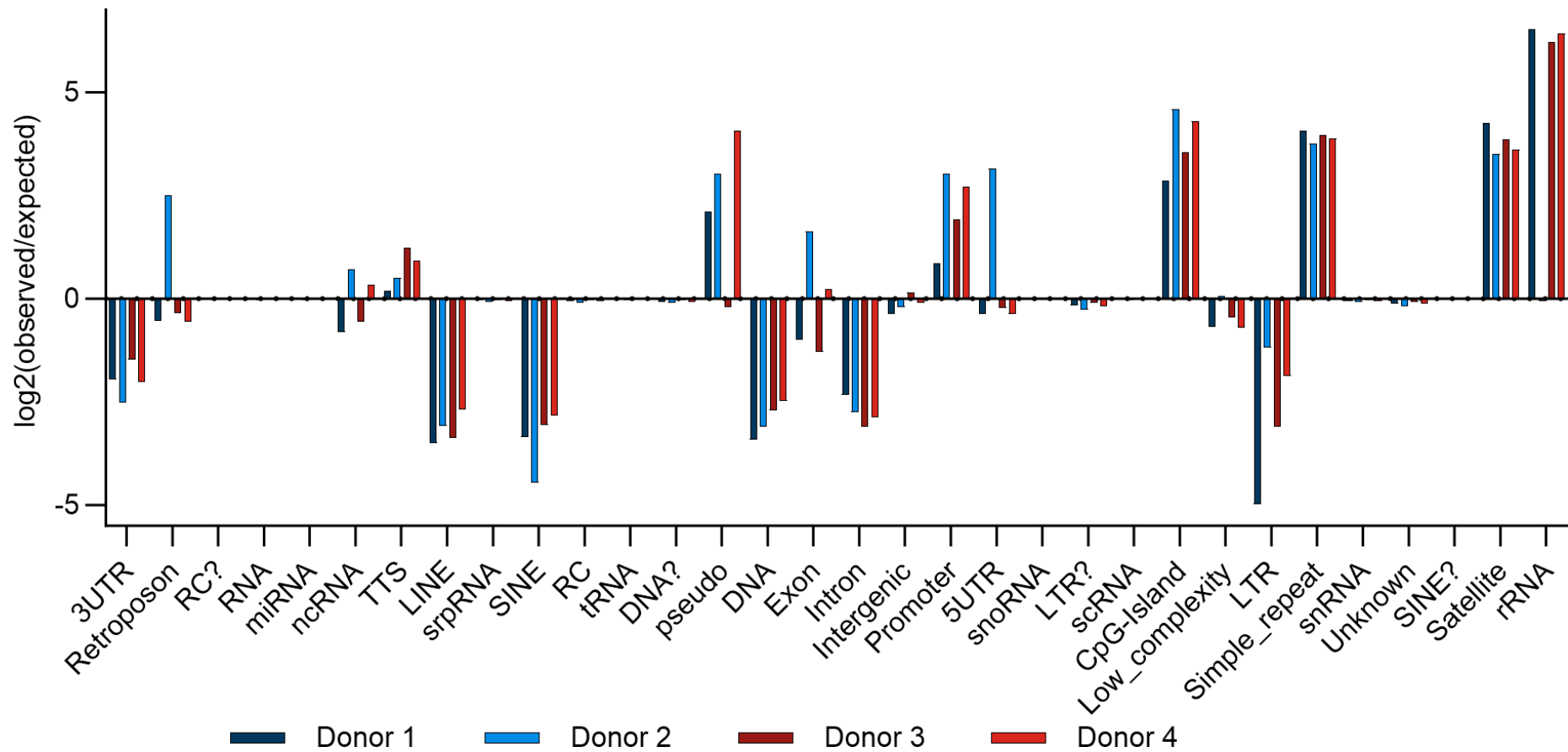


Figure 17: HOMER annotation results for PRM1 peaks of all four individuals. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed}/\text{expected})$ on the y-axis. Results for Donor 1 are shown in dark blue, results for Donor 2 are shown in light blue, results for Donor 3 are shown in dark red and results for Donor 4 are shown in red.

Most regions show the same peak pattern for all individuals. The genomic regions assigned to Retroposon, ncRNA, pseudo genes (pseudo), Exon, Intergenic and 5' UTR show great individual differences for one or two individuals compared to the others. Regions with a higher peak distribution compared to the reference genome are rRNA, Satellite, Simple Repeat and CpG-Island regions. Regions with a lower peak distribution compared to the reference genome are 3' UTR, LINE, SINE, DNA and Intron regions. About half of the genomic regions show no or almost no differences in the peak distribution compared to the reference genome.

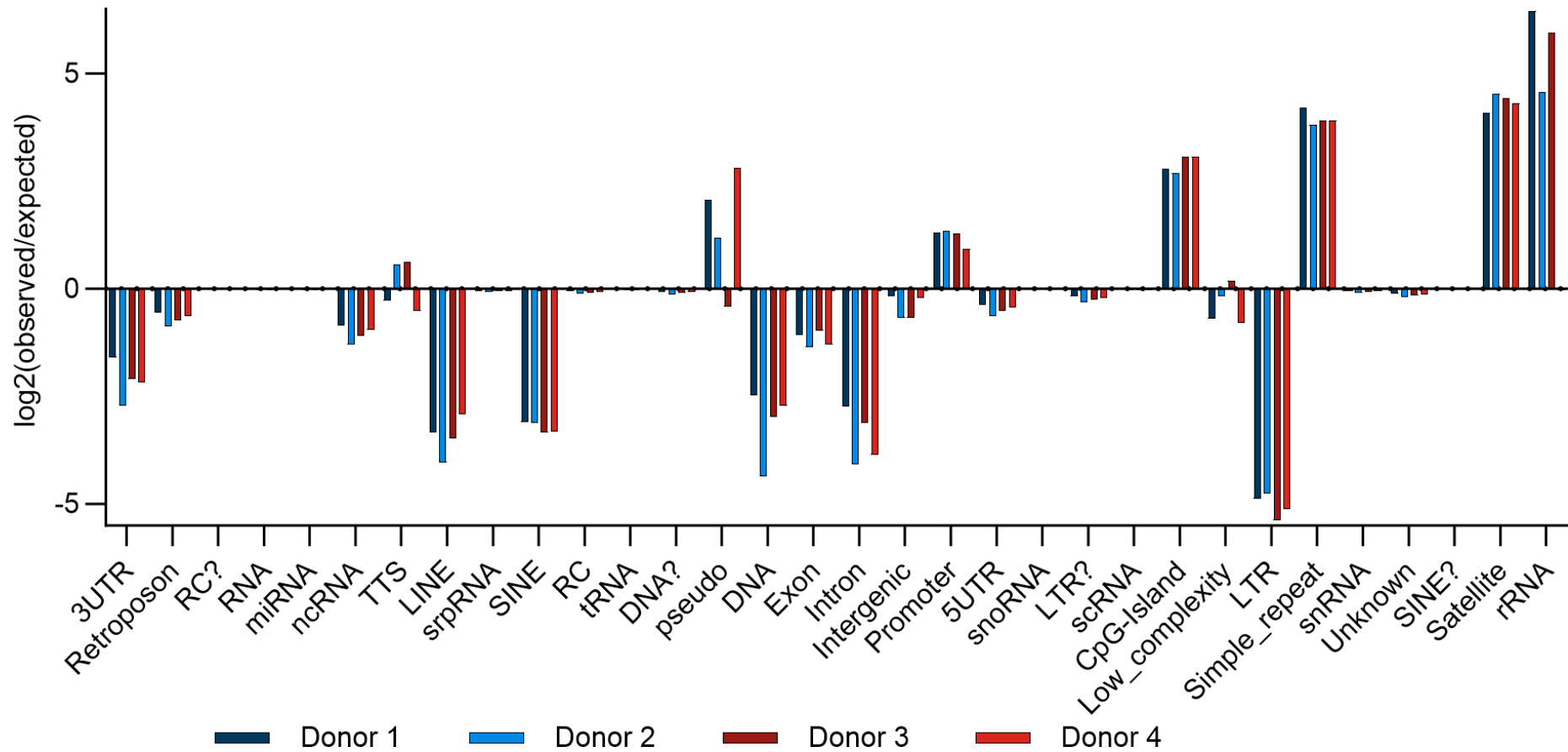


Figure 18: HOMER annotation results for H2B peaks of all four individuals. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed}/\text{expected})$ on the y-axis. Results for Donor 1 are shown in dark blue, results for Donor 2 are shown in light blue, results for Donor 3 are shown in dark red and results for Donor 4 are shown in red.

Most regions display the same peak pattern for all individuals and the peak distribution shows great similarities to the PRM1 peak annotation. Regions with a higher peak distribution than the reference genome are the rRNA, Satellite, Simple Repeat and CpG-Island regions. Interestingly, all four samples show a higher peak count in the Promotor region compared to the reference genome. Regions with a way smaller peak distribution compared to the reference genome are the 3' UTR, LINE, SINE, DNA, Intron and LTR regions and to a lesser degree Retroposon, ncRNA, Exon, Intergenic and 5' UTR regions.

Not all peaks of each ChIP-Seq sample could be mapped to genomic regions. The total amount of peaks mapping to genomic regions are 611, 278, 409 and 602 for the H2B samples of Individual 1, Individual 2, Individual 3 and Individual 4, respectively. The number of PRM1 peaks mapping to genomic regions are 329, 536, 285 and 542 and the number of PRM2 peaks are 347 for Individual 1, 638 for Individual 2, 492 for Individual 3 and 407 for Individual 4. This results in about 70 % of mappable general peaks (67.12 % - 77.24 %) and about 78 % of the narrow peaks being mappable (74.03 % - 85.41 %).

All three proteins show very similar peak distribution values in 3' UTR, SINE, DNA and Simple Repeat regions in all four individuals and with slightly more variations between the chromatin constituting proteins in LINE, Satellite and rRNA regions (mean values of the four donors for each protein compared to the general mean value of all samples \pm 10 % confidence interval). For both protamines very similar peak distributions were found in 3' UTR, LINE, SINE, DNA and Simple Repeat regions (mean values of the four donors for each protein compared to the general mean value of all samples \pm 5 % confidence interval). Slightly more variance was found in Satellite and rRNA regions (10 % confidence interval). The peak distributions for H2B show more variance between individuals compared to the protamine peaks.

Regions collectively combined in the satellite region of the HOMER annotation include alpha satellite (α -satellite, ALR), beta satellite (β -satellite, BSR), gamma satellite II (GSATII) and HSATI. Other regions included in the group of satellite are Telomere-Associated Repeats (TAR1), SAR, SST1, as well as simple satellite repeats (CATTC)_n and (GAATG)_n. To the DNA region HSMAR2 could be assigned. The class of LINEs (long interspersed nuclear elements) in the HOMER annotation showed only LINE-1 retrotransposons, while the SINE (short interspersed nuclear elements) group includes mainly *Alu* elements.

Looking more closely at the peak distribution of the genomic regions, I could observe that the majority of peaks map to satellite regions (~28 % on average in the H2B samples, 35 % in the PRM1 samples and 50 % on average in the PRM2 samples), followed by intergenic regions (22 % on average in the H2B samples, 23 % in the PRM1 samples and 18 % in the PRM2 samples) and simple repeat regions (~17 % on average in the PRM1 and PRM2 samples and 14 % in the H2B samples). About 0.3 % of the peaks of all samples map to rRNA regions (PRM1 0.4 %). Interestingly, no PRM2 peaks map to Retroposon, ncRNA and 5' UTR region, while H2B and PRM1 do. Only two out of four individuals exhibited peaks assigned to retroposon regions in the H2B samples and only one individual exhibit peaks assigned to that region in the PRM1 samples. All H2B samples show peaks assigned to ncRNA regions, but only two individuals exhibit peaks mapped to that region in the PRM1 samples. Three out of four individuals show peaks in the H2B samples assigned to 5' UTR regions and only one person shows peaks in that region for the PRM1 samples. In all samples peaks were detected that were annotated to LINE and SINE regions (3 % of all peaks assigned to LINEs and 2 % assigned to SINEs on average), DNA (average of 0.5 % of all peaks in all samples), intron (mean percentage of 0.5 % of all peaks for all samples) and CpG island regions (mean percentage of 3.8 %). LTR regions show the highest number of peaks in H2B samples (except Individual 3 where H2B and PRM1 both show two peaks in that region) with an average of 3.7 % of all peaks compared to 1.8 % in PRM1 samples and 0.3 % in PRM2 samples.

To test if the RNAs found in the UNITAS outputs of mature sperm heads are potentially transcribed inside the sperm heads because they are part of the nucleohistone fraction or if they represent remnants of transcription in earlier stages of spermatogenesis in the nucleoprotamin fraction, an INTERSECT analysis was carried out. Bed files from different RNA classes were downloaded from LNCipedia (Version 5.2 GRCh38/hg38), the UCSC Table Browser (GRCh38/hg38), piRNADB and dashr2 (v2.0, hg38) and compared to the coordinates of the PRM1-, PRM2- and H2B-covered genomic regions applying the bedtools-INTERSECT routine.

The comparison of the peak coordinates of PRM1/PRM2 and H2B to the different RNA genomic regions showed great differences between the different genomic regions and examined proteins.

The highest number of overlaps with RNA genomic regions was found in the lncRNA regions with an average of 250.5 overlapping regions for H2B, 207.25 overlaps for PRM1 and 140.75 overlaps for PRM2 (individuals show great variance in the amount of overlapping regions for the three proteins). The second highest number of overlaps was scored with piRNA genomic regions, with 63.75, 43.5 and 22.25 overlaps on average for H2B, PRM1 and PRM2, respectively. Only minor overlaps could be found between the examined proteins and rRNA genomic regions and sno/miRNA genomic regions. No overlaps were found in the comparison with tRNA genomic regions. The results for each individual are shown in appendix tables 1 – 5.

To get a more graphical and statistical overview about the peak comparison of peaks that overlap with lncRNA in the H2B, PRM1 and PRM2 samples Venn diagrams were created using the Intervene Venn module. The results from the INTERSECT analyses with the ChIP data and the lncRNA data were used to create Venn diagrams to find overlapping regions between the four individuals for each protein.

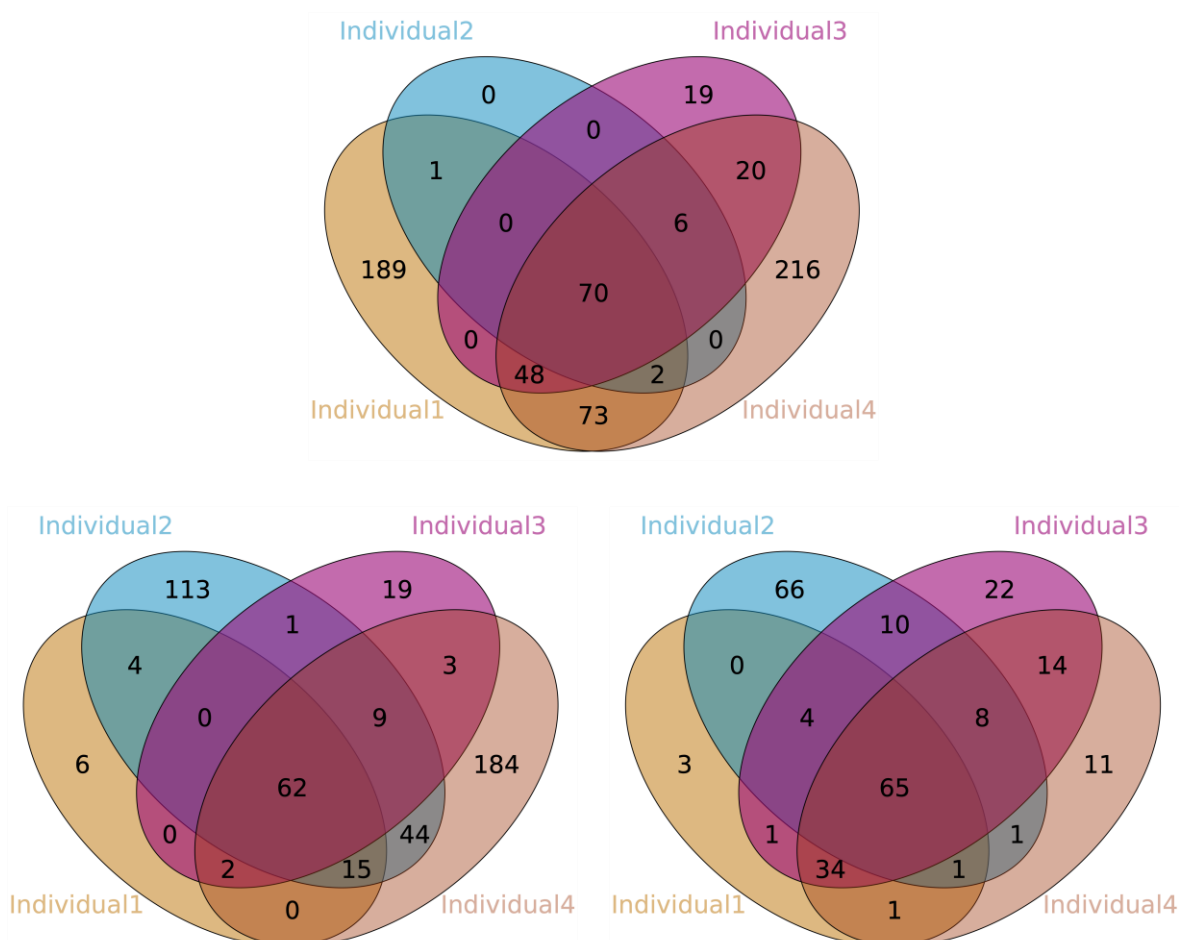


Figure 19: Venn diagrams of lncRNA overlaps between all four donors for H2B (upper diagram), PRM1 (lower left diagram) and PRM2 (lower right diagram). The Venn diagrams show the number of lncRNA-H2B (upper diagram), lncRNA-PRM1 (lower left diagram) and lncRNA-PRM2 (lower right diagram) overlapping peaks for each individual and the amount of overlapping peaks between individuals (overlapping regions of the ovals). The peaks of Individual 1 are represented in the sand-coloured oval, the peaks of Individual 2 in the cyan-coloured oval, the peaks of Individual 3 in the magenta-coloured oval and the peaks of Individual 4 are shown in the tan-coloured oval.

