DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
368

VIKTORIJA KUKUŠKINA

Understanding the mechanisms of endometrial receptivity through integration of ‘omics’ data layers

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
368

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
368

VIKTORIJA KUKUŠKINA

Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers

UNIVERSITY OF TARTU
Press

Institute of Molecular and Cell Biology, University of Tartu, Estonia

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular Biomedicine on May 14, 2020 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisors: Reedik Mägi, PhD
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Prof. Andres Metspalu, MD, PhD
Chair of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia
Estonian Genome center, Institute of Genomics, University of Tartu, Tartu, Estonia

Prof. Andres Salumets, PhD
Department of Obstetrics and Gynecology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Reviewer: Prof. Maido Remm, PhD
Chair of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Opponent: Prof. Stephan Beck, PhD
Chair of Medical Genetics, University College London, UCL Cancer Institute, London, UK

Commencement: 29.06.2020 at 11.00 am, at Room No 105, 23B Riia St., Tartu and via University of Tartu MS Teams

The publication of this dissertation is granted by the Institute of Molecular and Cell Biology at the University of Tartu.

This research was funded by grants IUT34-16 and IUT34-4 from the Estonian Ministry of Education and Research, by Enterprise Estonia, grant no EU48695, by the EU FP7-PEOPLE-2012-IAPP grant SARM (grant no 324509), by EU-FP7 Eurostars Program (grant NOTED, EU41564) and by the European Commission Horizon 2020 research and innovation programme under grant agreement 692065 (project WIDENLIFE).University of Tartu ASTRA project PER ASPERA, NIASC, Estonian Center of Genomics/ Roadmap II (project No. 2014-2020.4.01.16-0125). Data analyses were carried out in part in High Performance Computing Center of University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-353-9 (print)
ISBN 978-9949-03-354-6 (pdf)

University of Tartu Press
www.tyk.ee

"You could find out most things, if you knew the right questions to ask.
Even if you didn't, you could still find out a lot."

Iain M. Banks, The Player of Games

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS 9

ABBREVIATIONS 10

INTRODUCTION 12

1. LITERATURE REVIEW 14
   1.1. Journey from past to nowadays 14
   1.2. Overview of female reproduction 16
   1.3. Menstrual cycle 18
      1.3.1. Regulation of menstrual cycle 18
   1.4. Receptive endometrium and embryo implantation 21
   1.5. Transcriptomics methods 23
      1.5.1. Expression microarrays vs RNA sequencing 25
      1.5.2. RNA-seq data processing and analysis 25
      1.5.3. Factors affecting RNA-seq analysis 28
   1.6. Role of gene expression in endometrial receptivity context 32
      1.6.1. Endometrial receptivity microarray studies 32
      1.6.2. The era of ERA (Endometrial Receptivity Array) 33
      1.6.3. Endometrial receptivity RNA-seq studies 34
      1.6.4. Non-coding RNAs 35
   1.7. Methylomics methods 38
      1.7.1. Evolution of methods for studying DNA methylation 40
      1.7.2. Methylation sequencing and microarray data analysis workflows 42
      1.7.3. Factors affecting DNA methylation analysis 46
   1.8. Role of DNA methylation in endometrial receptivity context 47
   1.9. Integration of 'omics' layers 49

2. AIMS OF THE STUDY 51

3. RESULTS AND DISCUSSION 52
   3.1. Effect of cellular composition on endometrial gene expression 52
      3.1.1. Description of cohort, materials, and methods 52
      3.1.2. Differentially expressed genes with and without cell proportion adjustment 55
      3.1.3. Validation of cell proportions 58
   3.2. Correlation between potential endometrial receptivity marker genes and their associated miRNAs 60
      3.2.1. Description of cohort, materials, and methods 60
      3.2.2. Potential receptivity marker genes and their associated regulating miRNAs 61
   3.3. Methylation pattern differences between early-secretory and mid-secretory endometrium and its correlation with gene expression 65

3.3.1. Description of cohort, materials, and methods 65
3.3.2. Differential methylation in mid-secretory vs early-secretory endometrium 67
3.3.3. Correlation between methylation and gene expression levels, enrichment and pathway analyses 70
CONCLUSIONS 73
SUMMARY IN ESTONIAN 74
REFERENCES 77
ACKNOWLEDGMENTS 93
PUBLICATIONS 95
CURRICULUM VITAE 145
ELULOOKIRJELDUS 148

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, referred to in the text by Roman numerals (Ref. I to Ref. III).

I Suhorutshenko, M.[1], **Kukushkina, V.***, Velthut-Meikas, A., Altmäe, S., Peters, M., Mägi, R., Krjutškov, K., Koel, M., Codoñer, F.M., Martinez-Blanch, J.F., Viella, F., Simon, C., Salumets, A., Laisk, T. (2018). Endometrial receptivity revisited: endometrial transcriptome adjusted for tissue cellular heterogeneity. Hum Reprod. *33* (11), 2074–2086.

II Altmäe, S., Koel, M., Võsa, U., Adler, P., Suhorutšenko, M., Laisk-Podar, T., **Kukushkina, V.**, Saare, M., Velthut-Meikas, A., Krjutškov, K., Aghajanova, L., Lalitkumar, P.G., Gemzell-Danielsson, K., Giudice, L., Simon, C., Salumets, A. (2017). Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers. Sci Rep. *7,* 10077.

III **Kukushkina, V.***, Modhukur, V.*, Suhorutšenko, M., Peters, M., Mägi, R., Rahmioglu, N., Velthut-Meikas, A., Altmäe, S., Esteban, F.J., Vilo, J., Zondervan, K., Salumets, A., Laisk-Podar, T. (2017). DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Sci Rep. *7*, 3916.

The publications listed above have been reprinted with the permission of the copyright owners.

My contributions to the listed publications were as follows:

Ref. I Performed all necessary processing steps before analyses of biopsy RNA data, performed deconvolution, differential expression analyses and meta-analyses. Partially prepared figures and participated in writing the manuscript.

Ref. II Performed all necessary data preprocessing steps and differential expression analyses. Performed correlation analysis between expressions of mRNA and miRNA data, respectively. Revised the manuscript.

Ref. III Performed methylation and expression data preprocessing, and differential analyses. Performed correlation analysis between gene expression and methylation values, participated in preparation of the figures and in writing the manuscript.

[1] * – indicates joint First Authorship

ABBREVIATIONS

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AD	*Anno Domini* (the year Jesus was born)
BMI	Body mass index
BMIQ	Beta-Mixture Quantile (normalization)
CAGE	Cap analysis of gene expression
CCD	Charge coupled device
cDNA	Complementary DNA
CGI	CpG island
CNV	Copy number variation
COX-2	Cyclooxygenase 2
CpG	Cytosine-guanine dinucleotide site
CPM	Counts per million
DEG	Differentially expressed gene
DMR	Differentially methylated region
DNMT	DNA methyltransferase
ds-cDNA	Double-stranded complimentary DNA
EGF	Epidermal growth factor
ERA	Endometrial receptivity array
ES	Early-secretory
ESE	Early-secretory endometrium
EST	Expressed sequence tag
EV	Extracellular vesicle
FACS	Fluorescence-activated cell sorting
FC	Fold change
FDR	False discovery rate
FPKM	Fragments per kilobase of transcript per million fragments mapped
FSH	Follicle-stimulating hormone
GnRH	Gonadotropin-releasing hormone
GWAS	Genome-wide association study
HB-EGF	Heparin-binding EGF-like growth factor
IGF	Insulin-like growth factor
IL	Interleukin
ISVA	Independent surrogate variable analysis
IVF	*In vitro* fertilization
LCM	Laser capture microdissection
LH	Luteinizing hormone
LIF	Leukemia inhibitory factor
LIFR	Leukemia inhibitory factor receptor
lincRNA	Long intergenic noncoding RNA
lncRNA	Long non-coding RNA

MACS	Magnetic-activated cell sorting
miRNA	MicroRNA
mRNA	Messenger RNA
MS	Mid-secretory
MS-SNuPE	Methylation-sensitive single nucleotide primer extension
MSE	Mid-secretory endometrium
NGS	Next generation sequencing
NOOB	Normal-exponential out-of-band (background correction)
PBC	Peak based correction
PCA	Principal component analysis
PCOS	Polycystic ovary syndrome
PCR	Polymerase chain reaction
piRNA	PIWI-interacting RNA
PMD	Partially methylated domain
QC	Quality control
qPCR	Quantitative PCR
RIF	Repeated implantation failure
RNA-seq	RNA sequencing
RP-HPLC	Reverse-phase high-performance liquid chromatography
RPL	Recurrent pregnancy loss
RRA	Rank aggregation method
rRNA	Ribosomal RNA
SAGE	Serial analysis of gene expression
siRNA	Short interfering RNA
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SQN	Subset quantile normalization
SSPA	Single-strand conformation polymorphism analysis
STRT-seq	Single-cell tagged reverse transcription sequencing
SVA	Surrogate variable analysis
SWAN	Subset-quantile within array normalization
TGF	Transforming growth factor
T-UCR	Transcribed ultraconserved regions
TPM	Transcripts per million
TSS	Transcription start site
UTR	Untranslated region
WOI	Window of implantation

INTRODUCTION

Human reproduction is a sophisticated matter and can be affected by many factors, such as health status, accompanying diseases, genetic background, environment, and lifestyle. According to the World Health Organization, over 10% of females in a stable relationship are suffering from involuntary infertility/subfertility worldwide. The central process of female fertility – the menstrual cycle – is a complex process that is regulated by various factors, results in the maturation and release of oocytes, and structural changes in the inner uterine lining (endometrial tissue) in anticipation of the implanting embryo. The embryo implantation process requires precise regulation, right timing, and cooperation between the endometrium and the embryo, otherwise, implantation fails. The endometrium is most receptive to the embryo during a relatively short period of time, called the window of implantation. To reach this receptive state, the orchestrated coordination of many factors, including different 'omics' data layers is needed. The main 'omics' disciplines include transcriptomics and methylomics. Perhaps the most well-known is transcriptomics, which is associated with transcription and can be studied using gene expression array or RNA sequencing technologies. On the other hand, methylomics deals with DNA methylation processes and related mechanisms, and also uses arrays and sequencing as research tools. Usually, knowledge of one 'omics' layer is not enough to see the whole picture, as all the processes are interconnected, where DNA methylation regulates gene expression and vice versa, or transcription of one RNA is regulated by another RNA and so on. Thus, the integration of several 'omics' data layers is necessary for a better understanding of complex processes, such as endometrial receptivity.