The Venn diagrams show the peak comparisons of the lnc-H2B (upper Venn diagram), lnc-PRM1 (lower left Venn diagram) and lnc-PRM2 (lower right Venn diagram) hits between individuals. Each oval represents the overlapping peaks between protein of interest and lncRNA peaks from one individual with the corresponding number of peaks, whereas overlapping regions show the number of peaks that overlap between two, three or all four individuals. The peaks of Individual 1 are represented in the sand-coloured oval, the peaks of Individual 2 in the cyan-coloured oval, the peaks of Individual 3 in the magenta-coloured oval and the peaks of Individual 4 are shown in the tan-coloured oval. 70 peaks overlap

between all four individuals of the lnc-H2B peaks, 62 peaks overlap completely between the lnc-PRM1 peaks and 65 peaks overlap between the lnc-PRM2 peaks for all individuals.

Out of these total overlaps (70 for H2B, 62 for PRM1 and 65 for PRM2) all lnc-H2B overlapping peaks overlap with at least one of the protamines, while lnc-PRM1 overlapping peaks and lnc-PRM2 overlapping peaks both have two unique peaks, and 13 overlapping regions are protamine-specific. The majority of peaks that overlap with lncRNA peaks also overlaps between the three protein samples (62 total overlaps). The overlapping regions are distributed along Chr3, Chr4, Chr6, Chr7, Chr8, Chr17, Chr18, Chr21, Chr22, ChrX and ChrY. The overlapping regions are shown in the Venn diagram below.

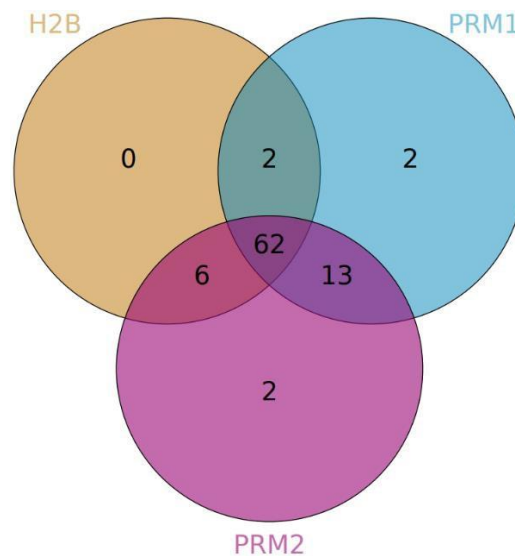


Figure 20: Venn diagram of the total lncRNA overlaps (overlaps that were found in all individuals) between H2B (sand-coloured circle), PRM1 (cyan-coloured circle) and PRM2 (magenta-coloured circle). Numbers show unique and overlapping peaks.

2.2.2.6 RNA-binding capacity of protamines

Protamines are intrinsically disordered proteins that have the ability to bind RNA. This protein-RNA-interaction was more precisely examined, more specifically the interaction with lncRNA since these RNAs were found in the complete RNA cargo of sperm heads and a lot of H2B-, PRM1- and PRM2-associated genomic regions overlap with lncRNA genomic regions. Therefore, RNA immunoprecipitation was done with sperm head samples of four individuals with the same antibodies as used for the above-mentioned ChIP-analyses. RNA-Seq was carried out by Novogene with a strategy to enrich for lncRNA enrichment, including

rRNA removal and omitting size fractionations. The obtained sequences were HISAT2-mapped to hg38 to determine the quality of the datasets and resulted in a mean percentage of alignable reads of 18.13 % (individual alignment rates differ greatly between datasets, ranging from 1.96 % to 54.16 %). The obtained bam files were mapped to the lncRNA gtf file obtained from LNCipedia using FEATURECOUNTS and were corrected for the overall alignment rate. The individual results of the 100 most abundant lncRNA hits are shown in appendix tables 6 – 9. Additionally, the bam files were also mapped to tRNA and sno/miRNA gtf files obtained from the UCSC Table Browser. The numbers for tRNA and sno/miRNA overlaps are shown in appendix figures 11 and 12.

2.2.2.7 Protamines associate with lncRNA in an unspecific manner

The number of sequences mapping to lncRNA genes and the total number of lncRNA genes found in the individual datasets differ greatly with 3,054,954 hits mapping to 17,277 lncRNA genes (Individual 1 H2B sample) to 22,148 hits mapping to 360 lncRNA genes (Individual 3 H2B sample; other samples show less lncRNA genes but higher hit numbers). The lncRNA hits make up between 2.44 % (Individual 3 H2B) and 17.69 % (Individual 1 H2B) of the alignable sequences.

To get a quantitative impression on the lncRNA profiles, the READCOUNT outputs were sorted according to read number from highest to lowest for each individual and compared to one another. With that I reproduced the overall qualitative pattern of lncRNA reads with respect to their abundance: Two lncRNA genes were found in all samples (except Individual 2) in the upper 25 % read counts of all hits, those are lnc-LRR1-1 and lnc-NEMF-1. One lncRNA gene was found in the upper 25 % read counts of all protamine samples (except Individual 2), this lncRNA is lnc-CCNB1IP1-1. The majority of lncRNAs found in each sample differ greatly in terms of abundance, showing both qualitatively and quantitatively a high variance in the lncRNA distribution. The amount of overlapping lncRNA genes for the 50 most abundant lncRNA genes between individuals are shown in appendix table 13.

Only a minor amount of sequences mapped to tRNA genes in all samples (0 – 5 tRNA genes with 92 – 4,038 sequences in total for one dataset). The exception is the H2B data of Individual 3, here 88 tRNA genes were found. The individual results differ greatly in terms of quality and quantity of the tRNA distribution and are shown in appendix table 10. The tRNA hits make up between 0.00 % and 0.0256 % of the alignable sequences.

A slightly higher amount of sequences mapped to sno/miRNA genes. Between 177 sequences (Individual 3 H2B) and 3540 sequences (Individual 1 PRM2) mapped to four (Individual 2 PRM2) to 310 (Individual 1 H2B) sno/miRNA genes. These hits make up between 0.0233 % and 0.1384 % of the alignable sequences. The sno/miRNA genes show no specific pattern regarding their association with the examined proteins. The individual number of hits are shown in appendix table 12.

2.2.3 Discussion

During the late stages of spermatogenesis, histones are massively replaced in round spermatids by transition proteins, which are then replaced in elongating spermatids by protamines, resulting in a highly condensed chromatin packaging in elongated spermatids and mature spermatozoa. In humans around 85 % of the nucleohistones are replaced by protamines (Steger et al., 2000) leading to transcriptionally silenced chromatin. To confirm the findings of Steger et al. and to get a more detailed picture at which time point of spermatogenesis and in which ratio the PRM1 and PRM2 genes are transcribed, local BLAST routines were performed using SRA data of different cell types during spermatogenesis (BioProject PRJNA310976) with the protamine mRNA sequences obtained from the NCBI CCDS database as database. In all cell types PRM1 and PRM2 transcripts could be found, with more PRM2 transcripts in early stages of sperm development and slightly more PRM1 transcripts in mature sperm cells. The abundance of PRM1 and PRM2 transcripts is relatively low before the first meiotic division, high in round spermatids and decreases again in mature sperm cells. These results are consistent with the results of Steger et al. The high abundance in round spermatids correlates with the initiation of the histone-to-protamine transition (Steger et al., 2000). In round spermatids the PRM1 and PRM2 genes are transcribed, whereupon the transcripts are translated in elongating spermatids, resulting in a sharp decrease in the amount of transcripts in mature sperm cells and the condensation of chromatin. Interestingly, PRM1 and PRM2 transcripts belong to the 20 most abundant transcripts in sperm heads, showing a much higher abundance in the READSCOUNTS analysis compared to the local BLAST analysis. These dissimilar findings might be caused by the fact that BLAST may miss some matches. The BLAST strategy is expected to find most matches but sacrifices high sensitivity in order to gain fast results. Also, the quality of the BLAST results depends on the chosen parameters and the quality of the input samples. As the READSCOUNTS annotation ranked hY1 as the most abundant transcript in sperm heads, which was also found using BLAST routines but not to that extent, it might be assumed that the annotation using the human reference genome and READSCOUNTS results in a more accurate result.

To find out if the proteomic situation correlates with the transcript patterns observed, I used antibodies against PRM1 and PRM2 to check for the presence of protamine proteins in sperm heads. To this end, dot blots were created using a sperm head sample and a sample of total ejaculate. To gain more information about the residual histones in sperm heads, an antibody against the core histone H2B was used. The presence of H2B in sperm heads and the

functionality of the antibody were checked by western blotting. Furthermore, the dot blot and western blot approaches were used to confirm the quality and binding ability of the antibodies used in the ChIP-Seq and RIP-Seq experiments. The western blot did show a result for H2B (even though the disaggregation of the ladder showed some difficulties [the 15 kDA bar is more widened than the other bars]). According to the supporting data of the used antibody, H2B does have a molecular weight of 14 kDA. In the performed western blot the size of H2B is around 15 kDa which aligns with the known size of this protein. SDS PAGES for PRM1 and PRM2 were performed but did not show a detection signal. On the other hand, the signal detection from the dot blot approaches showed clear signals of the antibodies in the sperm head and total ejaculate samples. Thus, it could be verified that the used antibodies were suitable for the following experiments. SDS PAGES cannot be used for every protein and does have some limitations like the resolution of modified histones or highly basic proteins like protamine, for that the IEP is more basic than the applied buffer system (the isoelectric point of H2B is 10.32, of PRM1 12.08 and of PRM2 11.9, respectively). Furthermore, proteins are denatured prior to electrophoresis, which disrupts their native conformation and can lead to altered protein structures and binding properties. Using a native approach without denaturing detergents allows an improved binding of the antibody to the protein (Glenney et al., 1983; Littauer et al., 1986). Another limiting factor is the molecular weight of proteins. SDS PAGE is not suitable for effectively separating proteins with very low molecular weight or proteins with similar molecular weight in general. Proteins with low molecular weight might migrate too quickly and appear as a single protein, or if the gel is running too long they might even run out of the gel. If the gel is not running long enough the proteins might not separate adequately as well. The conventional buffer systems that can be used for histones does not work for the highly basic protamines. Isolating protamines according to their isoelectric point using isoelectric focusing might be a more suitable method. Isoelectric focusing separate proteins according to their isoelectric point along a pH gradient under non-denaturing conditions (Smoluch et al., 2016).

In mature sperm cells most of the histones are replaced by protamines, therefore further transcription of these proteins is not necessary. In addition, the transition of histones to protamines is assumed to lead to a transcriptional inactivation of most genes.

Due to the incomplete histone-to-protamine transition a subset of genomic loci may escape complete removal of the histone packaging proteins, with epigenetic effects of the respective gene regulation. But not only residual histones, but also protamines might have an impact in epigenetic processes. Therefore, in this work I aimed to correlate the non-coding RNA

expression with the chromatin data of residual histones, represented by H2B, and PRM1 and PRM2 and investigate the possible role of chromatin components in the transgenerational inheritance of epigenetic information. To get a general overview about the histone and protamine chromatin package ChIP analyses were performed.

In all samples only small overall numbers of peaks could be detected. The relatively small number of peaks in the H2B samples can be explained with the exchange of histones to protamines and the only small number of residual histones remaining in the sperm nucleus. The small numbers for the PRM1 and PRM2 samples might result in the arrangement of those proteins with DNA and their size. Protamines contain arginine-rich anchoring domains that bind to the diester backbone of DNA (Brewer et al., 1999). Thereby, one protamine molecule binds in the DNA groove and binds one turn of the helix (Hud et al., 1993). PRM1 binds 10 – 11 bp of DNA and PRM2 binds 15 bp (Balhorn, 2007). The binding of protamine to DNA leads to a toroidal structure containing up to 60 kb of DNA, and the sperm nucleus can contain up to 50,000 toroidal loops (Hud et al., 1993). This tight package might result in a more even distribution of the protamine package along the DNA. This tight packaging might also influence the binding capacity of the antibody to all protamines in the sperm nucleus. As the ratio of PRM1 to PRM2 in sperm nuclei is supposed to be 1:1, the amount of protamines per toroidal loop should be around 5,000 (60,000 bp DNA per toroidal loop*0.5/10 bp DNA bound by PRM1 + 60,000 bp DNA per toroidal loop*0.5/15 bp DNA bound by PRM2) and the total number of protamines should be around 250,000,000 (5,000 protamines*50,000 toroidal loops). Additionally, as there are about 50,000 toroidal loops in one sperm nucleus and one toroidal loop binds about 60,000 bp DNA the amount of DNA that is bound by all protamines should be around 3 billion bp ($3 \times 10^9 = 3$ Gigabase (Gb) pairs; 60,000 bp DNA x 50,000 toroidal loops). As the human genome has a size of about 6.4 billion bp around half of the genome should be covered by protamines. This could not be replicated in the ChIP-Seq experiments, which supports the assumption that peak calling could not accurately predict all proteins in the samples due to their even distribution along the genome. As peaks represent regions in the genome that are enriched with aligned reads, an even distribution of the reads generated during ChIP-Seq could result in only small signals that might be falsely assigned to background noise and therefore be excluded. Depending on the amount of proteins binding to one specific area, the size of the predicted peaks are more narrow or more broad. If too many proteins are located in the same genomic area next to each other the resulting peak would be very broad, probably not passing the threshold of the used peak calling program and therefore assigned as background noise. This

assumption is further supported by the fact that the generated peaks range from 210 nt to 1700 nt and therefore can include more than just one protein binding site. However, a general trend can be seen for the ChIP analyses in that H2B, PRM1 and PRM2 bind to the same genomic regions and that the majority of the predicted peaks for those proteins do overlap. These overlaps might be caused by the size of the peaks ranging from 210 nt to 1700 nt. Histones build octamers containing two copies of each core histone H2A, H2B, H3 and H4 that wrap around 146 – 147 bp DNA to condensate the DNA into nucleosomes (Schon et al., 2019) while, PRM1 binds 10 – 11 bp of DNA and PRM2 15 bp of DNA (Balhorn, 2007). Therefore, several histone and protamine binding sites might be included in the same peak. The Fraction of Reads in Peaks (FRiP) is a measurement of how many reads are found within one peak in proportion to the total reads and can be used as a quality measure for ChIP-Seq. A high number of reads within one peak indicates that the majority of reads are located in specific enriched regions and has a high signal-to-noise ratio. On the other hand, a low number of reads within one peak indicates the location of the majority of reads in non-specific regions, implying a low specificity. According to the ChiLin pipeline (Qin et al., 2016) a good FRiP score is $\geq 1\%$. The majority of my ChIP-Seq data is below this threshold, implying a non-specific and broad sequence distribution.

HOMER is an annotation tool that identifies genomic regions where more sequencing reads are found than it would be expected by chance. Regions where less sequences were detected than would have been expected are found in 3' UTR, LINE, SINE, DNA, Intron and LTR regions while more sequences were found in CpG Island, Simple Repeat and rRNA regions than would have been expected. The HOMER annotation of PRM2 shows the most conserved patterns, with less sequences found in ncRNA and Exon regions than would have been expected in all four samples.

Regions collectively combined in the satellite region of the HOMER annotation include alpha satellite (α -satellite, ALR), beta satellite (β -satellite, BSR), gamma satellite II (GSATII) and HSATI. Those are all found in the centromeric regions (Altemose et al., 2022). Other regions included in the group of satellite are Telomere-Associated Repeats (TAR1), which are located in the telomere-proximal subtelomeric region (Dubocanin et al., 2022; Kwapisz & Morillon, 2020), SAR, SST1, as well as simple satellite repeats (CATTC)_n and the reverse complement (GAATG)_n. Satellites are highly repeated non-coding sequences (Talbert & Henikoff, 2022). α -satellites are AT-rich tandem repeats composed of 171 bp monomers that build the centromeres of all chromosomes in humans. Additionally, most

chromosomes contain classical human satellites 2 and/or 3 (HSat2 and HSat3) which are derived from the simple repeat (CATTC)_n (Altemose et al., 2022).

Interestingly and regarding the ribosomal RNA annotation, only 5S rRNA, 5.8S rRNA and 28S rRNA were assigned by the HOMER annotation to the individual samples. Those three rRNAs are all part of the 60S ribosomal subunit. There are 12 gene clusters in the human genome of diploid cells encoding rRNA: the 5S rRNA is transcribed from a cluster on chromosome 1q42, while the other three rRNA molecules (18S, 5.8S and 28S) are produced from a 47S precursor transcript expressed from clusters on chromosomes 13p12, 14p12, 15p12, 21p12 and 22p12 (Stults et al., 2008). Interestingly, according to the T2T-CHM13 assembly these clusters are located in peri/centromeric regions (Altemose et al., 2022). Mature 5S rRNA has a length of 121 nt, 18S rRNA is 1,870 nt long, 5.8S_{S/L} rRNA is 157 nt and 162 nt long and 28S rRNA has a size of more than 5,000 nt (Aubert et al., 2018). Each rDNA repeat on chromosomes 13, 14, 15, 21 and 22 has a size of about 43 kb and contains a ~13.3 kb coding region and a ~30 kb intergenic spacer. The clusters contain about 300 to 400 copies of the rDNA genes on a haploid genome (Yu et al., 2015) and show great length variability between and within individuals and range from 50 kb to more than 6 Mb (Stults et al., 2008). Considering the size of the precursor rRNA and the mature rRNAs, the arrangement of the rRNA genes in the precursor rRNA and the size of the peaks assigned to rRNA ranging from 210 nt to 670 nt it is obvious that the peaks have to be located inside the rRNA cluster and are not just falsely assigned due to their location close to centromeric satellites.