The current thesis is focused on integrating some of the 'omics' layers in the context of healthy endometrium during the transition from pre-receptive to receptive states using bioinformatic tools. In the literature overview part, I touch upon the mechanisms underlying the female menstrual cycle and embryo implantation. Next, I take a look at research methods used in obtaining and handling transcription and methylation datasets, tell about the pros and cons of these methods, and discuss the possible problems that could arise during data handling. Also, I give an overview of how 'omics' data has been used in the context of female reproductive studies.

In the experimental part, I show the integration between whole-tissue and cell-type-specific RNA sequencing datasets, introduce whole tissue deconvolution and present the results we obtained. Secondly, I evaluate predicted gene-microRNA (miRNA) interactions through the integration of expression datasets of miRNAs and their associated target genes. And finally, I present an analysis of methylation data from endometrial tissue, obtained with Infinium Human Methylation 450K BeadChip, and its correlation with gene expression. The dataset used in these studies is unique, as endometrial whole-tissue methylation, messenger RNA (mRNA) and miRNA expression data originate from the same

individuals/samples, and additionally represent paired sampling i.e. both pre-receptive and receptive samples were collected from the same individual within one menstrual cycle.

The current thesis focuses on only pairwise integration of data layers; however, additional layers would provide an even better understanding of processes underlying endometrial receptivity. Hopefully, our studies encourage other researchers in the field to use more data layers in further endometrial studies.

1. LITERATURE REVIEW

1.1. Journey from past to nowadays

Our understanding of the female reproductive system and related processes has evolved in time. The oldest known description of the ovary was made by ancient Greeks in the first century AD. Ancient Greek physician and philosopher Galen described female uterus as a male penis and testicles that are placed inside the body. While in middle ages master Nicolaus described it as an organ that is hollow and villous inside and smooth outside, is divided into seven compartments and has two openings. The knowledge about female anatomy was broadened by Leonardo da Vinci with his precise drawings and descriptions (Figure 1). Sadly, da Vinci also described the female reproductive system through the male reproductive system. In the 17th and 18th centuries, the female and male reproductive systems were finally separated and viewed as distinct ones.

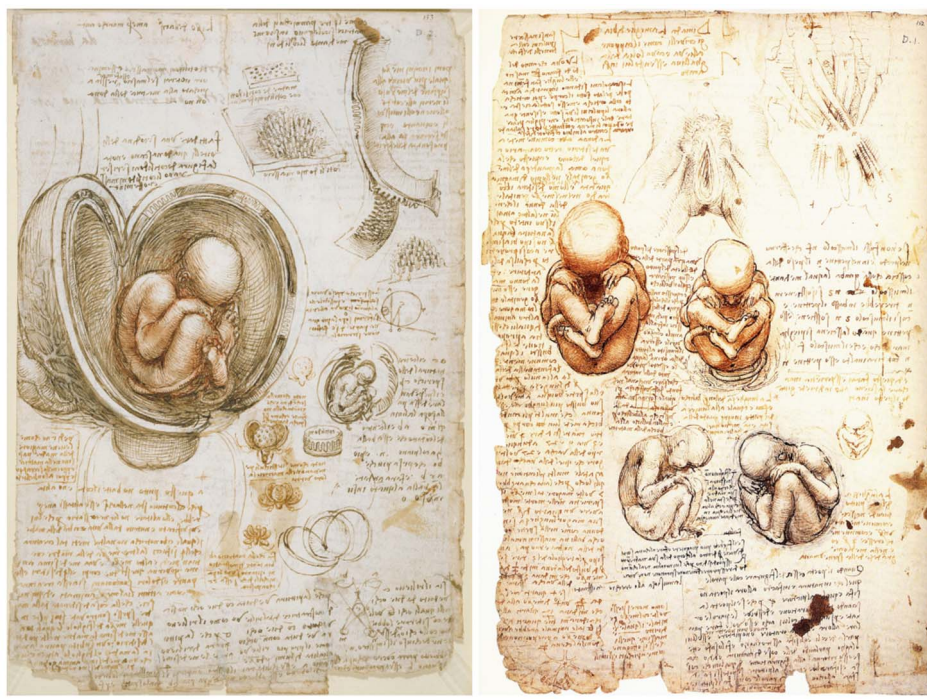

Figure 1: Leonardo da Vinci – Studies of the Foetus in the Womb (Da Vinci, 1510). Copyright: The Royal Collection © Her Majesty Queen Elizabeth II. Source: Wikimedia commons.

In parallel to an increased understanding of female reproduction and fertility, new scientific methods evolved. It could be said that the journey to the micro-world started in the late 16th century with the invention of the first microscopes, when the microscopic breakthrough was made by Antonie van Leeuwenhoek in the 1670s. That resulted in the discovery of the mammalian egg cell by C.E. von Baer in 1826/27 (Von Baer, 1827), description of the human egg cell by Edgar Allen in 1928/29 (Pratt et al., 1929), and observation of the fertilization process (fusion of egg and sperm) by Oskar Hertwig in 1876 (Hertwig, 1876). The start of the genetics era in the second half of the 19th century, with the trait inheritance studies on garden peas by Gregor Mendel (Mendel, 1865) and the first discovery of nucleic acids (Miescher, 1871) moved the female reproduction studies to a molecular level, making the understanding of things even more complex. In the 20th century, genetics and associated methodology evolved rapidly, resulting in the discovery of the DNA structure in 1953 (Watson and Crick, 1953) and invention of the first sequencing method in 1970, which was further adapted and improved by Frederick Sanger (Wu, 1994). The new methodologies contributed to rapid progress in the field of reproductive studies, resulting in first attempts in *in vitro* fertilization (IVF) in the late 60s, and in the birth of the first IVF child in 1978 (Bates and Bowling, 2013). Since then the understanding of various aspects of human reproduction has improved and IVF is currently a relatively common practice. For example, according to the Estonian National Institute for Health Development, in Estonia, nearly 3% of newborns were conceived via IVF (Tervise Arengu Instituut 2015a, 2016b).

The RNA distinguishing from DNA (Allen, 1941) and RNA sequencing (Min Jou et al., 1972), invention of the polymerase chain reaction (PCR) and PCR based methods set the stage for a whole-genome sequencing boom and bloom of ‘omics’ studies. Nowadays, a full range of ‘omics’ studies, such as genomics, transcriptomics, proteomics, metabolomics and more, lead to new research directions where these layers are integrated to provide a bigger and more coherent picture of biological processes (and process malfunctions in diseases) taking place within a single organism, tissue or even a single cell. Such deep insight into biological processes, where all these different ‘omics’ levels are working in orchestrated harmony, gives a possibility to manage reproduction and fertility problems on a personal level, improving diagnostics and treatment outcomes and thus leading to a healthier and happier society.[2]

[2] Historical overview is mainly based on books “A History of the Life Sciences” by Magner L.(Magner, 2002) and “Making sex: body and gender from the Greeks to Freud” by Laqueur T. (Laqueur, 2003); and web-materials “A history of the male and female genitalia” by Findlen P. (Findlen, 1998)

1.2. Overview of female reproduction

The female reproductive system consists of external and internal organs. The outer part is called vulva and forms female external genitalia, while the inner part is divided into the vagina, uterus, fallopian tubes and ovaries. The vagina is a fibromuscular tubular structure connecting the vulva and the uterus. The uterus is a muscular organ that nourishes and supports the growing embryo, and anatomically includes a body and cervix, where the cervix is a narrow part of the uterus that forms a connection with the vagina (Figure 2). The uterine body has three major layers: outer serous membrane or perimetrium, the middle muscular layer (myometrium), and inner uterine lining – endometrium, which plays an important role during embryo implantation, and consists of basal and functional layers. The fallopian tubes connect the uterus with the ovaries and facilitate gamete and embryo transport. The ovaries hold the female ovarian reserve, are the site of folliculogenesis and also a source of sex hormones governing the maturation of the endometrium.[3]

The newborn female already has 1–2 million primary oocytes (Himelstein-Braw et al., 1976), which decrease in number with aging, and by puberty, there are about 300–500 thousand follicles left (Richardson et al., 1987). During the reproductive lifespan, only 400–500 oocytes complete the ovulation process (Richardson et al., 1987), and only one egg is usually released from the ovary each month. At puberty, the hypothalamus starts producing gonadotropin-releasing hormone (GnRH) which signals the pituitary gland to release gonadotropins, luteinizing (LH) and follicle-stimulating hormones (FSH). The increasing levels of gonadotropins stimulate the ovaries to produce steroid hormones – progesterone and estrogen, which are important menstrual cycle regulators (Bates and Bowling, 2013). Menarche, that on average takes place at the age of 13, indicates the start of the female reproductive period and the beginning of a series of menstrual cycles. The cycle series last until menopause, which usually starts at the age of 50, and indicates the end of the reproductive period. The average menstrual cycle is 28 days long and is usually divided according to ovarian/follicle state to follicular and luteal phases, or by the endometrial state to proliferative and secretory phases (Bates and Bowling, 2013; OpenStax, 2016). Ovulation indicates the transition from the follicular to the luteal phase, and also from the proliferative to the secretory phase. After ovulation, if the oocyte is fertilized, it starts rapid division and on the 4th – 5th day after fertilization transforms into a blastocyst. The blastocyst enters the uterine cavity and through communicating with the endometrium using different factors – for example, extracellular vesicles – attaches to the best area suitable for implantation and later invades the endometrium, thus establishing a

[3] Female reproductive system anatomy description is based on "Handbook of Clinical Neurology" book part "Anatomy and physiology of genital organs – women" by Graziottin and Gambini (Graziottin and Gambini, 2015); "Physiology, Female Reproduction" publication by Rosner and Sarao (Rosner and Sarao, 2019); and "Anatomy and Physiology" web-book part "Anatomy and Physiology of the Female Reproductive System" (OpenStax, 2016).

pregnancy. If the oocyte is not fertilized on time, the uterine lining and unfertilized egg are shed via menstrual bleeding.[4]

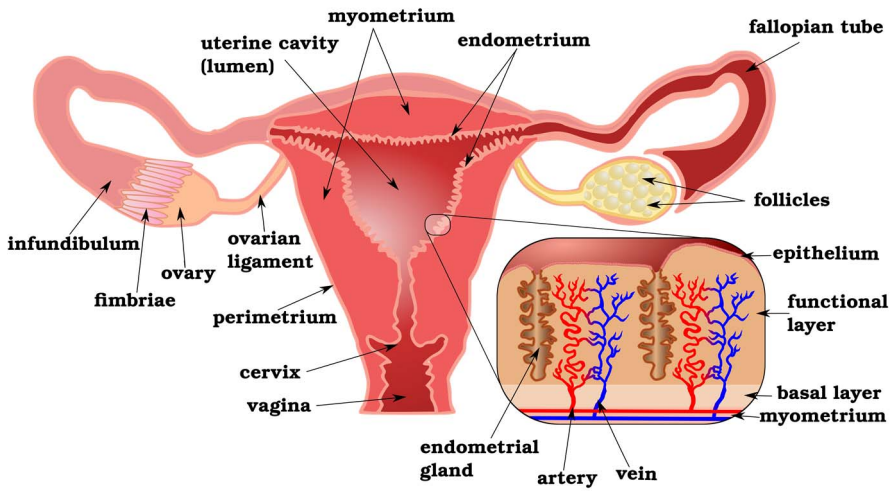

Figure 2. The female reproductive tract. The female reproductive tract includes the vagina, uterus, fallopian tubes and ovaries. Cervix is the junction between the uterus and vagina. Ovaries contain growing follicles, which are released after maturation during ovulation into the abdominal cavity. The fallopian tubes have filaments (*fimbriae*) attached to the ends (*infundibulum*) that catch the released oocytes and guide them into the tubes. The uterus consists of three tissue layers: outer membrane – perimetrium, muscular layer – myometrium and the inner uterine lining, which is partially shed during menstruation – the endometrium. The endometrium is divided into basal (cell source for the functional layer) and functional layer (is shed during menstruation). The functional layer grows and matures during the menstrual cycle, developing glands and a capillary network (maximized: endometrium in late proliferative – early secretory phase). Drawn by Anton Kukuškin.

There are many reasons why human reproduction could fail in the early stages. Some of the failures occur because of problems with sperm, egg or embryo quality; problems with endometrial maturation or implantation timing; embryo-endometrium communication failure; various diseases – for instance, endometriosis (Houshdaran et al., 2016; Sampson, 1927; Sanchez et al., 2014), polycystic ovary syndrome (PCOS), premature ovarian insufficiency, and recurrent pregnancy loss (RPL) (Lucas et al., 2016). When women fail to get pregnant

[4] Description of the menstrual cycle and its related processes are mainly based on review publication by Bates and Bowling (Bates and Bowling, 2013) and "Anatomy and Physiology" web-book part "Anatomy and Physiology of the Female Reproductive System" (OpenStax, 2016).

within one year of regular intercourse, infertility is diagnosed and medical actions will follow. First of all, the cause is determined, and then a proper treatment is chosen depending on the cause of infertility. There are several infertility treatments available, such as intrauterine insemination and IVF. If three or more IVF cycles fail, in which one or two morphologically high-quality embryos are transferred during each cycle, repeated implantation failure (RIF) is diagnosed (Koot et al., 2016; Ruiz-Alonso et al., 2013). Besides, fertility could be influenced by physiological factors, such as genetics, or by lifestyle factors (smoking), or infections by viruses or bacteria (Damario, 2014). Likely, a better understanding of the normal reproductive processes will also help to solve infertility issues.

To better understand the processes underlying what is collectively known as "female fertility", we will now look at different physiological factors that are critical for maintaining female reproductive potential.

1.3. Menstrual cycle

The menstrual cycle is the process lasting from the first day of menstrual bleeding to the first menstruation day of the next cycle and is controlled by hormones produced by the hypothalamus, pituitary gland and ovaries (Figure 3) (Reed and Carr, 2015). The average cycle length in humans varies from 25 to 30 (up to 34) days with a median duration of 28 days. The menstrual cycle length is relatively stable in healthy women at age 20–40 years, and varies greatly within the first 5–7 years after menarche and 6–8 years before menopause (Sherman and Korenman, 1975). Menstrual bleeding usually lasts 3–6 days in 80% of women (Mihm et al., 2011; Reed and Carr, 2015). The luteal phase is usually 14 days long (Rosner and Sarao, 2019), but may vary from 7 to 19 days, while the length of the follicular phase varies more: from 10 to 23 days, and is the main determinant of menstrual cycle length.

1.3.1. Regulation of menstrual cycle

The menstrual cycle starts with the first day of bleeding, which is initiated with progesterone withdrawal (Figure 3. Steroid hormones) (Mihm et al., 2011; Reed and Carr, 2015) associated with the degenerating corpus luteum from the previous cycle. At the same time, the cohort of growing follicles is recruited (Figure 3. Gonadotropins) (Reed and Carr, 2015; Sherman and Korenman, 1975). The follicle theca cells (outer layer) bind LH and as a result start androgen production, while follicle granulosa cells (inner layer) bind FSH and as a result convert androgen to estrogen (Bates and Bowling, 2013), thus decreasing FSH level by negative feedback and raising estrogen levels. As a response to estrogen, endometrium starts growing and proliferating (Figure 3. Endometrial cycle) (Gómez et al., 2015; Nair and Taylor, 2010). LH level starts rising by the mid-follicular phase as a positive response to increased estrogen levels. The

FSH level reaches its minimum the day before ovulation, and LH reaches its maximum level on ovulation day (day ~14 of a 28-day cycle) (Figure 3. Gonadotropins) (Reed and Carr, 2015). Beside FSH and LH, other components play an important role during the follicular phase, such as activins (Sherman and Korenman, 1975), insulin-like growth factor 1 and 2 (IGF1 and IGF2) (Di Pietro et al., 2013; Zhou et al., 1994), various protein kinases (Makieva et al., 2018), and epidermal growth factor, which collectively assist oocyte development and maturation, while the plasma proteins, pituitary hormones, steroids, and non-steroidal ovarian factors regulate microenvironment of the ovary (Mihm et al., 2011). Among others, miRNAs also play a regulatory role in follicular selection and ovulation (Tesfaye et al., 2018).

Ovulation occurs in about 10–12 hours after the LH peak initiated by the rise of estradiol produced by the preovulatory follicle (Pauerstein et al., 1978). As the beginning of surge occurs roughly 34 – 36 hours before ovulation, this is used to predict ovulation timing. The LH surge stimulates progesterone synthesis (Figure 3. Gonadotropins), where progesterone stabilizes endometrial lining and in the following secretory phase induces the transformation of endometrial epithelium and decidualization of stromal fibroblasts, inhibits LH and FSH, and initiates granulosa cell luteinization (Figure 3. Steroid hormones) (Bates and Bowling, 2013; Houshdaran et al., 2014).

After ovulation, luteinized granulosa cells combine with theca-lutein cells forming corpus luteum (Figure 3. Ovarian cycle), that synthesizes progesterone and estrogen from cholesterol and prepares the endometrium for embryo implantation (Gómez et al., 2015; Murphy, 2004; Nair and Taylor, 2010), inducing specific structural and functional changes in endometrial epithelial cells – that include plasma membrane (Murphy, 2004) and cytoskeleton modifications (Martín et al., 2000; Thie et al., 1995) – and also decidualization process in stromal cells (Irwin et al., 1989). In response to corpus luteum secretion, capillaries start the invasion of the granulosa cell layer (Riesewijk, 2003), and on the 8^{th} or 9^{th} day after ovulation, vascularization achieves its peak (Figure 3. Endometrial cycle). Corpus luteum lifespan depends on LH secretion, and if pregnancy does not occur, the tissue undergoes luteolysis and forms scar tissue (Figure 3. Ovarian cycle). The decline of corpus luteum starts on days 9 – 11 after ovulation and results in steroid hormone (including progesterone) concentration decrease. At the same time stroma is filled with immune cells like macrophages, T-cells and natural killer cells (King et al., 1989; Nair and Taylor, 2010). Long exposure to progesterone leads to endometrial tissue thinning and atrophy (Figure 3. Endometrial cycle) (Nair and Taylor, 2010). Due to progesterone withdrawal, the constriction of spiral arterioles takes place and results in decreased blood flow, causing tissue ischemia. Prostaglandin release in endometrium causes contractions of uterine muscles which helps to get rid of degraded tissue.