For H2B, PRM1 and PRM2 ChIPs of sperm heads HOMER collectively observed less LINES and SINEs as expected. The class of LINES (long interspersed nuclear elements) in the HOMER annotation showed only LINE-1 retrotransposons, while the SINE (short interspersed nuclear elements) group includes mainly *Alu* elements. Transposable elements have been found to be expressed in the germline and may regulate gene expression (S. Zhou et al., 2023). Silencing of LINE-1 activation early in pre-implantation mouse embryos has shown to lead to developmental delay and that the re-activation of LINE-1 transposable elements is associated with fertility across generations (Lisner & Kimmins, 2023). Furthermore, LINE-1 retrotransposons are responsible for the mobilisation of *Alu* elements (Kohlrausch et al., 2022). *Alu* elements are also found to play a role in gene expression (Chen & Yang, 2017). Interestingly, LINE-1 elements were found to interact with a spermiogenesis-specific histone variant (HILS1) (Lisner & Kimmins, 2023).

Regarding an intergenerational effect due to the nucleoprotamine and nucleohistone association with LINEs and SINEs, it is striking to see, that LINE-1s are (together with other interspersed repeats) abundantly transcribed in the zygote and early developmental stages. Apparently and among others LINE-1s are key regulators of the zygote genome activation (ZGA) e.g. in mice as antagonists to MERVL sequences (Ansaloni et al., 2023). It is therefore tempting to speculate that the underrepresentation of LINEs in H2B-, PRM1- and PRM2-ChIPs might represent a kind of poising for transcription-like effect in the paternal pronucleus of the zygote. Surely more work is needed extending our experiments on H2B, e.g. analysing activating histone marks and their association with LINEs in sperm heads, to better understand the respective epigenetic inheritance linked to the male pronucleus.

Looking more closely at the peak distribution of the genomic regions, the number of peaks was more precisely looked at. The majority of peaks map to satellite regions (~28 % on average in the H2B samples, 35 % in the PRM1 samples and 50 % on average in the PRM2 samples), followed by intergenic regions (22 % on average in the H2B samples, 23 % in the PRM1 samples and 18 % in the PRM2 samples) which aligns with the finding of Yamaguchi et al. (Yamaguchi et al., 2018) in mice where H4 is predominantly localised in intergenic regions and H3K9me3 is preferentially associated with heterochromatin and satellite regions, possibly with the known role of H3K9me3 mediating repression of repetitive elements (Nicetto & Zaret, 2019). However, Yamaguchi et al. also found histones in repeat regions which align with my results that 14 % of the H2B peaks were assigned to simple repeat regions (17 % in PRM1 and PRM2). As histone-coverage points towards active transcription, it is well known that centromeric transcripts exist. Centromeric transcription is even of vital importance for the proper formation of CENP-A containing centromeric chromatin and downstream functions (Kixmoeller et al., 2020). These so-called centromere-derived RNAs or cenRNAs were identified several decades ago, however only recently received more attention e.g. in the context of promoting tumorigenesis (reviewed in Corless et al., 2020).

About 0.3 % of the peaks of all samples map to rRNA regions (PRM1 0.4 %). In eukaryotic cells only a fraction of rDNA genes is active, while others are silenced due to epigenetic marks (Yu et al., 2015). In the HOMER annotation only a small number of peaks (0 – 3) was assigned to rRNA regions compared to other genomic regions (as mentioned in the paragraph above). The $\log_2(\text{observed/expected})$ value is relatively high compared to other genomic regions (e.g. higher than the \log_2 value for the satellite region). This result hints towards the

interpretation that the H2B-, PRM1- and PRM2-ChIP data contains more DNA sequences mapping specifically to rRNA genomic regions than we would normally expect. Considering the Fraction of Reads in Peaks, the distribution of the sequences seems more broad and less specific. Therefore, other annotation tools should be used to confirm or contradict the HOMER annotation. Interestingly, there are no genomic regions showing protamine-specific peak enrichment to the exclusion of H2B peak enrichment. This can be explained by the fact that residual histones are embedded between protamine toroidal loops (reviewed by Jenkins & Carrell, 2012) and are probably evenly distributed along the genome.

The sometimes-contrasting results between the examined proteins and the individuals, even though an overall uniform pattern can be seen, might result from the fact that bulk analyses were performed with a great number of sperm heads. It might be that sperm cells from the same individual differ from one another regarding their histone and protamine localisation or that not all sperm heads containing in the ejaculate samples were fully matured and completed their histone-to-protamine transition. The high number of sperm cells per sample most likely resulted in an average distribution of the proteins along the genome. This might also explain the low number of peaks in the ChIP experiments for PRM1 and PRM2. If each sperm cell shows slightly different histone and protamine patterns, the obtained sequences associated with the protein of interest might overlap broadly and lead to unclear peaks during the annotation. To obtain more sensitive results and to verify if those slight differences between individuals are due to biological variance or if this variance can be found even inside one individual single-cell analysis should be considered. Single-cell ChIP-Seq could help to distinguish if residual histones bind to the same genomic regions (or close by) as protamines or if the overlaps occur due to the high number of examined cells. Histone packaging is inevitably enclosed by protamine packaging and depending on the size of the DNA fragments generated during the chromatin immunoprecipitation and the resulting peak sizes during peak calling the obtained peaks will overlap to a certain degree but to find out to which degree they potentially overlap analysing single sperm heads might be a good option. One difficulty regarding this approach might be the accessibility of the sperm nuclei together with their differentiation. Grosselin et al. (2019) designed a microfluidics single-cell ChIP-Seq protocol including the compartmentalization of single cells with lysis reagents and micrococcal nuclease (MNase) in oil droplets, the production of hydrogel beads, the compartmentalization of single hydrogel beads in oil droplets and the one-to-one fusion of the droplets containing the digested nucleosomes with the hydrogel beads-containing droplets. To what extent the shearing of the sperm heads to open them up using

ultrasonication is functional or if the oil is more of a hindrance has to be tested. Sonicating the samples before the compartmentalization with the lysis reagents and MNase might hamper the proper labeling of the individual cells and therefore influence the possibility to distinguish one cell from another. Another suitable approach might be CUT&RUN sequencing (Cleavage Under Targets & Release Using Nuclease; Skene & Henikoff, 2017). CUT&RUN is similar to ChIP-Seq in that both identify genomic locations of chromatin-associated proteins but CUT&RUN has some advantages over ChIP-Seq. CUT&RUN uses Protein A and G-bound micrococcal nuclease (pAG-MNase) to selectively cleave antibody-labelled chromatin. It requires fewer cells and sequencing reads and shows lesser background. Using lesser cells could already give more information about general trends in chromatin packaging confirming or contradicting the ChIP results. As sperm heads and their chromatin have a very complex structure the functionality of the possible implementation of these methods has to be tested.

To look more closely at the protein packaging of genomic regions associated with different ncRNA classes and the influence they might have on the active transcription of those ncRNAs different ncRNA genomic regions were compared to the motifs of H2B, PRM1 and PRM2. The most overlapping regions were found with lncRNA genomic region, whereby the H2B samples had more overlaps with lncRNA genomic regions than the PRM1 or PRM2 samples. These findings suggest that a great part of the lncRNA genomic regions can be transcribed actively, respectively, early in the zygotic development. Only minor overlaps with rRNA regions could be found and no overlaps with tRNA regions which align with the results of the HOMER annotation. Interestingly, the annotation with INTERSECT resulted in more overlaps with rRNA regions than the HOMER annotation. The annotation with HOMER seems to be a proper tool to get a first general overview about the peak distribution and their associated gene regions but INTERSECT seems to get more precise results when looking at specific regions. If these distribution levels and patterns hold truth or if they result from the mean distribution of thousands of cells due to the bulk approach has to be tested.

Protamines are highly intrinsically disordered protein with 60 % of PRM1 and 100 % of PRM2 being intrinsically disordered. H2B on the other hand, has two intrinsically disordered regions that make up 42 % of the whole protein. Intrinsically disordered regions have a variety of different functions: they may act as entropic chains, for example as flexible linkers allowing the movement of domains which are positioned on either ends of the linker relative to each other or as spacers regulating the distance between domains, they are subject to posttranslational modifications which increases the interaction potential and the functional

state in which a protein can exist in a cell, they take part in chaperone activity assisting the proper folding of RNA and proteins, they might function as effectors interacting with other proteins and therefore modifying their activity, they might function as assemblers to bring multiple binding partners together to promote the formation of higher-order protein complexes or they might function as scavengers storing and neutralizing small ligands like ATP. Additionally, they expose short linear peptide motifs that enable interactions with structured domains in other proteins (Van Der Lee et al., 2014).

These disordered regions additionally possess the ability to bind RNA. To find out if these proteins bind to specific ncRNA classes including lncRNA, tRNA and sno/miRNA RNA immunoprecipitation experiments were carried out and the obtained sequences were compared to lncRNA, tRNA and snoRNA genes. Only minor tRNA genes could be assigned to the RNA sequences bound by H2B, PRM1 and PRM2 resulting in an average of 0.01 % of alignable sequences assigned to tRNAs. Slightly more sequences could be assigned to sno/miRNAs with an average of 0.07 % of the alignable sequences. Out of those three RNA classes most of the obtained RIP-Seq sequences could be assigned to lncRNAs. On average, 4.6 % of all alignable sequences could be assigned to lncRNA genes (H2B 6.5 %, PRM1 2.9 % and PRM2 4.6 %). The higher percentage of H2B-associated sequences assigned to lncRNA genes is caused by one individual that differs greatly from the other three individuals: 17.7 % of the alignable sequences of Individual 1 were assigned to lncRNA genes while 2.9 %, 2.4 % and 3.0 % of the alignable sequences of Individual 2, Individual 3 and Individual 4 were assigned to lncRNA genes respectively. The percentages of the PRM1-associated sequences assigned to lncRNA genes are more uniform with 2.7 %, 2.5 %, 2.9 % and 3.4 % for the individuals 1 to 4. The alignable sequences associated to PRM2 show again more quantitative differences between individuals; 3.3 % of the alignable sequences of Individual 1 were assigned to lncRNA genes, 2.7 % of the alignable sequences of Individual 2 could be assigned to lncRNA genes, 6.7 % of the alignable sequences of Individual 3 were assigned to lncRNA genes and 5.3 % of the alignable sequences of Individual 4 were assigned to lncRNA genes. The four individuals show great quantitative and qualitative variability regarding the hits to lncRNA genes. The binding capacity of intrinsically disordered protein regions seems to be not exclusively but predominantly assigned to lncRNAs. Two lncRNA genes were found in all samples (except Individual 2) in the upper 25 % read counts of all hits, those are lnc-LRR1-1 and lnc-NEMF-1. One lncRNA gene was found in the upper 25 % read counts of all protamine samples (except Individual 2), this lncRNA is lnc-CCNB1IP1-1. The majority of lncRNAs found in each sample differ greatly

in terms of abundance, showing both qualitatively and quantitatively a high variance in the lncRNA distribution. lnc-LRR1-1 is an RNA component of the signal recognition particle (SRP) 7SL1, which associates with the ribosome and targets freshly synthesized proteins to the endoplasmic reticulum for secretion or membrane insertion. lnc-NEMF-1 is also an RNA component of the 7SL, more precisely the 7SL3 RNA. lnc-CCNB1IP1-1 is the RNA component of the RNase P ribonucleoprotein, an endoribonuclease that forms the 5' termini of mature tRNA by cleaving tRNA precursor molecules. This diversity in the abundance of lncRNA genes across examined proteins and individuals suggest an unspecified binding of lncRNAs to the intrinsically disordered regions of H2B, PRM1 and PRM2. Those lncRNAs might be important for maintenance of a functional structure and stability of the proteins in the sense of solubility and phase transitions (Polymenidou, 2018) or they could play an important role in the zygote and therefore being selectively bound by histone and protamine.

3 General discussion

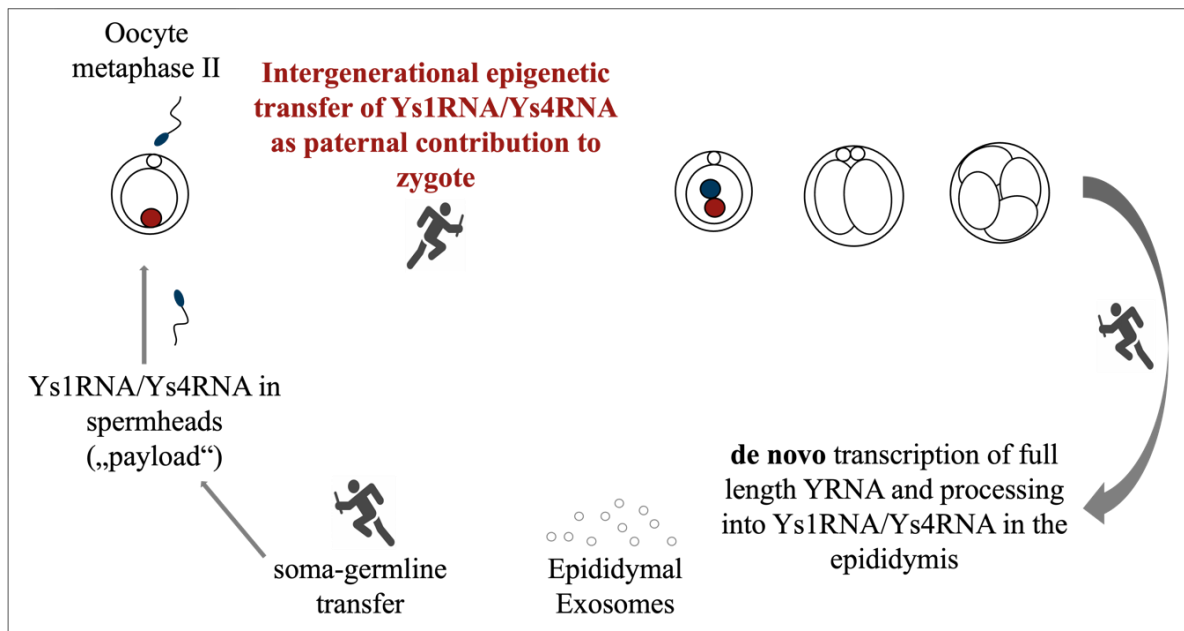
Epigenetic regulatory mechanisms are crucial for embryogenesis and development. They have the potential to mediate gene regulation across generations. Although, acquired traits can be written down in the epigenome of a cell, they cannot easily be transmitted from one generation to the next. For this to occur, epigenetic changes have to manifest in germ cells as well, being unaffected by erasure and reprogramming. The germline undergoes natural epigenetic reprogramming during embryonic development (Hajkova, 2011): epigenetic marks are removed and reset converting germ cells into stem cells. Without this reprogramming germ cells would retain the parental epigenetic memory, which would prevent the transmission of genetic information to the offspring (Sabour & Schöler 2012). My working hypothesis is that an erasure of epigenetic marks during the germline reprogramming might not be fully efficient, hence leading to the intergenerational inheritance of epialleles. Over the last decades the molecular and mechanistic correlates underlying the idea of acquired and possible heritable characters were intensively researched on. Self-sustaining feedback loops, chromatin-based mechanisms including DNA methylation and both coding and non-coding RNA as well as structural templating are generally accepted molecular mechanisms that have the potential to render environmental influence into gene expression regulation (Heard & Martienssen, 2014).

Summarising current research in metazoans, it is widely accepted that the oocyte carries the majority of relevant RNAs that function in the early embryo of most species. However, there is substantial and increasing evidence that sperm also carry a functional RNA payload. Theoretically, non-coding RNAs in the male germline can affect subsequent generations either directly via delivery to the zygote, or indirectly by directing chromatin or DNA modifications during spermatogenesis (Bošković & Rando, 2018). It is therefore of utmost importance to precisely describe the complete RNA cargo of male gametes and get experimental evidence to the functions of small RNAs in early embryo. In this study I did so by analysing the ncRNA cargo in sperm heads of six healthy men using RNA-seq and smallRNA-seq and evaluating the resulting readouts with different annotation tools. Certain trends could be observed: tRNAs and rRNAs are quite abundant in sperm heads with them being the predominant ncRNA classes in four out of 6 individuals (19.06 % of sequences mapped to rRNA on average with individual percentages ranging from 10.4 % to 29.6 % and an average of 9.76 % of the sequences mapped to tRNA with individual percentages ranging from 1.5 % to 15.2 %). One group of small ncRNA that was found in high abundance in the class of miscellaneous RNA is derived from Y RNA.

Human Y RNAs have a length around 100 nt (Gulìa et al., 2020). The 5' end and 3' end typically bind together to form a double-stranded stem domain divided in a lower and upper stem domain (Kowalski & Krude, 2015). The upper stem domain is important for chromosomal DNA replication and the lower stem domain has a Ro60 binding site and is, when bound to Ro60, involved in RNA stability and stress response (Kowalski & Krude, 2015). Y RNAs do not only exist in their full length but also in fragments between 25 nt and 35 nt (Röther & Meister, 2011). Y RNAs and Y RNA fragments are reported to be possible tumor biomarkers and play a significant role in several tumor types (Guglas et al., 2020). The fragmentation of Y RNAs, which is performed by RNase L (Donovan et al., 2017), is increasing in apoptotic cells and upon activation of the innate immune system (Nicolas et al., 2012; Rutjes et al., 1999) but was also found in non-apoptotic proliferating cells (Nicolas et al., 2012) and within extracellular vesicles (Driedonks & Nolte-T'Hoen, 2019).