The description above shows the menstrual cycle from usual and frequently repeating side – when the oocyte is not fertilized, and the cycle ends with menstrual bleeding. But there is another side, which ends with pregnancy and is discussed below.

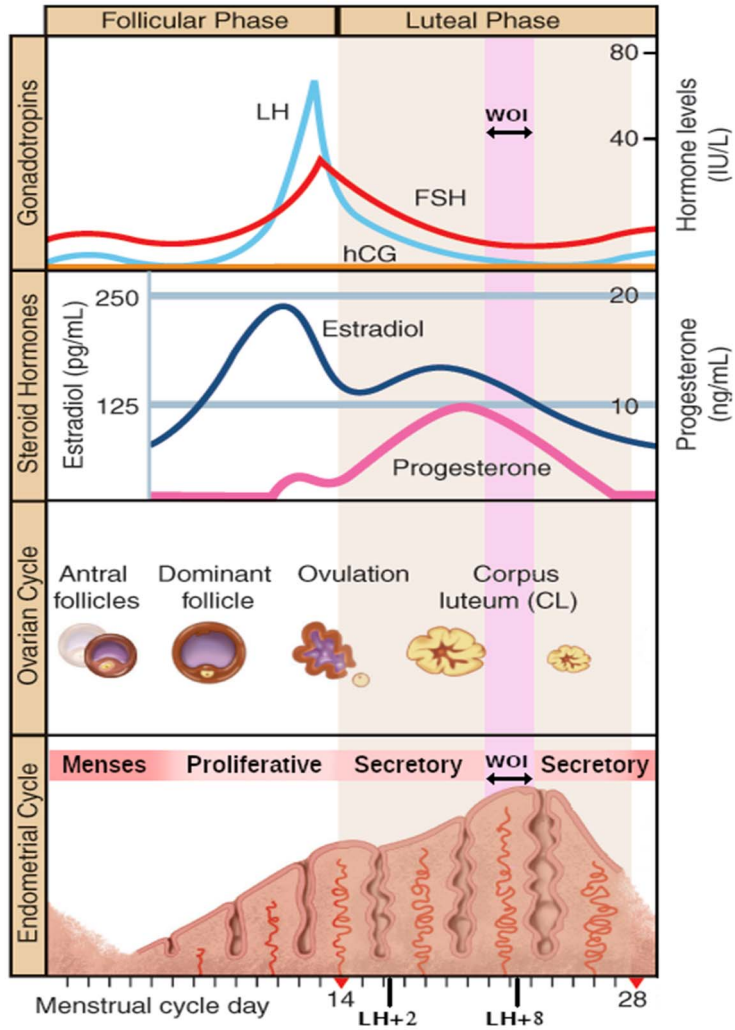

Figure 3. Average menstrual cycle. LH – luteinizing hormone (light blue); FSH – follicle-stimulating hormone (red); hCG – gonadotropin (orange); WOI – window of implantation; LH+2/+8 – indicates days after LH peak, where LH+2 indicates early-secretory (ESE) and LH+8 mid-secretory endometrium (MSE). "Gonadotropins" panel shows fluctuations of LH and FSH during the menstrual cycle, where LH peak occurs prior to ovulation. "Steroid hormones" panel shows changes in estradiol and progesterone. The "ovarian cycle" panel shows follicle maturation, oocyte release, and luteinization. "Endometrial cycle" panel shows endometrial growth, proliferation (Proliferative), vascularisation (Secretory) and shedding (Menses). Adapted from William Obstetrics 25e (Cunningham et al., 2018).

1.4. Receptive endometrium and embryo implantation

A properly functional mature endometrium is necessary for successful embryo implantation. The tissue itself can be divided into two layers – basal and functional layer (Figure 2) (Gómez et al., 2015). The basal layer is a source for the functional layer, consisting mainly of progenitor cells. The functional layer contains glands and stroma, grows from the basal layer and is shed every cycle (Figure 2). Endometrial tissue is composed of diverse cell types, mainly stromal and epithelial (glandular and luminal) cells, but also endothelial cells, leukocytes, other immune cells, and progenitor stem cells (Figueira et al., 2011).

The endometrium is receptive to the embryo during a short period of time called the window of implantation (WOI), which lasts about 48 hours (Gómez et al., 2015). WOI takes place in the mid-secretory phase at days ~7–10 after the LH peak of the 28-day menstrual cycle (Figure 3 Endometrial cycle) (Kao et al., 2002; Riesewijk, 2003). For the endometrium to reach its receptive state, the tissue undergoes vast structural changes each cycle, such as vascularization; a shift in proportions between stromal and epithelial cells, where epithelial cells become more dominant; decidualization of stromal cells; and functional and structural changes of epithelial cells. The embryos that implant during this time, show 84% chance for continuing pregnancy, whereas embryos implanted on day 11 after LH peak show only 18% chance (Kao et al., 2002; Wilcox et al., 1999).

The embryo implantation process takes place in three stages: apposition, adhesion, and invasion (Achache and Revel, 2006; Egashira and Hirota, 2013; Kao et al., 2002). During the apposition stage, the embryo changes its polarity and the zona pellucida (embryonal coating) is shed. Adhesion and invasion steps are quite self-explanatory: during adhesion, the embryo adheres to the endometrium, and during invasion invades into the maternal tissue. Embryo implantation involves a complex sequence of signaling events, such as interactions between cells, and between cells and the extracellular matrix. These events are mediated by lectins, integrins, interleukins (IL), matrix-degrading enzymes, growth factors (GF), cytokines, enzyme inhibitors and others (Figure 4) (Achache and Revel, 2006; Di Pietro et al., 2013; Kao et al., 2002; Paul et al., 2018; Singh et al., 2011). The maternal immune response is also regulated during the implantation step, to prevent severe immunological response to embryo (Gómez et al., 2015; Palomino et al., 2018), but at the same time protect the mother from potential pathogens. The embryo-maternal communication is based on interactions between ligands and their receptors, for example, leukemia inhibitory factor (LIF) and LIF-receptor (LIFR) play a role as communication signals between uterine and embryonic cells, thus being crucial for successful implantation (Cavagna and Mantese, 2003; Paul et al., 2018). The non-coding RNAs also have an important role in embryo-maternal communication. One of these non-coding RNAs are miRNAs, that are secreted by both embryo and endometrium, and are involved in embryo-maternal communication during the implantation step (Paul et al., 2019; Vilella et al., 2015). To protect signaling molecules from degradation in extracellular space, they are packed

into extracellular vesicles (EV). EVs are eukaryotic cell-derived vesicles of different sizes that carry signaling molecules like miRNAs and other non-coding RNAs, mRNAs, and proteins. They are found in various biofluids and they mediate intercellular communication between cells and tissues (Giacomini et al., 2017; Homer et al., 2017; Machtinger et al., 2015). Elisa Giacomini and coauthors have shown that endometrial stromal and epithelial cells uptake blastocyst EVs derived from IVF embryo spent media of 5-day old embryos, and that vesicles absorbed by trophoblasts enhance trophoblast adhesive capacity, providing evidence of an intracellular mechanism of embryo-maternal communication in humans (Giacomini et al., 2017). Additionally, there is some evidence that endometrial microbiota also influences the success of embryo implantation and placental bacterial community could play some role in communication between mother and embryo (Kyono et al., 2018; Moreno et al., 2016).

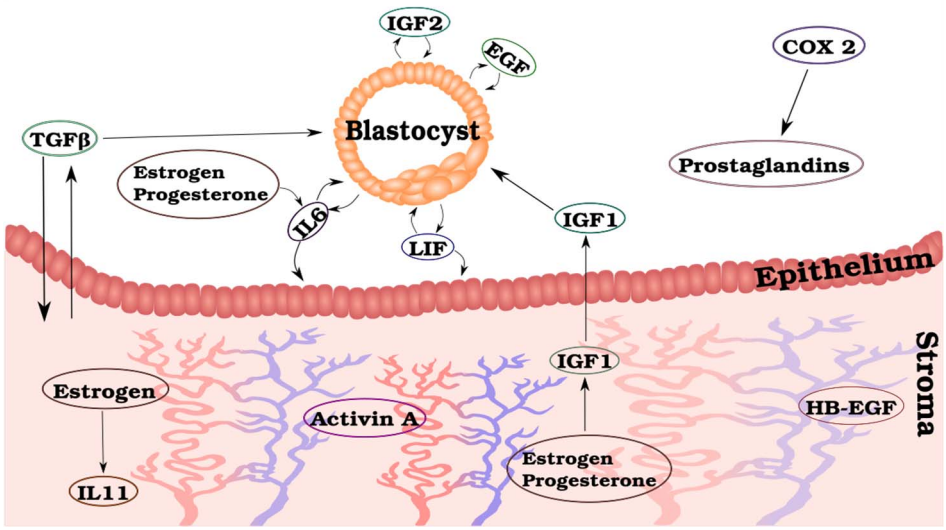

Figure 4. Visualization of some factors involved in implantation. IL11 and Activin A promotes endometrial decidualization. HB-EGF promotes endometrial cellular proliferation, decidualization, and glandular secretion. Estrogen and progesterone regulate IGF1 production in the endometrium, regulate IL6 secretion and facilitate blastocyst attachment. TGFβ in endometrium promotes proliferation, decidualization, implantation, and remodulation of the endometrium. In the case of an embryo, TGFβ increases invasiveness, promotes trophoblast adhesion and pre-/post-implantation development of the embryo. COX-2 regulates prostaglandins, that increase vascular permeability and adhesiveness of endometrium, and promotes embryo implantation. Based on Singh *et al.* (Singh et al., 2011).

In conclusion, successful embryo implantation depends on various factors like blastocyst quality, endometrial receptivity state, synchronization of embryo developmental stages, and successful communication between maternal and embryonal cells (Kao et al., 2002; Singh et al., 2011).

1.5. Transcriptomics methods

It is shown that inadequate uterine receptivity is the reason for two-thirds implantation failures in IVF cycles, and one third is due to embryo factors (Hu et al., 2014). Knowing receptivity markers would help to improve IVF success rates, and one way of doing this is to study transcriptomic changes during the transition from early-secretory (ESE) to mid-secretory endometrium (MSE) in healthy fertile women.

The first attempts to study the whole transcriptome were undertaken in the early 1990s, and by the end of the decade, transcriptomics became a widespread discipline (Hrdlickova et al., 2017; Lowe et al., 2017). The most popular methods to study the transcriptome are microarrays and RNA sequencing (RNA-seq), but which were preceded by expressed sequence tag (EST) and serial/cap analysis of gene expression (SAGE/CAGE). Both microarray and RNA-seq are based on the methods used for DNA analysis, thus before transcriptome analysis, the RNA is reversely transcribed to complementary DNA (cDNA) (Figure 5) (Hrdlickova et al., 2017; Lowe et al., 2017).

The samples usually used for transcriptome studies are tissue biopsies or "bulk" tissue, blood, cultured cells or single cells. Biopsies are used because each tissue has its own characteristic expression signature. While blood is the easiest tissue to get, it usually lacks transcripts originating from tissues/cells other than blood cells, so it is better to get a biopsy from the tissue of interest rather than use blood. The cultured cells or single cells are mainly used to observe cell-specific expression patterns, especially in the case of complex tissues such as cancers or endometrium (Kanter and Kalisky, 2015). Endometrial receptivity expression studies have been conducted using mainly microarrays (Altmae et al., 2010; Borthwick et al., 2003; Díaz-Gimeno et al., 2014; Kao et al., 2002; Mirkin et al., 2005; Riesewijk, 2003), but recently some studies have used RNA-seq (Hu et al., 2014; Marí-Alexandre et al., 2016) or single-cell RNA-seq (Lucas et al., 2018).

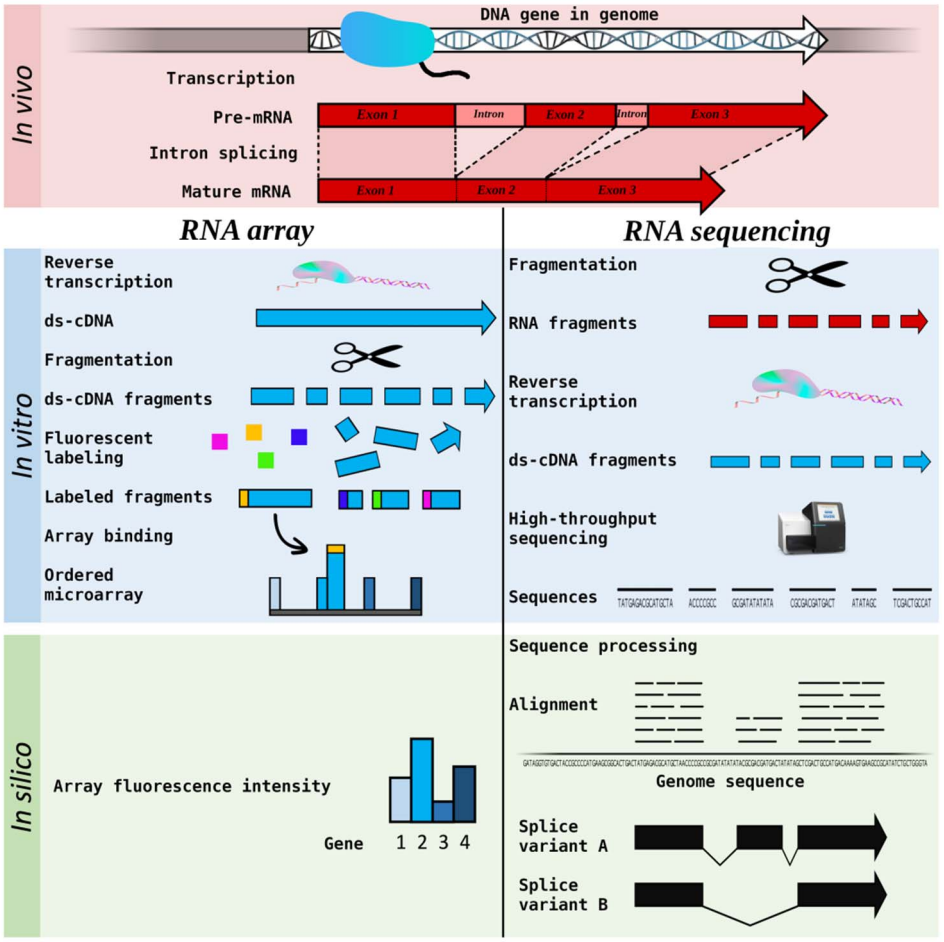

Figure 5. Comparison between RNA microarray and RNAseq technologies. In eukaryotes, genes are transcribed and spliced to produce mature mRNA (red). In *in vitro* analysis, the RNA is extracted, and in the case of microarrays, reverse transcribed to cDNA (blue) and then fragmented to double-stranded cDNA (ds-cDNA) fragments, while for RNA-seq, the extracted RNA is fragmented and then fragments are reverse transcribed, resulting in ds-cDNA fragments. During the following step, array ds-cDNA fragments are labeled with fluorescent tags (orange, dark blue, green, pink) and labeled fragments bind to the ordered array of complementary oligonucleotides, while RNA-seq ds-cDNA fragments are sequenced with high-throughput, short-read sequencing methods. In the last step – *in silico* analysis, arrays fluorescence intensities are measured, showing an abundance of sequences that represent genes of interest, or RNA-seq sequences are aligned to a reference genome or *de novo* assembled to reconstruct transcribed genome regions. Later this data could be used for gene annotation, expression level measurements, and detection of splice variants. Adapted from Lowe *et al.* (Lowe et al., 2017).

1.5.1. Expression microarrays vs RNA sequencing

Many studies compared RNA-seq with expression arrays and found that the advantages of microarrays are low price and easier data analysis. On the other hand, RNA-seq is more precise and sensitive to fluctuations of low gene expression. It has also a broader range of detectable transcripts and their isoforms due to saving strandedness and is capable of identifying single nucleotide variations (SNV) and genetic polymorphisms. In comparison to microarrays, RNA-seq needs less RNA material (RNA-seq ~1 ng, array ~1 μg), could be used without a reference genome and assembled *de novo*, thus it is a good tool for measuring genome-wide expression even in organisms without any reference data (Hrdlickova et al., 2017; Lowe et al., 2017; Marioni et al., 2008; Zhao et al., 2014). Microarrays require special reference transcripts for probes, and so are not suitable for poorly studied organisms and detecting novel transcripts (Lowe et al., 2017). One advantage of RNA-seq is the DNA sequencing base, hence the progress of genome sequencing technologies also results in the progress of RNA-seq methods.

In spite of all advantages, RNA-seq has also some shortcomings like high variability between technical replicates with low coverage (McIntyre et al., 2011). In contrast to microarray studies, where probe annotations are provided by the manufacturer, data processing and analysis protocols are well established, and the results are limited to already known transcripts, RNA-seq needs properly chosen analysis methods, depending on analyzed species, data quality, availability of samples, and scientific question (Schurch et al., 2016; Seo et al., 2016; SEQC/MAQC-III Consortium, 2014). For example, for differential analysis edgeR (McCarthy et al., 2012; Robinson et al., 2010) and DESeq2 (Love et al., 2014) are recommended, in case there are <12 replicates available, and DESeq (Anders and Huber, 2010) if more than 12 (Schurch et al., 2016).

It has been shown that current microarrays, like Affymetrix HTA 2.0, are as good as RNA-seq in detecting protein-coding RNAs and some non-coding RNAs, and in some cases, where there is a need to find the slightest changes in gene expression especially for low abundant transcripts, could even outperform RNA-seq, thus it is advised to use microarrays for differential expression analysis of known genes, while RNA-seq for thorough analyses with detection of unknown genes, and both platforms together in the case of alternative splicing analysis (Nazarov et al., 2017).

1.5.2. RNA-seq data processing and analysis

Before downstream analysis, such as differential expression analysis can take place, the raw RNA-seq reads should be prepared properly. The preparation steps usually include quality check; adapter trimming; removal of low-quality reads, where the low quality corresponds to base call accuracy; alignment to known reference or *de novo* assembly if necessary; quantification on the gene, exon or transcript level; and in the end differential expression analysis, if the aim was to find expression differences between conditions (Figure 6) (Lowe et al., 2017).