Interestingly, hY1 was found to be the most abundant RNA by mapping the sperm head transcripts to the human reference genome hg38 and is almost completely absent in human oocytes. However, not the complete Y RNA sequences could be traced in sperm heads but sequences of predominantly 30 nt and 31 nt generated from the 5' end of the source Y RNA. Additionally, these Y RNA fragments possess a 2'-O-methylation at their 3' end. Tracing the time point when YsRNAs are enriched in sperm cells testis RNA and epididymis data was examined. I found that almost no YsRNAs are found in testis and that the abundance of YsRNAs increases late in spermiogenesis. At the same time YsRNAs can be traced in the caput and corpus of the epididymis, with the most abundant YsRNA there being Ys4RNA. During spermiogenesis the spermatid nuclei decrease and epigenetic changes, in particular chromatin remodelling, take place. Here, YsRNAs are most probably integrated into the sperm head. YsRNAs are massively found in sperm seminal fluid exosomes. Interestingly, the abundance of the different YsRNAs depends on the size enrichment of the performed RNA-Seq: the 20 – 40 nt fractions contain predominantly Ys4RNA, while the 40 – 100 nt fraction contains predominantly Ys3RNA. These findings are interesting because, even though exosomes and epididymis do contain way more YsRNAs in general compared to sperm heads, the pattern of the different YsRNA distribution looks different. These findings suggest that there is a selective enrichment of Ys1RNAs and Ys4RNAs in sperm heads. In this context, I hypothesize that the intergenerational epigenetic transfer of Ys1RNA and Ys4RNA represents a paternal contribution to the zygote. hY1 and hY4 are transcribed and processed to Ys1RNA and Ys4RNA, respectively, in the epididymis and stored in exosomes. These exosomes are integrated into the sperm head where they then are integrated into the

zygote and passed on to the next generation. One known function of Y RNAs is their ability to bind to Ro60 due to a specific binding site in their stem loop structure. Ro60 binds to misfolded non-coding RNA, like pre-5S rRNA, and regulates their expression. Binding of Y RNA to this protein hinders the proper function of Ro60. Therefore, RNA bound to Ro60 was investigated on and Y RNAs could only be traced in very low number. I hypothesize that the YsRNAs in sperm cells hinder the proper folding of Y RNAs into their typical hairpin structure by binding to the stem domain and therefore ensure the functionality of Ro60. To summarize my results shortly, I propose a relay race of genetic and epigenetic inheritance taking place in the epididymis via vesicles as graphically outlined in the following:



The transcripts of sperm heads and testis differ greatly from one another. Although the majority of transcripts in sperm heads and testis are associated to post-transcriptional gene silencing, in testis this process is mediated by ncRNA (compare figure 6). Interestingly, a great part of the sperm head transcriptome is assigned to the sperm-specific proteins PRM1 and PRM2 as well as a transition protein that replaces histones temporarily before being replaced by protamine. To first describe the abundance of protamine along the human genome ChIP-Seq experiments with sperm heads samples of four individuals were carried out for PRM1 and PRM2. To compare the estimated localisation of protamine to residual histones ChIP-Seq was performed with an antibody against the core histone H2B. I found that the majority of peaks (enriched alignment reads in areas in the genome) generated via

peak calling overlap between all three investigated proteins and that they preferably bind to centromeric and telomeric region. Satellite DNA and short tandem repetitive sequence motifs are preferably found in these regions. Generally said, most peaks were found in genomic areas of repeating sequences. One striking result from the HOMER annotation was the high enrichment of motifs in rRNA genomic regions, although the number of peaks assigned to that region was relatively small in all samples compared to other genomic regions. As rRNA makes up about 20 % of the ncRNA in sperm heads, other ncRNAs that could be traced in higher abundance in sperm heads were looked at more closely. Therefore, the genomic regions transcribing rRNA, tRNA, sno/miRNA and lncRNA were compared to the peaks of H2B, PRM1 and PRM2. Only small overlaps could be found with rRNA and sno/miRNA genomic regions (no overlaps with tRNA genomic regions were found) but lncRNA genomic regions showed a high number of overlaps, especially in H2B samples, suggesting that these RNAs can be actively transcribed. Looking closely at the tertiary structure of those proteins showed that all three proteins have intrinsically disordered regions. H2B does have two intrinsically disordered regions, PRM1 has one intrinsically disordered region, while PRM2 is completely disordered and therefore referred to as intrinsically disordered protein. One of the functions of these intrinsically disordered regions is the ability to bind RNA. I found that the majority of ncRNAs bound to H2B, PRM1 and PRM2 belong to the class of lncRNAs. Interestingly, the lncRNAs bound by these proteins showed great variety between examined proteins but also between individuals. These findings suggest selected binding of lncRNAs to those proteins but an unspecific binding inside this RNA class. Those lncRNAs might be important for the structure and stability of histones and protamines. Also, they might play an important role in developmental processes in the zygote and therefore being selectively bound by histone and protamine.

Summarising the given results, ncRNAs are selectively enriched and transcribed in human sperm heads and constitute either an RNA payload with sequence specific function (Y RNA) or “unspecific” (lncRNA) cargo delivered to the zygote. To what extent YsRNA and lncRNAs have a direct or indirect impact on the proper development of the zygote needs to be further investigated on. However, it’s tempting to speculate that the molecular adaptive phenotype is associated with the RNA chaperoning capacity in the zygote, with a role of YsRNAs in reducing the Y RNA-Ro60-binding. At least metadata from in vitro reproductive assistance techniques (ICSI) point towards improved reproduction – or positive selection and adaptation – with sperm that underwent an epididymal passage, where the transfer between soma and germline takes place (Tournaye et al., 1996).

A second class of RNA payload delivered by the sperm head and into the zygote was characterised as lncRNA. In contrast to the situation I could find for YsRNA, the lncRNAs are different in their profiles ruling out a sequence-specific function. The exact relation between lncRNAs and histones and protamines should therefore be investigated in more detail and involve multiple samples to get an idea about the heterogeneous lncRNA profiles and their likely intergenerational epigenetic effect. Epigenetic marks in sperm cells like histone tail modifications, programmatic histone retention as well as DNA methylation and formation of DNA demethylation intermediates were shown to be associated with poor spermatogenesis, decreased fertility and fertilisation ability, embryo quality as well as pregnancy outcome (Jenkins & Carrell, 2012). Keeping that in mind, it is important to first get a better understanding if lncRNAs influence the structure and solubility of proteins or if the intrinsically disordered regions in histones and protamines just function as “sponge” for the lncRNAs until protamine-to-histone transition takes place in the zygote resulting in the release of those lncRNAs. To find out if lncRNAs are involved in structure maintenance knock down approaches using RNAi or antisense LNA GapmeRs should be applied in sperm cells of different developmental stages before and during histone-to-protamine transition. To what extent these approaches can be applied to sperm heads, considering their complex structure, has to be tested.

As ethic restrictions prohibit experiments with human embryos beyond 14 days of development only epigenetic marks that have a direct impact on the zygote can be traced and investigated on. Switching to model organisms like mice might be an adequate solution for some experimental approaches but unfortunately not for all. For example, the Y RNA abundance and distribution in mouse sperm cells differ greatly from that in human sperm cells. As the transcriptome, ncRNA cargo and chromatin packaging might differ greatly between different organisms it is of utmost importance to expand the research on human sperm heads to be able to take the biomedical implications of the epigenetic status into account. In this respect, my work provides a first general overview about the ncRNA cargo in human sperm cells that differs greatly from that in testis and human oocytes. This holds especially for paternally contributed YsRNA implying an important role of these ncRNAs for the proper embryonic development. Moreover, I could describe an epigenetic intergenerational effect of histone H2B and protamines. I found a correlation between sperm nucleoprotamine and nucleohistones, in that H2B and protamines largely overlap in covering the same regions of the sperm genome. Also, I could show that these proteins bind lncRNAs. These interactions were mostly unspecific, however with the potential of influencing the

sperm chromatin with respect to its physico-chemical properties. If there are defined lncRNAs that play an important role as paternal contribution and in embryonic development remains not settled due to the huge variability in the respective profiles and has thus to await further studies.

Literature

- Altemose, N., Logsdon, G. A., Bzikadze, A. V., Sidhwani, P., Langley, S. A., Caldas, G. V., Hoyt, S. J., Uralsky, L., Ryabov, F. D., Shew, C. J., Sauria, M. E. G., Borchers, M., Gershman, A., Mikheenko, A., Shepelev, V. A., Dvorkina, T., Kunyavskaya, O., Vollger, M. R., Rhie, A., ... Miga, K. H. (2022). Complete genomic and epigenetic maps of human centromeres. *Science*, 376(6588).
<https://doi.org/10.1126/science.abl4178>
- Ansaloni, F., Gustincich, S., & Sanges, R. (2023). In silico characterisation of minor wave genes and LINE-1s transcriptional dynamics at murine zygotic genome activation. *Frontiers in Cell and Developmental Biology*, 11.
<https://doi.org/10.3389/fcell.2023.1124266>
- Arimbasseri, A. G., & Maraia, R. J. (2016). RNA Polymerase III Advances: Structural and tRNA Functional Views. In *Trends in Biochemical Sciences* (Vol. 41, Issue 6).
<https://doi.org/10.1016/j.tibs.2016.03.003>
- Aubert, M., O'donohue, M. F., Lebaron, S., & Gleizes, P. E. (2018). Pre-ribosomal RNA processing in human cells: From mechanisms to congenital diseases. In *Biomolecules* (Vol. 8, Issue 4). <https://doi.org/10.3390/biom8040123>
- Balhorn, R. (2007). The protamine family of sperm nuclear proteins. In *Genome Biology* (Vol. 8, Issue 9). <https://doi.org/10.1186/gb-2007-8-9-227>
- Bali, P., Im, H. I., & Kenny, P. J. (2011). Methylation, memory and addiction. *Epigenetics*, 6(6). <https://doi.org/10.4161/epi.6.6.15905>
- Bao, J., & Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. In *Reproduction* (Vol. 151, Issue 5).
<https://doi.org/10.1530/REP-15-0562>
- Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. In *Cell* (Vol. 116, Issue 2). [https://doi.org/10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5)
- Bline, A. P., Goff, A. Le, & Allard, P. (2020). What is lost in the weismann barrier? In *Journal of Developmental Biology* (Vol. 8, Issue 4).
<https://doi.org/10.3390/jdb8040035>

- Bošković, A., & Rando, O. J. (2018). Transgenerational epigenetic inheritance. In *Annual Review of Genetics* (Vol. 52). <https://doi.org/10.1146/annurev-genet-120417-031404>
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell*, *128*(6). <https://doi.org/10.1016/j.cell.2007.01.043>
- Brewer, L. R., Corzett, M., & Balhorn, R. (1999). Protamine-induced condensation and decondensation of the same DNA molecule. *Science*, *286*(5437). <https://doi.org/10.1126/science.286.5437.120>
- Brunner, A. M., Nanni, P., & Mansuy, I. M. (2014). Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics and Chromatin*, *7*(1). <https://doi.org/10.1186/1756-8935-7-2>
- Burkhardt, R. W. (2013). Lamarck, evolution, and the inheritance of acquired characters. *Genetics*, *194*(4). <https://doi.org/10.1534/genetics.113.151852>
- Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaisch, A., Padovan-Merhar, O., Regev, A., Rinn, J. L., & Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology*, *16*(1). <https://doi.org/10.1186/s13059-015-0586-4>
- Chen, L. L., & Yang, L. (2017). ALU alternative Regulation for Gene Expression. In *Trends in Cell Biology* (Vol. 27, Issue 7). <https://doi.org/10.1016/j.tcb.2017.01.002>
- Cheng, H., Shang, D., & Zhou, R. (2022). Correction To: Germline stem cells in human (Signal Transduction and Targeted Therapy, (2022), 7, 1, (345), 10.1038/s41392-022-01197-3). In *Signal Transduction and Targeted Therapy* (Vol. 7, Issue 1). <https://doi.org/10.1038/s41392-022-01258-7>
- Corless, S., Höcker, S., & Erhardt, S. (2020). Centromeric RNA and Its Function at and Beyond Centromeric Chromatin. In *Journal of Molecular Biology* (Vol. 432, Issue 15). <https://doi.org/10.1016/j.jmb.2020.03.027>
- Cornwall, G. A., & Von Horsten, H. H. (2007). Sperm maturation in the epididymis: Role of segment-specific micro environments. In *The Genetics of Male Infertility*. https://doi.org/10.1007/978-1-59745-176-5_13

- Donovan, J., Rath, S., Kolet-Mandrikov, D., & Korennykh, A. (2017). Rapid RNase L–driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery. *RNA*, *23*(11). <https://doi.org/10.1261/rna.062000.117>
- Driedonks, T. A. P., & Nolte-T’Hoen, E. N. M. (2019). Circulating Y-RNAs in extracellular vesicles and ribonucleoprotein complexes; Implications for the immune system. In *Frontiers in Immunology* (Vol. 10, Issue JAN). <https://doi.org/10.3389/fimmu.2018.03164>
- Dubocanin, D., Cortes, A. E. S., Ranchalis, J., Real, T., Mallory, B., & Stergachis, A. B. (2022). Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin. *BioRxiv*.
- Duttke, S. H., Chang, M. W., Heinz, S., & Benner, C. (2019). Identification and dynamic quantification of regulatory elements using total RNA. *Genome Research*, *29*(11). <https://doi.org/10.1101/gr.253492.119>
- Dym, M., Kokkinaki, M., & He, Z. (2009). Spermatogonial stem cells: Mouse and human comparisons. In *Birth Defects Research Part C - Embryo Today: Reviews* (Vol. 87, Issue 1). <https://doi.org/10.1002/bdrc.20141>
- Glenney, J. R., Glenney, P., & Weber, K. (1983). Mapping the fodrin molecule with monoclonal antibodies. A general approach for rod-like multidomain proteins. *Journal of Molecular Biology*, *167*(2). [https://doi.org/10.1016/S0022-2836\(83\)80336-2](https://doi.org/10.1016/S0022-2836(83)80336-2)
- Gong, J., Wang, P., Liu, J. C., Li, J., Zeng, Q. X., Yang, C., Li, Y., Yu, D., Cao, D., & Duan, Y. G. (2022). Integrative Analysis of Small RNA and mRNA Expression Profiles Identifies Signatures Associated With Chronic Epididymitis. *Frontiers in Immunology*, *13*. <https://doi.org/10.3389/fimmu.2022.883803>
- Gou, L. T., Dai, P., Yang, J. H., Xue, Y., Hu, Y. P., Zhou, Y., Kang, J. Y., Wang, X., Li, H., Hua, M. M., Zhao, S., Hu, S. Da, Wu, L. G., Shi, H. J., Li, Y., Fu, X. D., Qu, L. H., Wang, E. D., & Liu, M. F. (2014). Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Research*, *24*(6). <https://doi.org/10.1038/cr.2014.41>
- Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyat, F., Frenoy, O., Pousse, Y., Reichen, M., Woolfe, A., Brennan, C., Griffiths, A. D., Vallot, C., & Gérard, A. (2019). High-throughput single-cell

- ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nature Genetics*, 51(6). <https://doi.org/10.1038/s41588-019-0424-9>
- Guglas, K., Kołodziejczak, I., Kolenda, T., Kopczyńska, M., Teresiak, A., Sobocińska, J., Bliźniak, R., & Lamperska, K. (2020). YRNAs and YRNA-derived fragments as new players in cancer research and their potential role in diagnostics. In *International Journal of Molecular Sciences* (Vol. 21, Issue 16). <https://doi.org/10.3390/ijms21165682>
- Gulìa, C., Signore, F., Gaffi, M., Gigli, S., Votino, R., Nucciotti, R., Bertacca, L., Zaami, S., Baffa, A., Santini, E., Porrello, A., & Piergentili, R. (2020). Y RNA: An overview of their role as potential biomarkers and molecular targets in human cancers. In *Cancers* (Vol. 12, Issue 5). <https://doi.org/10.3390/cancers12051238>
- Hajkova, P. (2011). Epigenetic reprogramming in the germline: Towards the ground state of the epigenome. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 366, Issue 1575). <https://doi.org/10.1098/rstb.2011.0042>
- Hargitt, G. T. (1944). What is germ plasm? *Science*, 100(2599). <https://doi.org/10.1126/science.100.2599.343>
- Heard, E., & Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: Myths and mechanisms. In *Cell* (Vol. 157, Issue 1). <https://doi.org/10.1016/j.cell.2014.02.045>
- Heller, C. G., & Clermont, Y. (1963). Spermatogenesis in man: An estimate of its duration. *Science*, 140(3563). <https://doi.org/10.1126/science.140.3563.184>
- Hendrick, S. S. (1981). Self-disclosure and marital satisfaction. *Journal of Personality and Social Psychology*, 40(6). <https://doi.org/10.1037//0022-3514.40.6.1150>
- Hess, R. A., & De Franca, L. R. (2008). Spermatogenesis and cycle of the seminiferous epithelium. *Advances in Experimental Medicine and Biology*, 636. https://doi.org/10.1007/978-0-387-09597-4_1
- Hofgärtner, F. J., Schmid, M., Krone, W., Zenzes, M. T., & Engel, W. (1979). Pattern of activity of nucleolus organizers during spermatogenesis in mammals as analyzed by silver-staining. *Chromosoma*, 71(2). <https://doi.org/10.1007/BF00292823>

- Hud, N. V., Allen, M. J., Downing, K. H., Lee, J., & Balhorn, R. (1993). Identification of the Elemental Packing Unit of DNA in Mammalian Sperm Cells by Atomic Force Microscopy. *Biochemical and Biophysical Research Communications*, *193*(3).
<https://doi.org/10.1006/bbrc.1993.1773>
- Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-interacting RNA: Its biogenesis and functions. *Annual Review of Biochemistry*, *84*. <https://doi.org/10.1146/annurev-biochem-060614-034258>
- Jan, S. Z., Vormer, T. L., Jongejan, A., Röling, M. D., Silber, S. J., de Rooij, D. G., Hamer, G., Repping, S., & van Pelt, A. M. M. (2017). Unraveling transcriptome dynamics in human spermatogenesis. *Development (Cambridge)*, *144*(20).
<https://doi.org/10.1242/dev.152413>
- Jenkins, T. G., & Carrell, D. T. (2012). Dynamic alterations in the paternal epigenetic landscape following fertilization. In *Frontiers in Genetics* (Vol. 3, Issue JUL).
<https://doi.org/10.3389/fgene.2012.00143>
- Jeon, H., Lee, H., Kang, B., Jang, I., & Roh, T. Y. (2020). Comparative analysis of commonly used peak calling programs for chip-seq analysis. *Genomics and Informatics*, *18*(4). <https://doi.org/10.5808/GI.2020.18.4.E42>
- Johnson, L., Varner, D. D., Roberts, M. E., Smith, T. L., Keillor, G. E., & Scrutchfield, W. L. (2000). Efficiency of spermatogenesis: A comparative approach. *Animal Reproduction Science*, *60–61*. [https://doi.org/10.1016/S0378-4320\(00\)00108-1](https://doi.org/10.1016/S0378-4320(00)00108-1)
- Karolchik, D., Hinricks, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC table browser data retrieval tool. *Nucleic Acids Research*, *32*(DATABASE ISS.). <https://doi.org/10.1093/nar/gkh103>
- Khan, A., & Mathelier, A. (2017). Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*, *18*(1).
<https://doi.org/10.1186/s12859-017-1708-7>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8). <https://doi.org/10.1038/s41587-019-0201-4>