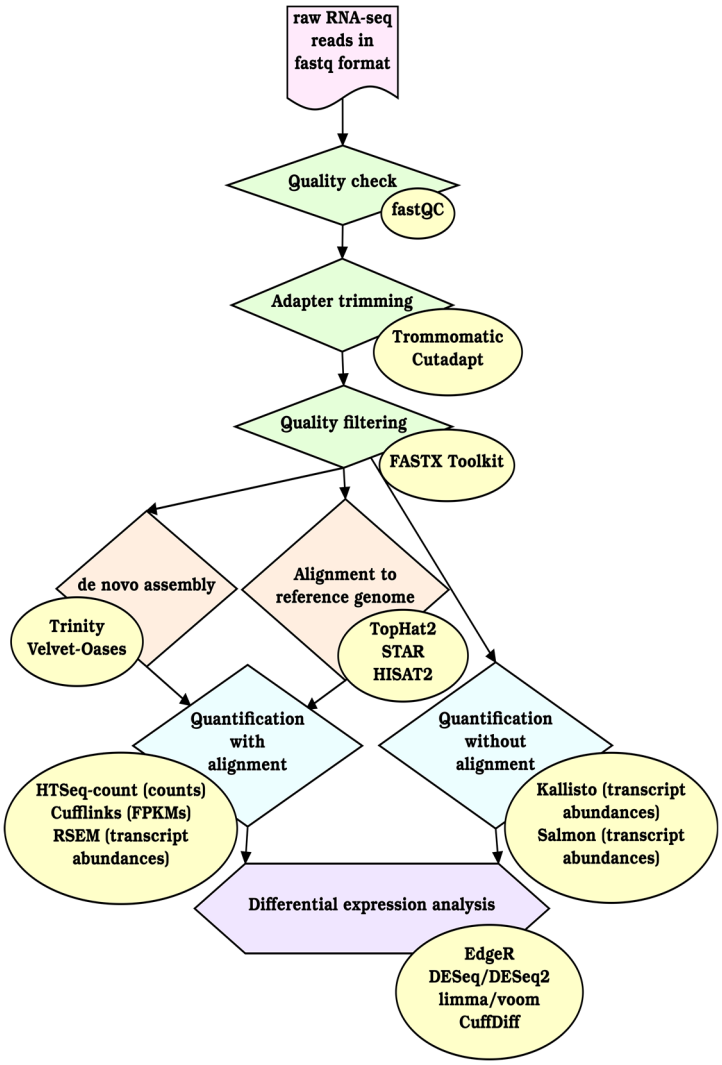

Figure 6. Preprocessing, -mapping and -analysis scheme for RNA-seq data. Data is colored in pink, preprocessing steps are green, alignment/assembly steps – orange, quantification steps – blue, tools – yellow, and differential expression analysis is purple. First, the raw reads are quality checked, next if necessary are trimmed and quality filtered. Then aligned to the reference genome or assembled *de novo*. The reads may be quantified with or without alignment/assembly step. After quantification, the expression of transcripts is analyzed, providing differentially expressed transcripts as a result.

First of all, a quality check of raw reads is needed to get an overview of read base quality, adapter sequence presence, the amount and origin of overrepresented sequences (Lowe et al., 2017), and one of the most popular tools for this step is the FastQC software (Andrews, 2010). This step helps to correct data for further steps before alignment, as low-quality data influences alignment quality,

and thus may lead to poor analysis results. The trimming step is necessary to remove adapters or low-quality bases from the beginning and end of the read, thus eliminating incorrect alignment. During the trimming step, it is also possible to remove overrepresented sequences, depending on the trimming tool. For example, Trimmomatic (Bolger et al., 2014) also includes a low-quality read removal option. The removal of low-quality reads improves further alignment – it lowers the amount of multi-mapped or wrongly mapped reads. One of the tools that could be used for this step is the FASTQ quality filter tool from the FASTX-Toolkit (Hannon lab, 2009).

When the data is filtered and trimmed, it is ready for the next step – alignment to a suitable reference or for *de novo* assembly (Lowe et al., 2017). Many tools are available for alignment, the most popular of them are TopHat2 (Kim et al., 2013) or STAR (Dobin et al., 2013), and for *de novo* assembly, Trinity (Grabherr et al., 2011) or Velvet-Oases (Schulz et al., 2012). The alignment step can be skipped using alignment-free quantification tools like Kallisto (Bray et al., 2016) and Salmon (Patro et al., 2017).

Alignment is followed by quantification of expression on the gene, exon, or transcript levels (Lowe et al., 2017). One broadly used tool to quantify expression on gene or exon levels is the HTSeq-count script from the HTSeq package (Anders et al., 2015). This script provides counts that can be further used in differential expression analysis. Another tool – Cufflinks – provides FPKMs (fragments per kilobase of transcript per million fragments mapped) (Trapnell et al., 2010, 2012), in which case the transcript length is taken into account, in contrary to count data that does not account for transcript length. The RSEM (Li and Dewey, 2011) package is the best for *de novo* assembled genomes, as it quantifies transcript abundances and does not rely on reference genomes.

RNA-seq data can be used for a wide variety of analyses, most commonly for differential expression analysis between some study groups; allele specific-expression analysis; or splicing site analysis (Lowe et al., 2017). Most popular tools for the differential expression analysis are: for the FPKM based analysis – Cuffdiff, a tool from Cufflinks package (Trapnell et al., 2010, 2012); for the count data, several R packages mentioned above, like edgeR, DESeq/DESeq2, DEGSeq (Wang et al., 2010), limma/voom (Costa-Silva et al., 2017; Law et al., 2018; Lowe et al., 2017) are available. Data normalization is carried out before analysis, usually, some normalization methods are included in the package. The normalization is necessary for accounting for factors that could influence mapped read numbers, such as gene length, GC-content, sequencing depth, thus increasing variability between and within samples (Evans et al., 2018). The differential analysis may be carried out using various statistical models depending on the package (Costa-Silva et al., 2017), for example, generalized linear models are used by edgeR (McCarthy et al., 2012), DESeq/DESeq2 (Anders and Huber, 2010; Love et al., 2014), baySeq (Hardcastle and Kelly, 2010), and limma (Ritchie et al., 2015). After obtaining a list of differentially expressed genes (DEGs), at least some of them should be validated with another method, like qPCR (Costa-Silva et al., 2017; Lowe et al., 2017).

1.5.3. Factors affecting RNA-seq analysis

When planning an optimal study design, it must be kept in mind that RNA-seq analysis results can be affected by biological factors stemming from study sample donor's phenotype, life-style and environment; or technical factors from sample collection, storage and preparation phases; or data preprocessing and analysis methods. In the early steps of sample collection and RNA preparation, the time between blood/biopsy collection and RNA isolation, or RNA isolation protocols, reagents used and their quality, laboratory and personnel habits and experience – all of them are a possible source of batch effects, that may lead to technical variability resulting in gene expression differences bigger than the differences between measurable conditions (Leek et al., 2010; Schurmann et al., 2012; Williams et al., 2014). In the next steps, the nucleic acid amplification, platforms used for that, sample preparation conducted on separate date and time, library, and even sequencing lane affect further analysis outcome through different batch effects (Leek et al., 2010; Schurmann et al., 2012; Williams et al., 2014). Thus, it is important to eliminate, minimize or at least take into account technical effects on early analysis steps, for example, by indexing and multiplexing samples to minimize line effects (Williams et al., 2014), using the same protocols and reagents for sample preparation, and/or reducing the time between tissue collection and RNA isolation to minimize possible RNA degradation. Recording changes in sample preparation could help to take these differences into account during the analysis step (Leek et al., 2010; Schurmann et al., 2012). The extracted total RNA on average has 90% of ribosomal RNA (rRNA) and only 2% of mRNA, so there is a need for mRNA enrichment using polyA selection, or rRNA level reduction using rRNA depletion methods (Hrdlickova et al., 2017; Nazarov et al., 2017; Sheng et al., 2016). While the polyA selection leads to 3' bias in the distribution of reads, the depletion may lead to unpredictable changes at exon-level (Nazarov et al., 2017). These biases could be reduced during counting or normalization steps by using appropriate statistical models (Tuerk et al., 2017).

After sequencing comes the data preprocessing step that includes quality control and other preparation steps before mapping. It is necessary to check the GC-content, overrepresented sequences, adapter content, and base quality, because all these parameters could affect the downstream sequence alignment (Lowe et al., 2017; Sheng et al., 2016; Williams et al., 2014). Data can be improved before sequence alignment by trimming out the adapters, removing bad quality bases in the read start/end, or filtering out overall bad quality reads (Lowe et al., 2017; Sheng et al., 2016). The alignment step could affect analysis due to falsely aligned reads or multi-mapped reads. It is important to use options suitable for dataset and analysis of interest, and check the quality of aligned data (Sheng et al., 2016; Williams et al., 2014). Quality check gives a good overview of the numbers of unmapped and mapped reads, insert size, multi-mapped reads and it helps to detect 3' or 5' bias, GC bias, and batch effects (Sheng et al., 2016). During the counting step, only uniquely mapped reads should be counted. The choice of counting tool is made depending on the

aims and analysis type, because in some cases transcript length may affect analysis results (Lowe et al., 2017; Nazarov et al., 2017; Sheng et al., 2016; Williams et al., 2014). The read count approach is usually used for differential expression analysis on gene or exon level.

The analysis results in the case of 'bulk' tissue could be influenced by cellular composition, which varies from sample to sample (Gong et al., 2011; Schelker et al., 2017; Shen-Orr and Gaujoux, 2013). Several approaches have been developed to address this issue, either using physical cell sorting methods (such as flow cytometry) to obtain pure cell fractions for transcriptomic analysis, or computational deconvolution, which uses various statistical approaches to calculate estimated cellular fractions.