- Kixmoeller, K., Allu, P. K., & Black, B. E. (2020). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. In *Open Biology* (Vol. 10, Issue 6). <https://doi.org/10.1098/rsob.200051>
- Kohlrausch, F. B., Berteli, T. S., Wang, F., Navarro, P. A., & Keefe, D. L. (2022). Control of LINE-1 Expression Maintains Genome Integrity in Germline and Early Embryo Development. In *Reproductive Sciences* (Vol. 29, Issue 2). <https://doi.org/10.1007/s43032-021-00461-1>
- Kotaja, N. (2013). Spermatogenesis, Mouse. In *Brenner's Encyclopedia of Genetics* (pp. 529–532). Elsevier. <https://doi.org/10.1016/B978-0-12-374984-0.01461-3>
- Kowalski, M. P., & Krude, T. (2015). Functional roles of non-coding Y RNAs. In *International Journal of Biochemistry and Cell Biology* (Vol. 66). <https://doi.org/10.1016/j.biocel.2015.07.003>
- Kubota, H., & Brinster, R. L. (2018). Spermatogonial stem cells. In *Biology of Reproduction* (Vol. 99, Issue 1). <https://doi.org/10.1093/biolre/iocy077>
- Kuksa, P. P., Amlie-Wolf, A., Katanic, Z., Valladares, O., Wang, L. S., & Leung, Y. Y. (2019). DASHR 2.0: Integrated database of human small non-coding RNA genes and mature products. *Bioinformatics*, 35(6). <https://doi.org/10.1093/bioinformatics/bty709>
- Kumar, P., Kuscu, C., & Dutta, A. (2016). Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). In *Trends in Biochemical Sciences* (Vol. 41, Issue 8). <https://doi.org/10.1016/j.tibs.2016.05.004>
- Kwapisz, M., & Morillon, A. (2020). Subtelomeric Transcription and its Regulation. In *Journal of Molecular Biology* (Vol. 432, Issue 15). <https://doi.org/10.1016/j.jmb.2020.01.026>
- Lerner, M. R., Boyle, J. A., Hardin, J. A., & Steitz, J. A. (1981). Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science*, 211(4480). <https://doi.org/10.1126/science.6164096>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7). <https://doi.org/10.1093/bioinformatics/btt656>

- Lismer, A., & Kimmins, S. (2023). Emerging evidence that the mammalian sperm epigenome serves as a template for embryo development. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-37820-2>
- Littauer, U. Z., Giveon, D., Thierauf, M., Ginzburg, I., & Ponstingl, H. (1986). Common and distinct tubulin binding sites for microtubule-associated proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 83(19). <https://doi.org/10.1073/pnas.83.19.7162>
- Liu, Y. (2008). A new perspective on Darwin's Pangenesis. In *Biological Reviews* (Vol. 83, Issue 2). <https://doi.org/10.1111/j.1469-185X.2008.00036.x>
- Lyons, S. M., Fay, M. M., & Ivanov, P. (2018). The role of RNA modifications in the regulation of tRNA cleavage. In *FEBS Letters* (Vol. 592, Issue 17). <https://doi.org/10.1002/1873-3468.13205>
- MacFarlane, L.-A., & R. Murphy, P. (2010). MicroRNA: Biogenesis, Function and Role in Cancer. *Current Genomics*, 11(7). <https://doi.org/10.2174/138920210793175895>
- Mäkelä, J. A., & Toppari, J. (2017). Spermatogenesis. In *Endocrinology (Switzerland)*. https://doi.org/10.1007/978-3-319-44441-3_13
- Maraia, R. J., Sasaki-tozawa, N., Driscoll, C. T., Green, E. D., & Darlington, G. J. (1994). The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. *Nucleic Acids Research*, 22(15). <https://doi.org/10.1093/nar/22.15.3045>
- Maraia, R., Sakulich, A. L., Brinkmann, E., & Green, E. D. (1996). Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic Acids Research*, 24(18). <https://doi.org/10.1093/nar/24.18.3552>
- Marques, B., Matos, R., & Martinho, R. G. (2018). Regulation of the oocyte epigenome during prophase i arrest. In *Encyclopedia of Reproduction*. <https://doi.org/10.1016/B978-0-12-801238-3.64457-4>
- Misell, L. M., Holochwost, D., Boban, D., Santi, N., Shefi, S., Hellerstein, M. K., & Turek, P. J. (2006). A stable isotope-mass spectrometric method for measuring human spermatogenesis kinetics in vivo. *Journal of Urology*, 175(1). [https://doi.org/10.1016/S0022-5347\(05\)00053-4](https://doi.org/10.1016/S0022-5347(05)00053-4)

- Nicetto, D., & Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. In *Current Opinion in Genetics and Development* (Vol. 55). <https://doi.org/10.1016/j.gde.2019.04.013>
- Nicolas, F. E., Hall, A. E., Csorba, T., Turnbull, C., & Dalmay, T. (2012). Biogenesis of y RNA-derived small RNAs is independent of the microRNA pathway. *FEBS Letters*, 586(8). <https://doi.org/10.1016/j.febslet.2012.03.026>
- Nikitina, T. V., & Tishchenko, L. I. (2005). RNA polymerase III transcription machinery: Structure and transcription regulation. In *Molecular Biology* (Vol. 39, Issue 2). <https://doi.org/10.1007/s11008-005-0024-x>
- Nilsson, E. E., Maamar, M. Ben, & Skinner, M. K. (2020). Environmentally induced epigenetic transgenerational inheritance and the weismann barrier: The dawn of neo-lamarckian theory. In *Journal of Developmental Biology* (Vol. 8, Issue 4). <https://doi.org/10.3390/jdb8040028>
- Oliva, R., & Castillo, J. (2011). Proteomics and the genetics of sperm chromatin condensation. In *Asian Journal of Andrology* (Vol. 13, Issue 1). <https://doi.org/10.1038/aja.2010.65>
- Page, S. L., & Hawley, R. S. (2003). Chromosome choreography: The meiotic ballet. In *Science* (Vol. 301, Issue 5634). <https://doi.org/10.1126/science.1086605>
- Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., Lei, L., Han, C., Ning, L., Cao, Y., Zhou, Q., Chen, Q., & Duan, E. (2012). A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. In *Cell Research* (Vol. 22, Issue 11). <https://doi.org/10.1038/cr.2012.141>
- Perez, M. F., & Lehner, B. (2019). Intergenerational and transgenerational epigenetic inheritance in animals. In *Nature Cell Biology* (Vol. 21, Issue 2). <https://doi.org/10.1038/s41556-018-0242-9>
- Piucò, R., & Galante, P. A. F. (2021). piRNADB: A piwi-interacting RNA database. *BioRxiv*.
- Polymenidou, M. (2018). The RNA face of phase separation. In *Science* (Vol. 360, Issue 6391). <https://doi.org/10.1126/science.aat8028>

- Qin, Q., Mei, S., Wu, Q., Sun, H., Li, L., Taing, L., Chen, S., Li, F., Liu, T., Zang, C., Xu, H., Chen, Y., Meyer, C. A., Zhang, Y., Brown, M., Long, H. W., & Liu, X. S. (2016). ChiLin: A comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. *BMC Bioinformatics*, *17*(1). <https://doi.org/10.1186/s12859-016-1274-4>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6). <https://doi.org/10.1093/bioinformatics/btq033>
- Raina, M., & Ibba, M. (2014). TRNAs as regulators of biological processes. In *Frontiers in Genetics* (Vol. 5, Issue JUN). <https://doi.org/10.3389/fgene.2014.00171>
- Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. In *Nature Reviews Molecular Cell Biology* (Vol. 19, Issue 3). <https://doi.org/10.1038/nrm.2017.104>
- Röther, S., & Meister, G. (2011). Small RNAs derived from longer non-coding RNAs. In *Biochimie* (Vol. 93, Issue 11). <https://doi.org/10.1016/j.biochi.2011.07.032>
- Rutjes, S. A., Van Der Heijden, A., Utz, P. J., Van Venrooij, W. J., & Pruijn, G. J. M. (1999). Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. *Journal of Biological Chemistry*, *274*(35). <https://doi.org/10.1074/jbc.274.35.24799>
- Sabour, D., & Schöler, H. R. (2012). Reprogramming and the mammalian germline: The Weismann barrier revisited. In *Current Opinion in Cell Biology* (Vol. 24, Issue 6). <https://doi.org/10.1016/j.ceb.2012.08.006>
- Scheuren, M., Möhner, J., & Zischler, H. (2023). R-loop landscape in mature human sperm: Regulatory and evolutionary implications. *Frontiers in Genetics*, *14*. <https://doi.org/10.3389/fgene.2023.1069871>
- Schon, S. B., Luense, L. J., Wang, X., Bartolomei, M. S., Coutifaris, C., Garcia, B. A., & Berger, S. L. (2019). Histone modification signatures in human sperm distinguish clinical abnormalities. *Journal of Assisted Reproduction and Genetics*, *36*(2). <https://doi.org/10.1007/s10815-018-1354-7>
- Shen, Y., Yu, X., Zhu, L., Li, T., Yan, Z., & Guo, J. (2018). Transfer RNA-derived fragments and tRNA halves: biogenesis, biological functions and their roles in

- diseases. In *Journal of Molecular Medicine* (Vol. 96, Issue 11).
<https://doi.org/10.1007/s00109-018-1693-y>
- Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *ELife*, 6.
<https://doi.org/10.7554/eLife.21856>
- Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., & Bohnsack, M. T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. In *RNA Biology* (Vol. 14, Issue 9).
<https://doi.org/10.1080/15476286.2016.1259781>
- Smoluch, M., Mielczarek, P., Drabik, A., & Silberring, J. (2016). Online and Offline Sample Fractionation. In *Proteomic Profiling and Analytical Chemistry: The Crossroads: Second Edition*. <https://doi.org/10.1016/B978-0-444-63688-1.00005-7>
- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., deMassy, B., Mikkelsen, T. S., & Kaessmann, H. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Reports*, 3(6).
<https://doi.org/10.1016/j.celrep.2013.05.031>
- Steger, K., Pauls, K., Klonisch, T., Franke, F. E., & Bergmann, M. (2000). Expression of protamine-1 and -2 mRNA during human spermiogenesis. *Molecular Human Reproduction*, 6(3). <https://doi.org/10.1093/molehr/6.3.219>
- Stults, D. M., Killen, M. W., Pierce, H. H., & Pierce, A. J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Research*, 18(1).
<https://doi.org/10.1101/gr.6858507>
- Tafforeau, L., Zorbas, C., Langhendries, J. L., Mullineux, S. T., Stamatopoulou, V., Mullier, R., Wacheul, L., & Lafontaine, D. L. J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Molecular Cell*, 51(4).
<https://doi.org/10.1016/j.molcel.2013.08.011>
- Talbert, P. B., & Henikoff, S. (2022). The genetics and epigenetics of satellite centromeres. *Genome Research*, 32(4). <https://doi.org/10.1101/gr.275351.121>

- Tanphaichitr, N., Sobhon, P., Taluppeth, N., & Chalermisarachai, P. (1978). Basic nuclear proteins in testicular cells and ejaculated spermatozoa in man. *Experimental Cell Research*, *117*(2). [https://doi.org/10.1016/0014-4827\(78\)90148-9](https://doi.org/10.1016/0014-4827(78)90148-9)
- Tournaye, H., Liu, J., Nagy, P. Z., Camus, M., Goossens, A., Silber, S., Van Steirteghem, A. C., & Devroey, P. (1996). Correlation between testicular histology and outcome after intracytoplasmic sperm injection using testicular spermatozoa. *Human Reproduction*, *11*(1). <https://doi.org/10.1093/oxfordjournals.humrep.a019004>
- Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. In *Chemical Reviews* (Vol. 114, Issue 13). <https://doi.org/10.1021/cr400525m>
- Villegas, J., Araya, P., Bustos-Obregon, E., & Burzio, L. O. (2002). Localization of the 16S mitochondrial rRNA in the nucleus of mammalian spermatogenic cells. *Molecular Human Reproduction*, *8*(11). <https://doi.org/10.1093/molehr/8.11.977>
- Vojtech, L., Woo, S., Hughes, S., Levy, C., Ballweber, L., Sauteraud, R. P., Strobl, J., Westerberg, K., Gottardo, R., Tewari, M., & Hladik, F. (2014). Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Research*, *42*(11). <https://doi.org/10.1093/nar/gku347>
- Volders, P. J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., & Vandesompele, J. (2019). Lncipedia 5: Towards a reference set of human long non-coding rnas. *Nucleic Acids Research*, *47*(D1). <https://doi.org/10.1093/nar/gky1031>
- Weinhold, B. (2006). Epigenetics: the science of change. *Environmental Health Perspectives*, *114*(3). <https://doi.org/10.1289/ehp.114-a160>
- Yamaguchi, K., Hada, M., Fukuda, Y., Inoue, E., Makino, Y., Katou, Y., Shirahige, K., & Okada, Y. (2018). Re-evaluating the Localization of Sperm-Retained Histones Revealed the Modification-Dependent Accumulation in Specific Genome Regions. *Cell Reports*, *23*(13). <https://doi.org/10.1016/j.celrep.2018.05.094>
- Yang, Q., Li, R., Lyu, Q., Hou, L., Liu, Z., Sun, Q., Liu, M., Shi, H., Xu, B., Yin, M., Yan, Z., Huang, Y., Liu, M., Li, Y., & Wu, L. (2019). Single-cell CAS-seq reveals a class of

- short PIWI-interacting RNAs in human oocytes. *Nature Communications*, 10(1).
<https://doi.org/10.1038/s41467-019-11312-8>
- Yu, F., Shen, X., Fan, L., & Yu, Z. (2015). Analysis of histone modifications at human ribosomal DNA in liver cancer cell. *Scientific Reports*, 5.
<https://doi.org/10.1038/srep18100>
- Zhang, L., Lu, Q., & Chang, C. (2020). Epigenetics in Health and Disease. In *Advances in Experimental Medicine and Biology* (Vol. 1253). https://doi.org/10.1007/978-981-15-3449-2_1
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9).
<https://doi.org/10.1186/gb-2008-9-9-r137>
- Zhou, S., Sakashita, A., Yuan, S., & Namekawa, S. H. (2023). Retrotransposons in the Mammalian Male Germline. In *Sexual Development* (Vol. 16, Issues 5–6).
<https://doi.org/10.1159/000520683>
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1).
<https://doi.org/10.1038/s41467-019-09234-6>
- Zhu, L., Liu, X., Pu, W., & Peng, Y. (2018). tRNA-derived small non-coding RNAs in human disease. In *Cancer Letters* (Vol. 419).
<https://doi.org/10.1016/j.canlet.2018.01.015>

4 Appendix

4.1 Figures

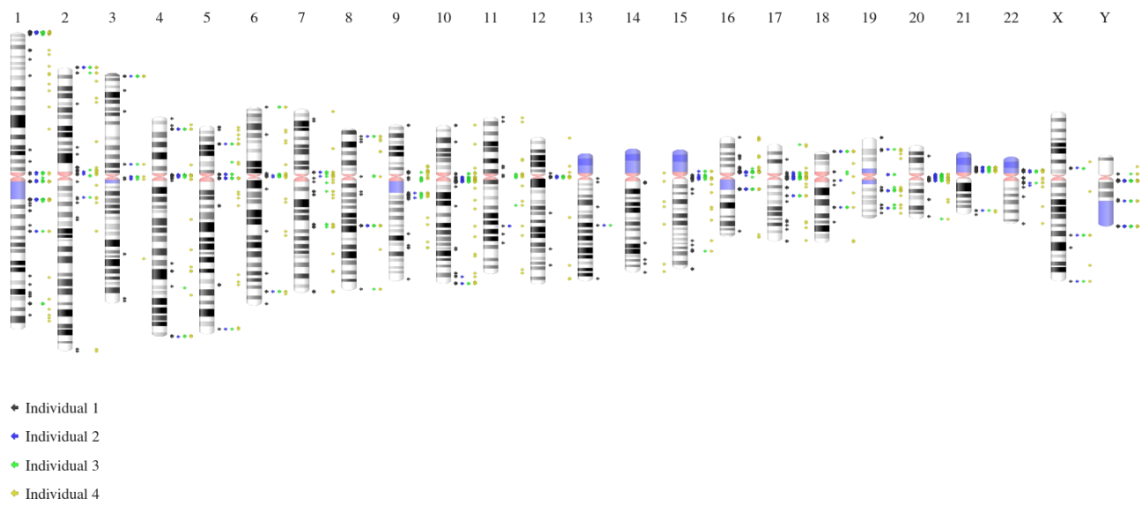


Figure 1: Idiogram of H2B peaks of all four individuals across the human genome. Peaks of Individual 1 are shown as black arrows, peaks of Individual 2 are shown as blue arrows, peaks of Individual 3 are shown as green arrows and peaks of Individual 4 are shown as yellow arrows.

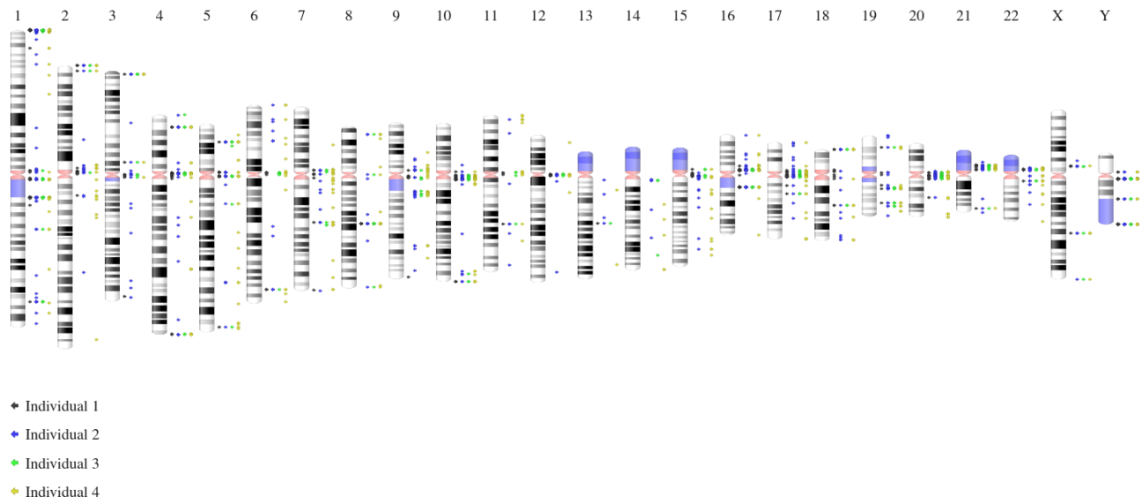


Figure 2: Idiogram of PRM1 peaks of all four individuals across the human genome. Peaks of Individual 1 are shown as black arrows, peaks of Individual 2 are shown as blue arrows, peaks of Individual 3 are shown as green arrows and peaks of Individual 4 are shown as yellow arrows.

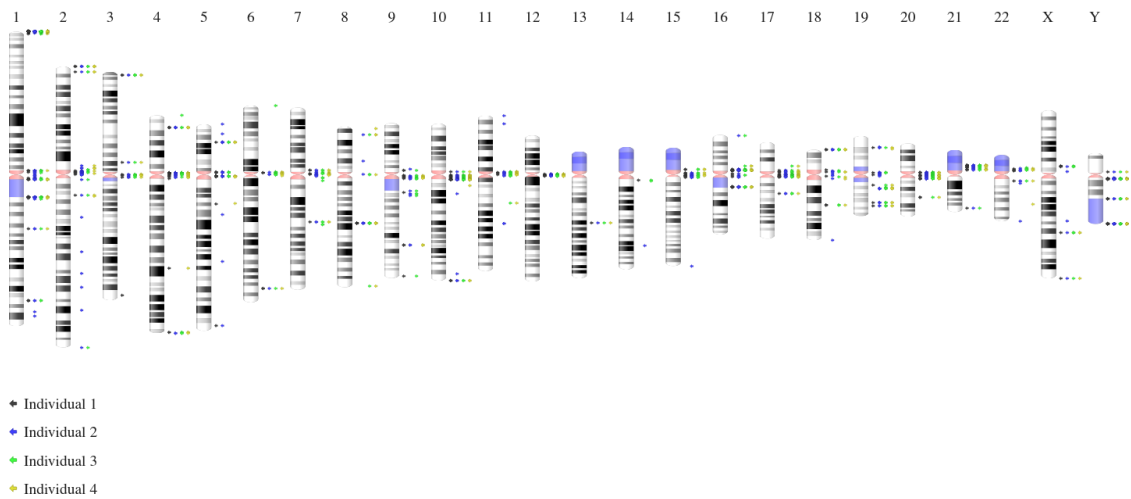


Figure 3: Idiogram of PRM2 peaks of all four individuals across the human genome. Peaks of Individual 1 are shown as black arrows, peaks of Individual 2 are shown as blue arrows, peaks of Individual 3 are shown as green arrows and peaks of Individual 4 are shown as yellow arrows.

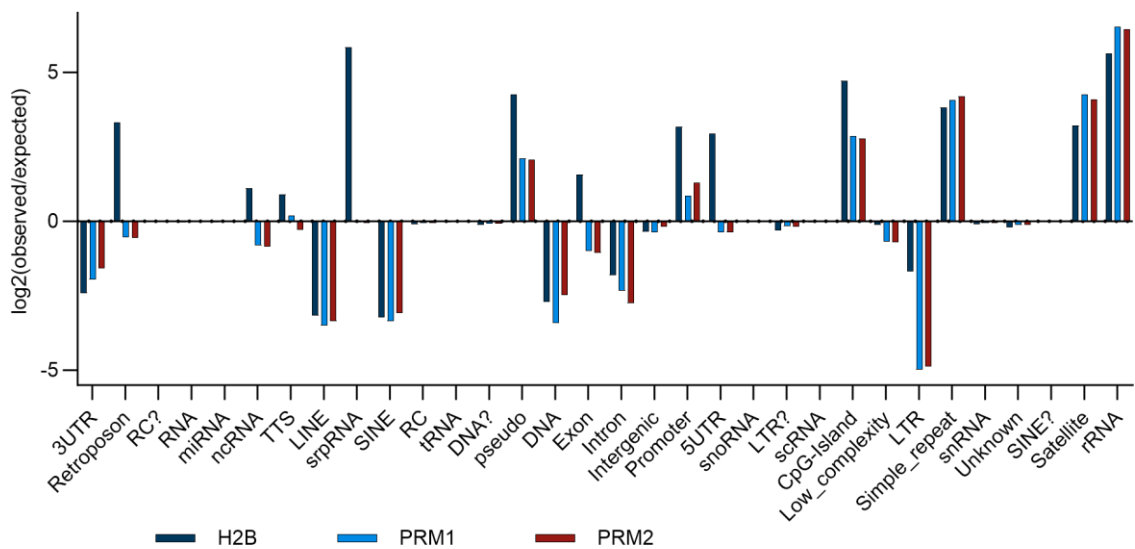


Figure 4: HOMER annotation results for H2B, PRM1 and PRM2 peaks of Individual 1. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed/expected})$ on the y-axis. Results for H2B are shown in dark blue, results for PRM1 are shown in light blue and PRM2 results are shown in dark red.

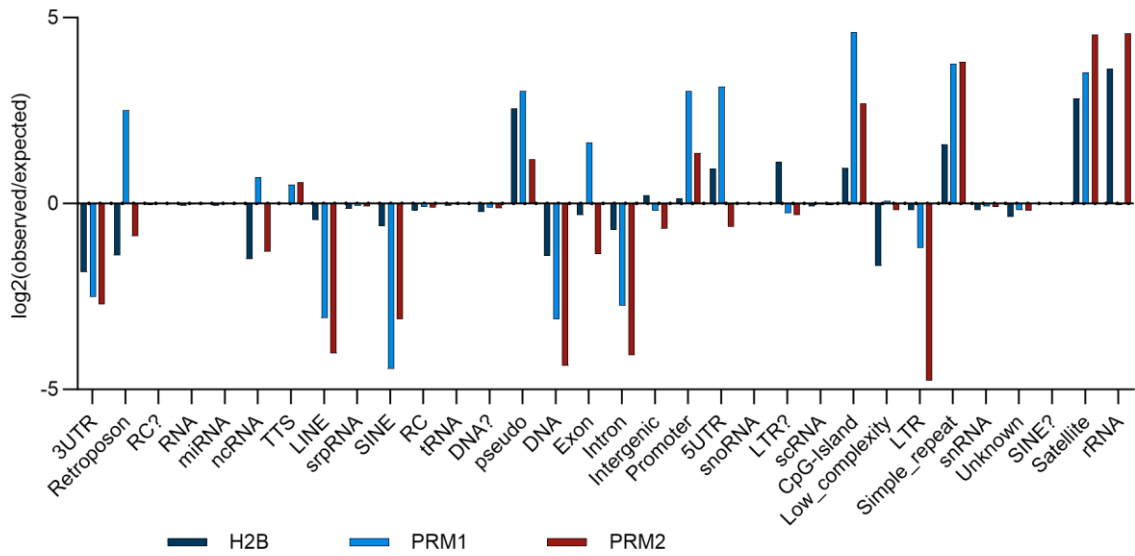


Figure 5: HOMER annotation results for H2B, PRM1 and PRM2 peaks of Individual 2. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed/expected})$ on the y-axis. Results for H2B are shown in dark blue, results for PRM1 are shown in light blue and PRM2 results are shown in dark red.

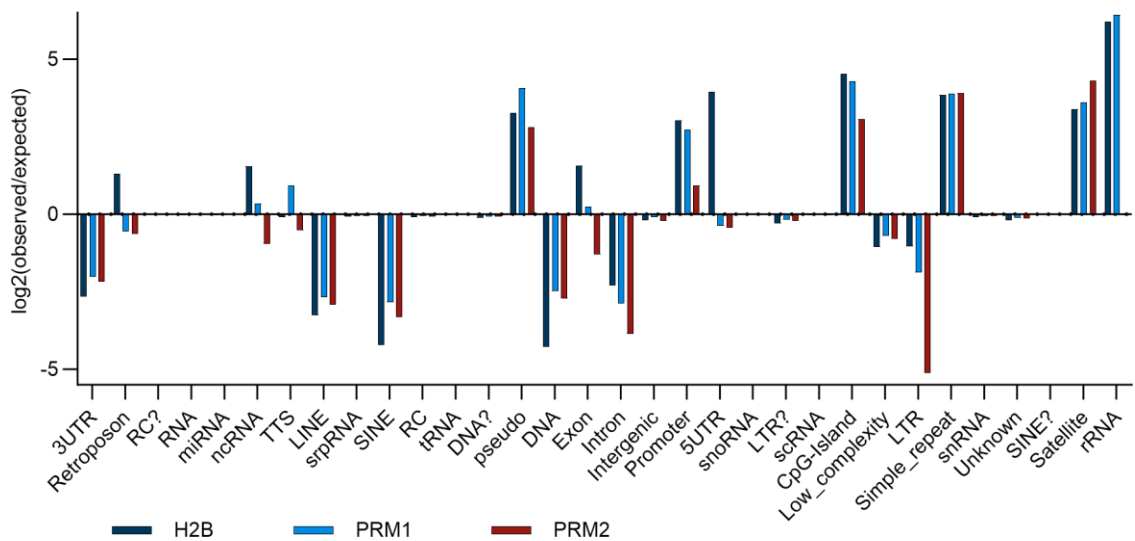


Figure 6: HOMER annotation results for H2B, PRM1 and PRM2 peaks of Individual 4. Genomic features are displayed on the x-axis and the peak ratio is given as $\log_2(\text{observed/expected})$ on the y-axis. Results for H2B are shown in dark blue, results for PRM1 are shown in light blue and PRM2 results are shown in dark red.

4.2 Tables

Table 1: Amount of overlaps with lncRNA genomic regions for each protein and individual

lncRNA	Individual 1	Individual 2	Individual 3	Individual 4
H2B	383	72	148	399
PRM1	89	276	99	365
PRM2	109	169	167	118

Table 2: Amount of overlaps with piRNA genomic regions for each protein and individual

piRNA	Individual 1	Individual 2	Individual 3	Individual 4
H2B	180	9	23	43
PRM1	10	66	32	66
PRM2	10	34	23	22

Table 3: Amount of overlaps with rRNA genomic regions for each protein and individual

rRNA	Individual 1	Individual 2	Individual 3	Individual 4
H2B	4	0	6	5
PRM1	4	0	5	5
PRM2	4	3	6	0

Table 4: Amount of overlaps with tRNA genomic regions for each protein and individual

tRNA	Individual 1	Individual 2	Individual 3	Individual 4
H2B	0	0	0	0
PRM1	0	0	0	0
PRM2	0	0	0	0

Table 5: Amount of overlaps with sno/miRNA genomic regions for each protein and individual

sno/miRNA	Individual 1	Individual 2	Individual 3	Individual 4
H2B	3	1	1	4
PRM1	1	4	1	3
PRM2	0	1	2	2

Table 6: lncRNA expression profiles for H2B-, PRM1- and PRM2-bound RNAs of Individual 1

H2B		PRM1		PRM2	
GeneID	Hits	GeneID	Hits	GeneID	Hits
lnc-LRR1-1	109280	lnc-LRR1-1	5423	lnc-LRR1-1	26238
lnc-NCOA3-5	91628	lnc-NEMF-1	2421	lnc-NEMF-1	18240
lnc-KDM8-1	56322	lnc-INPP4B-2	1528	MALAT1	9921
lnc-NCOA3-10	45340	lnc-FAM174A-6	1469	lnc-HLA-DMA-1	8300
lnc-TNK2-2	41559	lnc-BOD1-4	972	lnc-CDC42BPA-5	6659
lnc-PREX1-2	41497	lnc-DAPP1-2	972	lnc-HOOK1-4	6358
lnc-TP53BP2-8	16485	lnc-NECAB1-1	967	LINC00200	6315
LINC00917	16193	lnc-LYPD6-2	909	lnc-NEFL-1	6181
lnc-NEMF-2	15072	lnc-GJA10-24	887	lnc-ARHGAP10-5	6093
lnc-APOD-4	13717	lnc-ABCD4-1	850	LRRC75A-AS1	6002
lnc-ARF6-1	13592	lnc-KIFAP3-2	849	lnc-EVI5-1	5847
TUG1	12803	lnc-KBTBD3-3	846	lnc-U2SURP-3	5840
SORCS3-AS1	12654	LINC00355	775	lnc-CCNB1IP1-1	5775
lnc-USP53-3	12198	lnc-RPL39L-3	760	lnc-VWA3B-3	5764
lnc-SYTL4-2	10182	lnc-WRN-9	752	lnc-WFS1-3	5741
lnc-C5orf60-1	9927	lnc-MARCKS-1	737	lnc-CCDC96-1	5660
lnc-NEMF-1	9899	lnc-KDM4C-18	714	lnc-COL28A1-1	5594
CNTFR-AS1	9105	lnc-ESRRB-2	699	GABPB1-IT1	5594
LINC00664	9089	lnc-ANGPTL2-2	696	lnc-IL13RA2-3	5559
LINC00608	8807	lnc-REL-6	684	lnc-GPR65-8	5408
lnc-TDRD3-20	8736	LINC01516	681	lnc-MARF1-1	5332
PWRN1	8550	lnc-TAGLN-1	669	lnc-EHD3-3	5314
lnc-LAT-2	8549	lnc-ADRA1D-4	648	lnc-GLRX-4	5312
lnc-SCARB1-1	8125	lnc-DAD1-2	647	LINC00917	5208
lnc-RUBCN-2	8103	lnc-IFNK-10	627	LINC01714	5159
lnc-RFX2-1	7978	LINC01938	617	LINC00649	5125
lnc-ZNF716-10	7693	lnc-ATP6V0A1-4	614	lnc-ZNF800-4	5106
lnc-ARFGEF2-2	7515	ISPD-AS1	606	lnc-UBLCP1-2	5009
LINC00906	7184	PCBP1-AS1	606	lnc-FGGY-10	4917
lnc-DLK1-35	6991	lnc-ACTR2-3	597	lnc-ICA1-3	4806

lnc-ATP6V1D-8	6900	lnc-RORA-2	597	lnc-ERGIC1-1	4775
lnc-APOD-5	6890	lnc-DEIRA-1	596	lnc-CDKN2C-1	4668
lnc-VAX1-4	6812	lnc-ZNF780B-3	593	lnc-APOD-4	4342
lnc-FYTTD1-6	6711	LINC02275	588	lnc-PLCG2-6	4328
lnc-REN-2	6376	LINC01629	572	lnc-MAT2B-2	4313
LINC01005	6271	lnc-AASDHPPT-2	569	lnc-NGFR-3	4252
lnc-SERPIND1-1	6037	lnc-TREML2-2	564	lnc-RORA-2	4244
lnc-TMEM248-2	5965	lnc-AIG1-8	561	TMEM161B-AS1	4238
lnc-SHARPIN-4	5898	PLCE1-AS1	558	lnc-CCDC125-6	4178
lnc-RNF38-1	5789	lnc-GALNT6-4	557	LHFPL3-AS2	4126
lnc-CLEC18B-1	5671	lnc-LRP2BP-1	556	LINC02127	4098
lnc-BRCA1-3	5531	VIPR1-AS1	554	lnc-MLX-2	3978
ARHGAP22-IT1	5008	lnc-MUC1-2	552	ENTPD1-AS1	3773
lnc-BTD-2	4870	LINC01592	549	lnc-FBXL16-3	3741
lnc-RPIA-1	4864	lnc-STOML3-5	545	lnc-THOC3-5	3640
lnc-PASK-2	4832	lnc-IRF4-2	542	lnc-TMEM167B-1	3540
lnc-APBB1IP-1	4815	lnc-ROPN1B-13	541	lnc-DERL1-3	3499
lnc-MBD3-3	4762	lnc-FRG2-13	537	lnc-PPIL1-2	3484
lnc-METTL14-3	4757	lnc-CCNB1IP1-1	536	lnc-FILIP1L-8	3484
lnc-OR4F16-15	4738	lnc-PGM3-3	534	lnc-CHD2-23	3474
lnc-RAB34-1	4664	lnc-ARHGAP26-6	532	lnc-TSC22D1-2	3461
lnc-SIRT7-2	4610	lnc-PPP2R2B-3	511	lnc-CTAGE1-9	3437
LINC01471	4590	lnc-TMEM30B-1	510	lnc-TCP1-6	3422
lnc-PYGO2-2	4555	lnc-TBC1D22A-4	509	lnc-PTPA-3	3358
lnc-HLA-DMA-1	4552	LINC01010	506	LINC01358	3203
PTOV1-AS1	4544	lnc-TPBG-9	503	lnc-MISP3-1	3186
lnc-PLA2G1B-2	4362	lnc-KCNJ15-2	501	lnc-LRP8-3	3178
SPATA42	4340	lnc-LRRC38-6	497	lnc-PPP1R3B-11	3164
LINC01095	4330	lnc-GPR149-10	497	lnc-BRCA1-3	3151
lnc-OR4F29-7	4288	lnc-RNF166-1	494	MIR29B2CHG	3145
lnc-TMEM248-4	4186	lnc-IRF2BPL-6	494	lnc-LIPG-7	3144
lnc-ZNF727-17	4137	HIF1A-AS2	492	lnc-HAAO-7	3140
LINC01820	4091	lnc-CPXM2-1	489	SOX2-OT	3099