Physical cell sorting is more complicated and needs special instruments, but is more accurate in comparison to computational deconvolution. There are several techniques available to obtain enriched cellular populations: fluorescence-activated cell sorting (FACS), magnetic-activated cell sorting (MACS), laser capture microdissection (LCM), manual cell picking and microfluidic isolation (Figure 7) (Hu et al., 2016). FACS (Figure 7 A) is the most sophisticated and user-friendly technique, that allows simultaneous quantitative and qualitative single-cell analysis based on cell size, granularity and fluorescence antibody tagging of cell-specific antigens. The fluorescent probe labeled cells are exposed to laser while running through cytometry, so fluorescence detectors could distinguish between cell types based on user predefined characteristics. The droplets with cells of interest are then charged and the electrostatic system directs identically charged droplets into appropriate collection tubes (Hu et al., 2016). MACS (Figure 7 B) is another commonly used technique, where cells are labeled with magnetic beads wearing antibodies. The external magnetic field recognizes MACS bead conjugated antibodies, and thus polarize labeled cells while unlabeled cells are washed out (Hu et al., 2016). LCM (Figure 7 C) is based on microscopic visualization and the system consists of an inverted microscope, controlled microscope stage with vacuum capability, a solid-state near laser diode, laser controller, CCD (Charge Coupled Device) camera and monitor. The cells of interest are visualized with a microscope and then the transparent thermoplastic film on the cap above the cells is melted by laser impulse and fused with the targeted cells. The film with targeted cells is then removed, thus separating them from the "bulk" tissue (Hu et al., 2016). Manual cell picking (Figure 7 D) is also based on microscopic visualization, but instead of laser with melting film, the microscope is equipped with micro-pipettes, which allows isolating live cells (Hu et al., 2016). Alternatively, manual cell-picking can also be performed without micro-manipulation. Microfluidic chip (Figure 7 E) is based on fluid physico-chemical properties, that change in micro-channels. The chip channels are modified so they can bind targeted cells, while other cells are washed out with buffer (Hu et al., 2016).

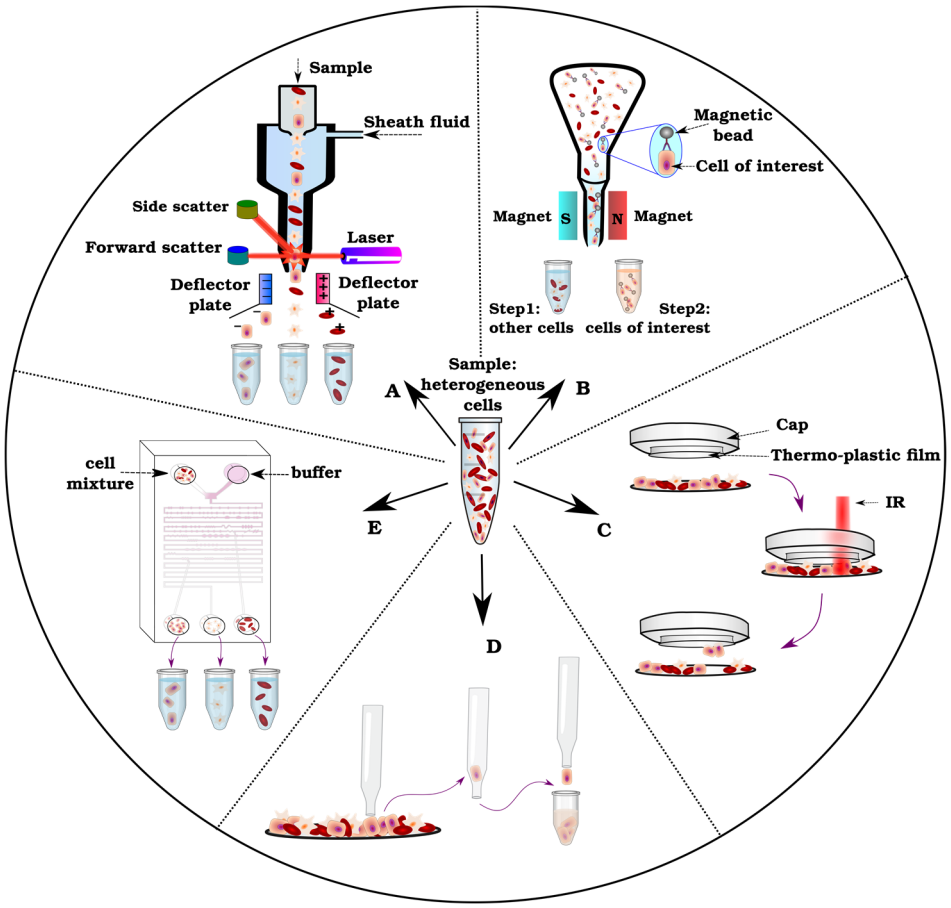

Figure 7. Schematic of five cell sorting methods. **A.** FACS. Labeled cells are exposed to laser as a stream of single-cell droplets. Then the fluorescence detector charges cells according to their fluorescence and light scatter characteristics. The electronic deflection system further guides the charged cells according to their charge into the collecting tubes. **B.** MACS. The fusion of antibody with magnetic beads is used to label cells of interest. Labeled cells are separated from the cell mixture with a magnetic field (S and N magnets) and collected to the tube. **C.** LCM. The cap with the thermoplastic film is moved above the cells. Then the infrared (IR) laser impulse is sent through the cap over cells of interest melting the film. The cells of interest adhere to the film and are removed with the cap from the cell mixture. **D.** Manual picking. Cells of interest are collected with a micromanipulator connected glass pipette under a microscope and transferred to the tube. **E.** Approximate microfluidic chip separation. In general: using a system of capillaries and buffer fluid, the cells are separated from cell mixture and collected to the collecting tubes. Based on Hu *et al.* (Hu et al., 2016).

There are several methods to computationally estimate cell fractions in the whole tissue, most of which are developed for whole blood samples (Abbas et al., 2009; Gong and Szustakowski, 2013; Liu et al., 2008a; Newman et al.,

2015; Qiao et al., 2012; Shen-Orr and Gaujoux, 2013). These statistical methods need only expression datasets (RNA-seq or microarray) and sometimes cell-specific RNA expression profiles, but in comparison to physical methods provide estimated cell fractions or arbitrary units (Gong and Szustakowski, 2013; Newman et al., 2015; Qiao et al., 2012; Schelker et al., 2017; Sturm et al., 2018). There are reference-free and reference-based methods, where first ones require a cell-type-specific signature and 'bulk' tissue expression data, while second ones need only 'bulk' tissue expression data (Avila Cobos et al., 2018; Teschendorff and Zheng, 2017). The cell-type-specific expression set usually includes cell-type-specific whole-genome expression patterns or some cell-type-specific signature genes (Qiao et al., 2012; Sturm et al., 2018; Teschendorff and Zheng, 2017).

In the case of the DeconRNASeq, the "bulk" tissue dataset should be limited only to transcripts that are present in cell-specific dataset (Gong and Szustakowski, 2013). Also, DeconRNASeq uses normalized transcriptional measurements from pure cell types and heterogeneous tissues, where measurements could be made on the gene, transcript or exon level, that are all provided by the user, so there are no limitations to cell and/or tissue types (Gong and Szustakowski, 2013). At the same time, CYBERSORT has its own set of reference genes for 22 hematopoietic cells and the user should provide only expression profiles for whole blood samples (Newman et al., 2015). The limitation of reference-based methods is reference availability and quality. For example, the overlap of signature genes between cell types could lead to spillover effect, where a high abundance of one cell type could lead to predicting higher abundance for another cell type (Sturm et al., 2018). The quality of the signature set also affects distinguishing between cell subpopulations (Qiao et al., 2012; Sturm et al., 2018). Another limitation is not taking into account cell-cell interactions, that could alter gene expression profiles (Teschendorff and Zheng, 2017).

The reference-free methods, such as surrogate variable analysis (SVA) or independent SVA (ISVA) that adjust the data to any type of confounder (Teschendorff and Zheng, 2017), deconvolute the expression data without prior knowledge about cell-types and can account for cell-cell interactions (Avila Cobos et al., 2018; Teschendorff and Zheng, 2017). Wang and colleagues provide a new deconvolution method implemented into CAM (convex analysis of mixtures) package that makes unsupervised deconvolution without using any cell type reference sets and needs only normalized gene expression values with *a priori* removed batch effects (Wang et al., 2016). The method is based on a theorem, where the scatter simplex (linear subspace) of mixed (bulk tissue) expressions is the compressed and rotated scatter simplex of pure (cell-type-specific) expressions. The expression of a separate gene in a cell-type is modeled as being linearly proportional to the abundance of that cell-type. CAM geometrically identifies vertices (corners) of mixed gene expression scatter simplex, where each vertex corresponds to cell-type-specific expression and contains cell-type signature genes. The method works accurately only if the number of heterogeneous samples is equal to or greater than the number of cell types in these

samples (Wang et al., 2016). LinSeed is another package for unsupervised deconvolution using a similar approach (Zaitsev et al., 2019).

The estimated cellular fractions obtained from deconvolution analysis can be used in the downstream differential expression analysis. In the bulk-tissue samples, tissue-averaged expression levels are measured, thus expression contribution of less abundant cell-types could be masked by that of more abundant cell-types (Avila Cobos et al., 2018); therefore, the adjustment according to cellular composition helps to prevent loss of signal from less abundant cell types, thus improving the quality of the analysis.

1.6. Role of gene expression in endometrial receptivity context

Studies into the transcriptomic changes taking place around the WOI have provided valuable insights into the underlying reproductive biology. These studies have analyzed the differences in expression between early-secretory (ES) and mid-secretory (MS) phases, and have usually used samples from women aged between 20–40 years as in this age interval menstrual cycle of healthy women is stable and women are most fertile (Hu et al., 2014; Kao et al., 2002; Koot et al., 2016; Liu et al., 2008b; Riesewijk, 2003; Ruiz-Alonso et al., 2013; Sherman and Korenman, 1975; Zhang et al., 2012a). The participants should have normal BMI, as bodyweight can influence fertility (Koot et al., 2016; Ruiz-Alonso et al., 2013). They are not allowed to take hormonal contraceptives for at least 3 months before the study, because the expression of many genes is influenced by hormonal levels (Hu et al., 2014; Koot et al., 2016; Liu et al., 2008b; Zhang et al., 2012a). Usually, participating women have at least one live-born child that indicates their fertility (Altmae et al., 2010). Ovulation tests are used to track the LH peak and histological dating according to Noyes's criteria (Noyes et al., 1950, 1975) is used to evaluate endometrial phase (Altmae et al., 2010; Kao et al., 2002; Riesewijk et al., 2003; Zhang et al., 2012a). The smoking status is also taken into account (Koot et al., 2016; Riesewijk et al., 2003).