LINC01921	4059	lnc-RASGRF1-6	482	lnc-CARNMT1-2	3058
lnc-FAM124B-4	4056	LINC01619	480	lnc-PTPRU-9	3033
lnc-FBXO43-4	4048	lnc-PTPRE-3	480	lnc-CPM-4	3028
lnc-LRRTM4-6	4045	lnc-EVI2A-3	477	lnc-PSMG1-11	3015
lnc-IGLL1-1	4042	lnc-C2orf91-9	475	lnc-ARHGAP28-7	3008
lnc-RNF39-4	4036	lnc-NREP-7	475	lnc-FAM193B-6	3005
lnc-ULBP2-2	3998	lnc-MNS1-8	470	lnc-VRK2-19	3001
lnc-TPGS2-3	3926	lnc-USP25-14	469	lnc-CDIPT-1	3000
lnc-GPR137B-1	3915	lnc-NEIL3-6	468	CNTN4-AS1	2988
lnc-FAM193B-6	3889	lnc-SLC7A9-4	465	ITGA9-AS1	2984
lnc-PPM1G-1	3868	LINC00886	464	lnc-MANSC1-4	2955
lnc-COIL-3	3852	lnc-ZNF486-12	464	lnc-PPFIA2-1	2953
lnc-STYXL1-2	3829	lnc-DNAJC6-2	463	lnc-C18orf54-2	2928
lnc-IL11RA-2	3750	lnc-AGMO-1	462	lnc-PDE6C-5	2901
lnc-OSBPL1A-1	3718	HDAC2-AS2	457	lnc-VGLL3-7	2889
lnc-ANKRD10-4	3649	lnc-PGPEP1-3	457	lnc-TNNT1-1	2888
LINC01606	3644	lnc-EBAG9-2	453	lnc-ABCA5-8	2886
lnc-TUT1-2	3597	lnc-RPS24-5	453	lnc-RPS24-16	2872
LINC00174	3586	lnc-SPI1-1	452	lnc-ZNF592-2	2835
lnc-C1QTNF3-AMACR-3	3550	lnc-KRT80-2	450	lnc-FLT3-2	2833
lnc-FBXO43-6	3528	lnc-HHAT-3	450	lnc-CEP170-13	2831
IBA57-DT	3466	lnc-DIAPH3-19	449	lnc-MYEF2-3	2828
PCAT6	3454	LINC00906	448	LINC01588	2808
PRNT	3450	MIR205HG	447	LINC01709	2802
lnc-THBS2-5	3449	lnc-PABPC4L-14	447	lnc-THBS1-2	2801
lnc-CEP170-13	3435	WWTR1-AS1	445	lnc-PROC-1	2796
lnc-ALG2-6	3428	lnc-TMBIM4-6	443	lnc-MGARP-1	2788
lnc-ST6GALNAC2-2	3423	lnc-CCL18-1	442	EGOT	2751
lnc-PGGT1B-8	3312	lnc-INTS2-4	433	lnc-PREX2-2	2751
lnc-APOD-6	3306	lnc-SLC1A1-7	433	lnc-ONECUT3-1	2748
lnc-CYTL1-1	3289	lnc-KDM1A-1	432	lnc-KCNC3-1	2742

lnc-CDH4-4	3260	lnc-CLCN3-8	431	lnc-ARHGAP26-4	2737
RGMB-AS1	3189	lnc-RASA1-24	429	HDAC11-AS1	2735
PWRN2	3188	lnc-ATP6V1D-8	427	lnc-AMZ2-5	2723
lnc-SORB3-1	3170	lnc-NR3C1-3	424	lnc-TMEM209-1	2711
lnc-RPRML-3	3163	lnc-FOXG1-16	422	LINC02233	2708
lnc-TP53I11-1	3151	lnc-ZAR1L-6	420	RGMB-AS1	2705

Table 7: lncRNA expression profiles for H2B-, PRM1- and PRM2-bound RNAs of Individual 2

H2B		PRM1		PRM2	
GeneID	Hits	GeneID	Hits	GeneID	Hits
lnc-ERMN-4	696	lnc-RNF39-4	7475	CHKB-DT	9562
LINC02104	630	lnc-IRX3-80	6009	lnc-IL22RA2-2	8083
lnc-CHST10-2	542	LINC01524	5982	lnc-DFFB-50	7787
ABCA9-AS1	536	lnc-MYOM2-4	5497	lnc-NOC2L-6	7178
lnc-PLA2G4A-7	533	LINC01914	5425	lnc-TMEM45B-2	6934
lnc-ZNF717-4	532	lnc-RBM12B-2	4944	lnc-TXLNB-3	6931
lnc-LOXL3-3	528	LINC00472	4912	LINC00929	6805
lnc-EIF2AK1-4	517	MIR646HG	4890	lnc-DHX35-6	6759
LINC01920	502	lnc-ADRA2C-5	4710	lnc-C17orf80-7	6692
lnc-CACNA1E-2	502	lnc-TBC1D22A-4	4653	lnc-ZNF680-4	6665
lnc-NBPF1-8	496	lnc-CUL2-3	4577	lnc-SHPRH-6	6605
lnc-GALNT15-8	496	lnc-IL12RB2-1	4563	lnc-IRX2-8	6555
lnc-FAM174A-6	474	lnc-NAPB-2	4542	lnc-CCDC17-1	6491
lnc-MARCH8-2	471	lnc-AP3S2-1	4108	lnc-ZNF781-2	6263
lnc-SPRY1-11	469	lnc-SPRY1-10	4046	lnc-TSHZ2-6	5919
lnc-GRM7-2	467	lnc-GMNC-5	3968	FIRRE	5906
lnc-GDF10-3	446	lnc-TBX5-3	3841	lnc-GRM4-1	5874
lnc-KLF5-11	437	lnc-C6orf222-4	3811	lnc-EZH2-1	5867
lnc-SAT2-1	436	lnc-KANSL1-4	3745	lnc-FANCI-5	5861
lnc-GRSF1-1	434	lnc-SEMA3A-2	3705	lnc-PLA2G4A-7	5850
lnc-SEL1L3-9	434	MIR4435-2HG	3656	lnc-EEPDI-49	5813
lnc-KLHL24-2	430	lnc-CNPY1-1	3651	lnc-HMBOX1-5	5805
lnc-PTPRU-4	415	lnc-ESM1-3	3633	lnc-RAN-4	5672

lnc-SLCO5A1-11	410	lnc-IBTK-6	3620	LINC00237	5667
lnc-XKRX-1	410	lnc-MFSD8-6	3612	lnc-TMTC3-18	5560
lnc-BHLHE22-10	408	LINC01317	3611	lnc-ALG2-6	5508
lnc-DCTD-65	406	lnc-MED27-1	3609	lnc-SLC26A11-1	5482
lnc-RBM11-14	391	lnc-IGSF6-1	3598	lnc-SEC31B-4	5438
lnc-SIPA1L1-2	389	lnc-MTFMT-1	3596	lnc-SLC5A2-2	5395
LINC01128	388	GAS5	3592	FZD10-DT	5318
lnc-HABP4-2	385	lnc-CTBP2-6	3587	lnc-TXNDC2-8	5317
lnc-MEF2A-10	382	LINC02099	3579	lnc-ZBTB7C-10	5309
lnc-KBTBD13-4	377	lnc-INTS9-2	3574	lnc-NID1-2	5215
lnc-ACACA-3	375	SUCLG2-AS1	3555	LINC02015	5203
lnc-STMP1-4	375	lnc-SLC38A4-1	3552	lnc-MSTN-2	5191
NR2F2-AS1	374	LINC00898	3540	lnc-RAD52-2	5098
lnc-PPP2R3C-1	366	KCNMA1-AS1	3529	lnc-CSRP2-2	5082
lnc-IRF2BP2-11	363	lnc-MEIS3-1	3524	lnc-CST9L-2	5037
lnc-EBF1-3	360	lnc-GPT2-1	3520	lnc-SSTR4-3	4881
LINC01666	356	RALY-AS1	3518	lnc-PIK3R1-11	4783
HNF1A-AS1	351	ST7-AS1	3505	lnc-COL28A1-6	4755
lnc-KCNRG-1	351	lnc-TGFBRAP1-15	3493	DLX6-AS1	4637
lnc-FGFR1OP2-1	345	lnc-XRCC2-13	3492	lnc-CCNB1IP1-1	4345
lnc-FAM160A1-1	343	lnc-CYP1B1-1	3490	lnc-SERP2-3	4330
lnc-NEBL-3	342	lnc-SOD1-9	3473	lnc-GPLD1-2	4151
lnc-TRPV2-4	331	ITFG2-AS1	3464	lnc-C19orf66-1	4103
LHFPL3-AS1	325	lnc-NBN-1	3463	LINC00326	4075
lnc-ACMSD-1	323	lnc-TMEM178A-6	3457	lnc-CEP170-13	3863
lnc-LZTS1-1	311	lnc-CEP41-1	3455	lnc-TBC1D22A-4	3857
lnc-SLC39A11-10	310	lnc-RACGAP1-1	3434	CBR3-AS1	3837
lnc-IRAK4-5	299	lnc-ALK-1	3424	lnc-UTP23-10	3700
lnc-MBOAT4-4	296	LINC01621	3418	lnc-NAT1-3	3662
lnc-SEC16B-7	293	lnc-CCDC125-1	3415	lnc-CD109-3	3593
lnc-ATP6V1D-8	291	lnc-MPHOSPH8-22	3403	RBM15-AS1	3554
lnc-TMEM167A-3	291	lnc-BTBD10-2	3399	TNFRSF14-AS1	3539
lnc-GADD45A-9	289	lnc-CCNJ-1	3398	lnc-FEZ2-7	3530

lnc-HERC1-2	289	lnc-DLK1-35	3395	FAM53B-AS1	3510
lnc-ATP13A3-1	288	lnc-CEBPB-12	3388	lnc-ITPA-2	3449
lnc-TBC1D2B-4	284	SOD2-OT1	3377	lnc-FAM174A-6	3408
lnc-UBE2H-2	279	lnc-CNGB1-1	3369	LINC01606	3390
lnc-CDK6-1	279	lnc-ESCO1-3	3352	lnc-GPR101-1	3388
lnc-KCNB2-5	279	lnc-WARS2-5	3348	lnc-PTMA-1	3362
lnc-ST8SIA4-3	272	LINC02454	3345	SPATA3-AS1	3308
KDM7A-DT	264	lnc-ATP6V1E2-2	3344	LINC00472	3302
lnc-ICA1-3	263	LINC01471	3343	lnc-TENM2-1	3286
lnc-ZSCAN10-3	261	lnc-NUCB2-4	3343	lnc-TMTC3-13	3275
lnc-CLASP2-5	257	lnc-HIST1H2AG-4	3342	lnc-DUSP4-6	3247
lnc-SLC6A4-151	256	LINC01615	3341	lnc-ANGPT2-2	3227
lnc-PAMR1-3	254	LINC01060	3334	LINC02226	3206
lnc-LRR1-1	253	lnc-MTA3-8	3332	lnc-ADGRL3-2	3204
lnc-NEIL3-9	250	CPB2-AS1	3326	LINC01410	3172
lnc-OR4F21-4	250	lnc-ADAMTSL4-7	3320	lnc-RAB23-1	3126
lnc-LHX2-8	250	lnc-ARHGAP29-2	3318	lnc-TEK-1	3069
lnc-PRSS23-4	248	lnc-IVNS1ABP-5	3314	lnc-COX7A2L-4	3027
lnc-CRYBA4-52	247	LINC02015	3312	lnc-ARHGAP29-5	2969
lnc-CDK17-10	246	lnc-MPP6-1	3308	lnc-B3GNT10-21	2965
lnc-RPRM-10	245	lnc-NKD1-3	3306	ACAP2-IT1	2872
lnc-RNF113B-8	244	lnc-CNNM1-3	3305	lnc-TMEM132E-1	2823
lnc-KIF18A-4	243	lnc-USP53-6	3304	lnc-GLRA4-2	2812
LINC01661	241	lnc-SLC9C2-5	3301	lnc-TENM4-8	2799
lnc-SCML1-1	236	lnc-SOX30-3	3295	lnc-CPNE4-1	2776
lnc-NXPH1-2	234	lnc-PTPRU-11	3291	lnc-GPC6-7	2724
LINC02422	233	lnc-RHD-2	3291	lnc-CTBS-4	2719
LINC01724	230	lnc-ZNF784-1	3283	lnc-PHF14-14	2718
lnc-CTIF-9	229	lnc-SELENOT-5	3271	lnc-PPDPFL-13	2715
lnc-MYOM1-11	228	lnc-C10orf120-1	3258	lnc-MTA2-4	2713
FTX	226	lnc-MEX3C-6	3258	lnc-GTF2B-15	2661
DAPK1-IT1	226	APTR	3252	lnc-MARCH4-2	2654
lnc-FAM92A-9	225	lnc-ADGRB3-6	3252	lnc-PASK-2	2606

lnc-IRS4-3	224	lnc-DAPK2-3	3247	lnc-CLDN20-3	2524
lnc-CBR1-2	224	LINC02018	3245	lnc-TLR10-7	2516
BAALC-AS1	222	lnc-PRPF18-10	3242	lnc-POLR3A-6	2490
lnc-ZNF391-5	222	lnc-RPLP1-12	3241	lnc-GALNT2-3	2375
lnc-TTLL4-2	221	lnc-SEMA6A-1	3235	lnc-HSD17B7-1	2365
LINC01963	219	lnc-TMEM135-5	3235	lnc-SLC26A8-1	2299
NPSR1-AS1	219	lnc-PPIL1-1	3235	lnc-BRD1-23	2290
lnc-TMEM178A-1	219	lnc-SOD2-6	3232	LINC00691	2230
lnc-ULK4-14	218	lnc-AJAP1-9	3226	lnc-TBC1D2B-7	2176
CDRT8	217	LINC00701	3224	lnc-ANKRD10-4	2070
MIR924HG	214	lnc-DTNBP1-16	3217	lnc-GRIK5-1	1912

Table 8: lncRNA expression profiles for H2B-, PRM1- and PRM2-bound RNAs of Individual 3

H2B		PRM1		PRM2	
GeneID	Hits	GeneID	Hits	GeneID	Hits
lnc-LRR1-1	2154	lnc-LRR1-1	5234	lnc-KCNE1B-158	6177
lnc-NEMF-1	570	lnc-NEMF-1	1930	lnc-LRR1-1	5600
lnc-CCNB1IP1-1	242	lnc-CCNB1IP1-1	561	lnc-NEMF-1	1901
lnc-OSBPL9-6	185	lnc-ARHGEF39-1	463	lnc-SLITRK1-8	491
LINC01372	180	MALAT1	380	lnc-CCNB1IP1-1	480
lnc-PRKN-18	172	LINC00599	281	NEAT1	461
lnc-IL17RA-4	165	lnc-KCNE1B-155	277	lnc-KCNE1B-155	396
lnc-PCSK9-1	158	lnc-WSCD1-3	263	lnc-ARHGEF39-1	391
lnc-FSCB-9	148	lnc-APOL1-14	262	lnc-PGAM2-2	327
lnc-RAI14-3	140	lnc-ZNF584-2	252	lnc-SLX1B-5	323
LINC01978	139	PTOV1-AS2	250	lnc-SNRPN-8	320
STX18-AS1	138	lnc-ANO2-10	244	lnc-DNAJC24-6	310
lnc-TNF-1	138	lnc-ICA1L-7	244	ADD3-AS1	308
lnc-ADRA1D-4	134	lnc-DLK1-35	239	lnc-EPS15-2	307
lnc-STRN3-14	133	lnc-H2AFY-3	231	lnc-GRAMD4-2	302
lnc-SFTPD-5	130	lnc-AUTS2-2	224	lnc-LNPK-2	298
lnc-ZNF585A-4	130	lnc-SELENOF-2	222	LINC01648	296
lnc-CXorf49-10	130	lnc-B3GALT1-1	222	LINC02102	293

LINC01585	129	lnc-ZNF783-1	220	MALAT1	292
LINC01128	128	lnc-AGT-6	216	LINC01562	292
lnc-GGCT-1	128	lnc-GABPA-20	214	LINC02411	290
LINC01322	127	lnc-PRKAA2-3	213	lnc-SLC25A25-4	290
lnc-GLYATL3-2	127	lnc-HTR1B-1	212	lnc-POF1B-3	288
lnc-TENM3-3	125	lnc-GPR107-1	210	lnc-OPALIN-1	283
LINC01105	124	MIR99AHG	209	lnc-SPANXB1-4	282
lnc-BAG5-2	124	lnc-LEF1-4	206	SAMMSON	280
lnc-MAPK6-17	123	lnc-ELAVL2-1	204	lnc-EVX1-14	280
lnc-ZNF517-3	123	lnc-HMGA1-5	204	lnc-ZNF131-2	278
LINC01422	122	lnc-SMIM15-8	203	lnc-ERICH1-19	277
lnc-CD72-1	118	TMPO-AS1	202	lnc-IL6ST-1	277
lnc-KCNE1B-155	118	lnc-RBP4-2	202	lnc-SEPT9-3	274
lnc-SNRPC-3	117	LINC01684	200	lnc-DBT-7	272
LINC01010	116	LINC02374	200	lnc-RASGRF1-2	271
lnc-MYOM2-4	115	RBM15-AS1	200	lnc-LCN9-4	270
LINC02147	115	lnc-TGM6-5	199	lnc-CHODL-4	268
lnc-CDH9-15	115	lnc-CNBD1-4	197	lnc-CD8B2-17	267
LINC02015	114	lnc-SIX3-3	196	lnc-PPIL2-4	264
lnc-RBBP6-4	112	KCNQ1OT1	196	lnc-MED15-1	263
lnc-EPHA7-4	111	LINC00391	195	lnc-KIF3A-66	263
lnc-C15orf41-18	111	LINC02027	194	lnc-ZBTB2-1	262
lnc-NTSR2-5	110	lnc-EML5-2	194	VLDLR-AS1	260
lnc-RRAS2-9	109	lnc-PLPP3-12	192	lnc-SLITRK5-1	260
DDX11-AS1	108	lnc-MFSD14B-3	191	lnc-MGARP-3	260
lnc-ZFHX4-9	107	lnc-GPX5-1	190	lnc-STC2-1	259
LINC01588	106	lnc-DHX37-15	190	lnc-KATNAL2-1	250
lnc-NFAM1-3	105	TBX18-AS1	188	lnc-ERRFI1-2	250
lnc-HAND1-5	102	LINC01060	185	lnc-YES1-8	247
lnc-LRRTM2-4	101	lnc-LMNTD1-1	185	lnc-DPRX-2	247
lnc-SHOC2-3	98	lnc-DPP8-1	185	lnc-CYP7A1-1	245
lnc-CMPK2-32	98	lnc-SUMF1-4	184	lnc-SMIM21-1	245
LINC01814	96	lnc-NDNF-1	184	lnc-HLA-C-2	242

lnc-EQTN-1	96	ATP6V0E2-AS1	182	lnc-PIEZO2-4	241
UBR5-AS1	96	lnc-WNT5B-3	182	lnc-RHOV-3	240
lnc-LRRC72-4	94	LINC01444	181	lnc-FBXO43-6	237
lnc-DTNBP1-16	93	LINC02301	180	lnc-IGSF11-5	236
lnc-TMUB2-1	93	lnc-GABPA-3	180	LINC02229	232
lnc-ACTR3B-6	93	lnc-OST4-9	180	lnc-SOCS6-10	232
MIR583HG	92	KTN1-AS1	178	lnc-SELENOW-4	232
lnc-FGF9-3	91	lnc-ANKRD46-5	177	lnc-KANK1-1	232
lnc-AIDA-2	91	lnc-TMEM156-2	176	lnc-ONECUT2-4	231
lnc-AIPL1-5	91	lnc-TRAF5-12	176	lnc-ESRP1-2	230
lnc-PCDH7-5	90	lnc-FBXO9-3	175	lnc-NUDC-1	230
lnc-MYBBP1A-1	90	lnc-NVL-1	171	lnc-TBC1D2B-4	229
lnc-SMYD3-5	90	lnc-TMEM132B-4	170	lnc-NOP9-1	226
lnc-ACTR2-3	88	AKT3-IT1	170	lnc-HAO2-4	224
LINC01441	87	LINC00484	167	lnc-CDH18-17	224
lnc-ZFYVE1-1	86	lnc-KCTD3-9	167	lnc-AHCY-2	224
lnc-GMIP-1	86	LINC01956	166	lnc-NIFK-6	222
lnc-OTOL1-12	86	lnc-CLDN10-9	166	lnc-SYT10-2	220
lnc-GLDC-6	85	lnc-ZEB2-22	166	lnc-EN2-3	219
lnc-STPG1-3	84	lnc-SP9-7	165	lnc-LMNTD1-3	217
lnc-MLXIP-12	84	lnc-SUMO1-8	165	LINC01126	217
lnc-CDKL2-2	84	CELF2-AS1	164	SGMS1-AS1	214
lnc-RPL11-6	84	lnc-PLEKHM1-3	163	lnc-ZNF726-4	213
lnc-PGPEP1-6	84	lnc-IRX3-80	162	lnc-STPG2-3	210
LINC01299	83	lnc-MSTN-2	162	lnc-DERA-7	208
RUNDC3A-AS1	83	lnc-PLCL1-3	162	lnc-ZNF705D-2	206
lnc-ANKRD20A4-7	83	lnc-CEP170-9	160	lnc-KLF6-18	206
lnc-ARMC6-1	82	lnc-LYPD6-2	159	lnc-DPF3-2	205
lnc-TSTA3-1	81	lnc-HNRNPA0-4	159	lnc-KCNE1B-3	204
lnc-EBF2-3	80	CLRN1-AS1	156	lnc-SENPA6-3	203
lnc-GRIP2-1	79	lnc-EPB41L4B-1	156	lnc-RUNX1T1-3	203
lnc-FAM49A-3	79	lnc-ADM-2	156	LINC00383	200
lnc-ATP6V1B2-3	79	lnc-DAPK3-2	153	lnc-PTPRU-11	199

lnc-AGT-6	78	LINC00709	152	lnc-POU4F2-4	199
lnc-PKD2-1	78	lnc-EFCAB12-2	151	lnc-KAT6A-3	199
lnc-ZNF385D-3	77	lnc-EVX1-14	148	lnc-ZNF345-29	198
lnc-VAX1-4	76	lnc-PLIN2-1	147	KLHL7-DT	198
lnc-FOXC2-1	76	LINC01718	146	lnc-ALDH1A2-4	197
lnc-MTF2-4	76	lnc-VCAN-1	146	lnc-FOXC1-10	194
lnc-RPL30-3	75	lnc-ZNF169-2	146	lnc-NKIRAS1-6	194
lnc-ERV3-1-10	74	lnc-WISP1-11	143	lnc-PABPC4L-21	193
lnc-THAP6-1	74	lnc-COL6A3-11	143	lnc-GGCT-1	192
LINCMD1	74	lnc-DCTD-65	142	lnc-AKAP7-1	192
lnc-UROD-2	74	lnc-SPRED2-21	141	lnc-KLF6-8	191
lnc-PPDPFL-5	74	lnc-FBXO11-1	140	lnc-FANCD2-1	189
RBM15-AS1	73	lnc-CCNB1IP1-4	139	lnc-LMBR1-11	188
lnc-PSMG4-11	73	LINC00589	137	lnc-ARPC5L-1	186
lnc-RPRML-3	72	lnc-BEND6-1	137	THUMPD3-AS1	182
lnc-EIF2AK3-4	72	LINC01687	135	lnc-EPN2-3	182

Table 9: lncRNA expression profiles for H2B-, PRM1- and PRM2-bound RNAs of Individual 4

H2B		PRM1		PRM2	
GeneID	Hits	GeneID	Hits	GeneID	Hits
lnc-LRR1-1	4495	lnc-KCNE1B-158	9846	lnc-LRR1-1	7874
lnc-NEMF-1	1285	lnc-LRR1-1	8391	lnc-NEMF-1	3864
lnc-KCNE1B-158	726	lnc-NEMF-1	3284	lnc-KCNE1B-158	1135
lnc-ARHGEF39-1	431	lnc-KCNE1B-155	1207	lnc-SNX2-1	977
lnc-CCNB1IP1-1	411	VIM-AS1	669	lnc-CCNB1IP1-1	915
lnc-KCNE1B-155	311	lnc-CCNB1IP1-1	633	lnc-SLC4A10-7	913
lnc-FZD4-1	292	GATA2-AS1	574	LINC01524	909
lnc-CGAS-3	224	lnc-MRPL14-4	561	lnc-ING1-1	850
MIR646HG	218	lnc-RECQL4-3	540	lnc-PCDH10-5	833
MALAT1	218	KCNQ1-AS1	537	lnc-GPBP1-3	816
LINC00479	206	FAM138B	520	lnc-HOMER1-6	813
lnc-SLIT2-1	199	lnc-ARHGEF39-1	506	lnc-ITGA7-1	804
LINC-PINT	192	lnc-POTEH-8	497	LINC00923	800

LINC00662	179	lnc-NUP214-3	493	lnc-ADGRL1-1	799
lnc-DDIAS-1	174	lnc-GRIP2-4	490	lnc-IL6-11	798
NR2F2-AS1	168	lnc-GPD2-2	488	lnc-DOC2A-1	794
PPP1R26-AS1	168	lnc-MPLKIP-7	487	lnc-UBE2L5-2	784
lnc-DDX18-2	165	lnc-GLCCI1-3	487	FAM53B-AS1	773
lnc-RIPOR1-1	164	lnc-TMEM52-2	482	LINC01550	765
lnc-KANSL1-7	164	lnc-TNRC6A-2	479	lnc-TUBB2B-6	763
lnc-C3orf58-8	161	NIFK-AS1	478	LINC01588	744
lnc-MAPK6-1	157	lnc-THOC3-2	463	lnc-HIST4H4-1	744
lnc-NXPH1-2	154	lnc-PDE6A-1	462	lnc-SHISA9-3	735
lnc-ADCY9-1	152	lnc-LYPLA2-1	459	lnc-ANKRD11-1	731
lnc-BCL2L11-1	148	lnc-TLE1-9	452	LINC02145	717
lnc-SRSF6-5	148	FAM66B	450	lnc-SEPT14-6	717
lnc-SLC29A3-3	147	lnc-EMC8-8	443	lnc-TP53TG3F-33	712
lnc-RIPK1-1	146	lnc-ERAL1-5	436	lnc-BTNL8-6	711
lnc-ZKSCAN8-3	146	LINC01394	434	TMEM202-AS1	702
ID2-AS1	145	lnc-CHST10-3	434	lnc-KCND2-2	696
ZNRF3-AS1	144	lnc-WNT7B-3	434	lnc-FXYD4-4	692
lnc-TMEM167B-1	143	MAGI1-IT1	430	GAS1RR	682
lnc-POLR1A-6	142	lnc-CCDC125-10	430	MALAT1	681
lnc-NTMT1-2	138	lnc-JADE2-2	424	lnc-MON1A-2	677
lnc-THEMIS-2	137	lnc-PYGO2-2	424	lnc-SNN-6	671
LINC02062	137	LINC00960	422	lnc-GSDMC-34	660
ZMIZ1-AS1	136	lnc-FAM153B-2	422	lnc-FAM84B-84	656
lnc-DMKN-1	134	FOXP4-AS1	418	lnc-N4BP2L2-4	656
lnc-SNPH-6	134	lnc-ZFP42-12	418	lnc-ITIH2-10	652
lnc-PABPC4L-9	133	LINC01376	416	lnc-C11orf95-1	647
lnc-DRD5-4	132	lnc-GATAD1-1	414	lnc-CPEB4-4	636
lnc-MON2-2	132	lnc-CDKN3-1	410	lnc-GLMP-1	636
lnc-RPS24-2	132	lnc-SMIM14-2	408	lnc-RGS5-1	633
lnc-FAM105A-1	132	lnc-PIP4K2A-1	408	ITPKB-IT1	633
lnc-KIF2B-5	132	LINC01776	407	LINC01914	596
lnc-IRS1-7	132	lnc-MRPS6-3	406	lnc-TLR4-1	595

lnc-WNT5B-7	130	lnc-CEACAM8-3	405	lnc-KCNE1B-155	584
lnc-SCML4-2	130	lnc-OTOL1-7	402	lnc-ANXA2R-1	567
lnc-GPX7-1	130	lnc-NPBWR1-7	398	lnc-PTPRU-11	566
lnc-STARD9-3	130	LMCD1-AS1	394	MAFTRR	561
LINC01776	128	lnc-IAH1-3	393	lnc-RBL2-2	548
lnc-DRD5-10	128	lnc-TMLHE-3	393	lnc-ANXA1-62	544
lnc-PDLIM3-3	128	lnc-PPM1E-4	385	lnc-SGTB-4	528
lnc-SC5D-4	128	LINC01615	384	lnc-SLC38A10-3	519
LINC00324	128	lnc-LRP5L-9	383	FIRRE	516
RFPL1S	127	lnc-KCNE1B-156	382	lnc-ARHGEF39-1	514
lnc-EFEMP1-3	126	lnc-PRKACG-1	379	lnc-UTRN-1	509
lnc-KDM8-1	126	lnc-FOXQ1-16	375	lnc-ZNF221-2	506
lnc-TTC30B-1	125	VAV3-AS1	373	TRDN-AS1	503
lnc-GUSB-13	125	lnc-TDRD3-11	366	lnc-CELF6-3	500
lnc-CRYBA4-52	123	lnc-LARP6-1	364	PAX8-AS1	480
LINC01033	122	lnc-RARB-1	363	lnc-TRMT61B-1	476
lnc-KDM2B-1	122	NRAV	361	LINC00894	473
lnc-LENG9-3	122	lnc-ACER2-7	360	NUTM2B-AS1	471
lnc-OSR2-4	122	lnc-SIRT4-5	358	MIR663AHG	467
lnc-MEX3B-1	120	MALAT1	356	lnc-SLC2A3-1	465
lnc-MYO16-9	120	LINC01320	353	LINC01568	463
lnc-GPR15-3	119	lnc-PAG1-5	350	lnc-ZNF37A-16	459
lnc-IGFBP3-5	118	lnc-DTNBP1-16	345	lnc-EPHA5-5	448
lnc-IGFBP2-8	117	lnc-IQCF6-2	345	UGDH-AS1	441
lnc-SLC25A17-2	117	lnc-C17orf51-1	344	lnc-MB-6	439
lnc-LSS-1	116	lnc-FAM174A-6	340	lnc-DMKN-1	433
LRRC75A-AS1	116	lnc-BOD1-1	338	CATIP-AS1	432
lnc-C2orf40-1	115	lnc-NUTM1-11	334	lnc-CDC40-5	429
lnc-POTEI-2	113	lnc-RNF135-1	333	lnc-COPG1-3	422
lnc-MKRN3-2	112	lnc-DRD5-28	332	lnc-TSFM-3	421
HNF4A-AS1	111	lnc-EPN2-3	324	lnc-SF3B2-1	418
lnc-OR10C1-1	111	lnc-AKR1E2-9	321	LINC00578	416
lnc-CCDC102B-7	109	lnc-RASA1-29	318	lnc-FBXW7-1	408

lnc-ADCY8-2	108	lnc-AKAP1-1	315	lnc-TMEM109-2	401
lnc-LYZL2-5	108	lnc-FAM160A1-11	314	lnc-TEX264-2	398
ADIRF-AS1	107	lnc-MSLN-2	309	lnc-ACTR10-7	396
lnc-PPP2R2C-1	106	lnc-PTMS-2	308	lnc-BDH2-3	389
lnc-SIRT4-5	106	lnc-C15orf41-18	307	lnc-WDR72-2	387
lnc-LINS1-5	106	lnc-KCNK13-3	307	CARMN	383
lnc-ZNF713-2	106	lnc-LRP5L-1	304	lnc-DTNBP1-16	374
lnc-COL6A3-11	104	lnc-TMEM259-5	304	lnc-SMIM14-5	370
MIR4300HG	103	lnc-APCDD1L-5	303	lnc-CYTL1-5	368
lnc-SCRIB-1	103	lnc-POU3F3-5	297	lnc-AES-8	361
lnc-LDAH-4	103	lnc-PSMA2-1	295	lnc-ACBD3-3	360
FBXO3-DT	102	lnc-CEP170-9	289	TEX41	359
lnc-TMPRSS2-83	102	lnc-ACOX3-5	285	lnc-FKTN-1	359
lnc-SAG-4	102	lnc-TMCO1-2	284	lnc-CDH18-17	356
lnc-ALB-7	102	lnc-AP4E1-4	284	lnc-RBM28-3	354
lnc-CCDC167-4	102	lnc-THAP4-3	281	lnc-SCGB2B2-131	352
lnc-CAV2-2	102	lnc-BBS2-2	281	lnc-MARVELD3-5	347
lnc-C12orf75-1	100	lnc-SLC2A12-5	280	lnc-CASP10-2	346
lnc-TNRC6A-3	100	lnc-DNAAF5-1	276	lnc-CCDC68-1	344
lnc-RIOX2-11	100	lnc-HAND2-3	272	lnc-LHFPL4-6	341
lnc-RNF24-2	98	lnc-ISX-2	272	lnc-MAST4-5	337

Table 10: Amount of overlapping lncRNA genes for the 50 most abundant lncRNA genes between individuals for H2B, PRM1 and PRM2

	H2B	PRM1	PRM2
Individual 1	48	47	46
Individual 2	48	50	49
Individual 3	45	45	44
Individual 4	45	45	44
Individual 1 & 2	0	0	0
Individual 1 & 3	0	0	0
Individual 1 & 4	0	0	0
Individual 2 & 3	1	0	0

Individual 2 & 4	1	0	0
Individual 3 & 4	2	2	2
Individual 1 & 2 & 3	0	0	0
Individual 1 & 2 & 4	0	0	0
Individual 1 & 3 & 4	2	3	3
Individual 2 & 3 & 4	0	0	0
all Individuals	0	0	1

Table 11: RIP-Seq hits with lncRNA for each protein and individual

		reads	lncRNA hits	lncRNA % hits alignable reads
Individual 1	H2B	22163408	2079062	17,69258
	PRM1	15730969	93536	2,9089913
	PRM2	25421321	527121	6,6953148
Individual 2	H2B	12636284	39757	2,8971061
	PRM1	44190756	803552	3,3574063
	PRM2	23285782	294986	5,3071108
Individual 3	H2B	26226818	12555	2,4423902
	PRM1	31692799	27740	2,7352428
	PRM2	33433524	44947	3,2551312
Individual 4	H2B	30210680	22947	2,9904166
	PRM1	32685372	58361	2,5326792
	PRM2	36297056	52841	2,6909298

Table 12: RIP-Seq hits with tRNA for each protein and individual

		reads	tRNA hits	tRNA % hits alignable reads
Individual 1	H2B	22163408	2557	0,0217598
	PRM1	15730969	291	0,0090502
	PRM2	25421321	1673	0,0212499
Individual 2	H2B	12636284	0	0
	PRM1	44190756	4038	0,0168716
	PRM2	23285782	0	0
Individual 3	H2B	26226818	0	0

	PRM1	31692799	146	0,014396
	PRM2	33433524	0	0
Individual 4	H2B	30210680	33	0,0043005
	PRM1	32685372	127	0,0055114
	PRM2	36297056	502	0,0255644

Table 13: RIP-Seq hits with sno/miRNA for each protein and individual

		reads	miRNA hits	miRNA % hits alignable reads
Individual 1	H2B	22163408	13843	0,1178024
	PRM1	15730969	2093	0,0650928
	PRM2	25421321	10893	0,1383592
Individual 2	H2B	12636284	553	0,0402973
	PRM1	44190756	12831	0,0536106
	PRM2	23285782	6174	0,1110768
Individual 3	H2B	26226818	177	0,0344327
	PRM1	31692799	361	0,0355956
	PRM2	33433524	860	0,0622825
Individual 4	H2B	30210680	824	0,1073824
	PRM1	32685372	1741	0,0755538
	PRM2	36297056	457	0,0232727