1.6.1. Endometrial receptivity microarray studies

Before the gene expression array era, the studies of transcription mainly concentrated on one or several genes. The invention of microarrays in 1990 and their wide-spread use made it possible to study the transcription of many genes at once (Marioni et al., 2008; Zhao et al., 2014). The first endometrial microarray studies were published in 2002, when Carson *et al*. and Kao *et al.* used *ca* 12k gene micro-arrays to find differences between ESE and MSE (Carson et al., 2002; Kao et al., 2002). They found DEGs associated with the extracellular matrix, signaling pathways, transcription factors (Carson et al., 2002; Kao et al., 2002), growth factors/cytokines (Carson et al., 2002), immune modulators (Kao

et al., 2002) and others. Their works were the first to show the complexity of the transition process between ESE to MSE. Some further studies tried to link gene expression differences with progesterone and estrogen level fluctuations by searching progesterone and estrogen receptor DNA binding sites in DEGs (Borthwick et al., 2003; Mirkin et al., 2005). These studies used a small number of samples, each time-point was collected from different women, and they did not take into account the effects of BMI, and smoking status (Borthwick et al., 2003; Carson et al., 2002; Kao et al., 2002). A more advanced sample collecting approach was used by the Riesewijk group – biopsies for ESE and MSE were collected from the 5 women within the same menstrual cycle, which helped to eliminate interindividual variability. Their DEGs showed some overlap with previous studies, where up-regulated DEGs were more overlapping with Kao *et al*. work and down-regulated DEGs showed bigger overlap with Carson *et al*. study (Carson et al., 2002; Kao et al., 2002; Riesewijk et al., 2003). The overlap between three studies was low due to differences in samples/sample collecting, microarrays used, analysis techniques used and are pointed out in further works (Riesewijk et al., 2003; Zhang et al., 2012). Further studies used a bigger number of samples (Mirkin et al., 2005), used animal models (Vallejo et al., 2010), looked into differences because of IVF (Liu et al., 2008b) or between fertile women and infertile patients (Altmae et al., 2010), found some overlap with the previous studies (Altmae et al., 2010; Liu et al., 2008b; Mirkin et al., 2005; Vallejo et al., 2010) and added new pieces of information into the endometrial regulation puzzle at the gene expression level. Still, the overlap between array studies remained low (Gómez et al., 2015; Zhang et al., 2012).

1.6.2. The era of ERA (Endometrial Receptivity Array)

Endometrial receptivity transcriptome studies have also had a practical output, as in 2010, Diaz-Gimeno and co-workers created a new tool for predicting endometrial receptivity and the WOI. The tool consisted of an array of 238 selected genes expressed at three stages of the endometrial cycle and a bioinformatic predictor. The bioinformatic predictor was based on a support vector machine algorithm and was trained for two conditions: endometrial receptivity (ES, MS, proliferative) and pathological classifications (implantation failure, hydrosalpinx) (Díaz-Gimeno et al., 2011). They also compared the test with classical histological dating according to Noyes criteria (Noyes et al., 1950, 1975). The ERA test made better predictions than histologists in most cases and was reproducible in time (Díaz-Gimeno et al., 2013). The ERA test was tested on recurrent implantation failure (RIF) patients, as the implantation failure may be caused by WOI misplacing. The implantation and pregnancy rates were higher after ERA test implementation, thus helping to plan embryo transfer according to personalized WOI timing (Ruiz-Alonso et al., 2013). However, RIF is still too complex to solve only with ERA testing and embryo quality check (Koot et al., 2016). In 2017, Encisco and colleagues came out with

another WOI timing test, that is based on RT-qPCR, and is named as ER Map®/ER Grade®. The test includes 40 genes that were selected from the literature based on the highest fold change between non-receptive and receptive phases, and additionally the genes that explained almost 100% of the sample variance according to PCA (principal component analysis) results. The overlap with the ERA gene set is relatively small, only *ANXA4, AQP3, ARG2, GPX3, MAOA, MT1H,* and *SCGB2A2* genes are present in both tests, but despite such small overlap, ER Map®/ER Grade® classification matched with ERA results in 97.59% of the training set and 91.67% of the testing set samples (Enciso et al., 2018).

1.6.3. Endometrial receptivity RNA-seq studies

Only some endometrial studies have used RNA sequencing, possibly because the method is more expensive and requires more complicated data preparation and analysis in comparison with array-based studies. The first differential expression study between ESE and MSE using the RNA-seq technique was made by Hu and colleagues (Hu et al., 2014). They used samples from 12 women, 6 per both ES and MS phase (6 of them – 3 samples per phase – were used for the main analysis, and the other 6 as validation set). They found 1,099 up-regulated and 1,237 down-regulated genes, which correlated with results from the validation set (2,212 genes overlapped; Pearson correlation between fold changes: R=0.9994). There was also gene overlap with previous array studies: with Zhang et al. (Zhang et al., 2016) 144 of the 148 genes (97%) and with Tapia et al. (Tapia et al., 2007) 55 of the 61 genes (90%) (Hu et al., 2014). The second study using RNA-seq for detecting transcription during the transition between ESE to MSE was more focused on noncoding RNAs. They found 3,297 significantly differentially expressed mRNAs, of which 1,408 overlapped with the previous study (Hu et al., 2014), showing that despite differences in sample sizing and methodology, there is bigger concordance between RNA-seq studies than array studies (Sigurgeirsson et al., 2016). The gene expression studies mentioned above have their limitations, most importantly small sample sizes. Rahmioglu *et al.* showed that at least 500 samples are needed for detection of 1.5 FC (fold change) gene expression in *ca* 80% of transcripts (Rahmioglu et al., 2017), therefore endometrial RNA-seq studies conducted thus far have been underpowered.

Most of the studies have used endometrial whole-tissue biopsies, where cell proportions vary, as the endometrium is a complex tissue with remarkable structural changes happening in different phases of the menstrual cycle, and also the biopsy size and intrauterine location vary. Cell composition should be taken into consideration during analysis, to avoid bias towards the expression pattern of cells with a larger proportion. The participating women's lifestyle, such as smoking, or some medical treatments as hormonal therapy, or women's age could drastically influence the results, so should also be taken into account.

Thus, there is still a need for more studies using larger sample sets and proper methodology to find more robust support for previous findings.

1.6.4. Non-coding RNAs

There are a lot of other non-coding RNA molecules transcribed from the genome together with coding RNAs, that play an important role in genome regulation (Chen et al., 2017a; Hüttenhofer et al., 2005; Moazed, 2009; Montazerian et al., 2018; Tesfaye et al., 2018; Trzybulska et al., 2018). Non-coding RNAs are grouped into small non-coding RNAs, like short interfering RNAs (siRNA); miRNAs; PIWI-interacting RNAs (piRNAs); and into long non-coding RNAs (lncRNAs), like long intergenic noncoding RNAs (lincRNAs) and transcribed ultraconserved regions (T-UCRs) (Hüttenhofer et al., 2005; Moazed, 2009; Trzybulska et al., 2018). For example, lncRNAs may regulate allele-specific expression, such as X chromosome inactivation and imprinting; may act as recruiter factors; regulate other lncRNAs (Cedar and Bergman, 2009; Hüttenhofer et al., 2005; Kung et al., 2013; Trzybulska et al., 2018); interfere with miRNA-mediated mRNA destabilization by masking miRNA binding-sites or acting as miRNA "sponges" (Kung et al., 2013). It has been shown that lncRNAs can also contribute to endometrial cancer development (Chen et al., 2017a; Sigurgeirsson et al., 2016). The small RNAs also may act at the DNA and chromatin level (Hüttenhofer et al., 2005; Moazed, 2009).

MiRNAs, small RNAs that recently drew interest in the reproductive context, are 18–24 bp short protein non-coding molecules, that can regulate intracellular gene expression locally in parent cells or may influence it distantly via transfer by blood flow, making them good disease biomarker candidates (Hull and Nisenblat, 2013; Montazerian et al., 2018; Paul et al., 2018; Tesfaye et al., 2018; Trzybulska et al., 2018). miRNAs are synthesized in three steps: first, primary miRNA is transcribed from the genome; second, it is cleaved to pre-miRNA; and third, it is transported into the cytoplasm, where it becomes a mature miRNA (Figure 8) (Hull and Nisenblat, 2013; Montazerian et al., 2018; Paul et al., 2018; Tesfaye et al., 2018). To regulate gene expression, the mature miRNA binds to the Argonaut proteins in the RNA-induced silencing complex (Hull and Nisenblat, 2013; Hüttenhofer et al., 2005; Moazed, 2009; Paul et al., 2018; Tesfaye et al., 2018; Trzybulska et al., 2018). One miRNA could regulate hundreds of transcripts, while one transcript could be regulated by several miRNAs (Altmäe et al., 2013; Hull and Nisenblat, 2013; Seitz, 2009). These small molecules mainly down-regulate the expression of mRNAs by cleavage and deadenylation or up-regulate by targeting gene promoters (Montazerian et al., 2018; Paul et al., 2018). The miRNAs are circulating in blood flow as part of microvesicles or lipoprotein complex compound, thus they are protected from degradation by RNases (Hull and Nisenblat, 2013; Montazerian et al., 2018; Trzybulska et al., 2018).

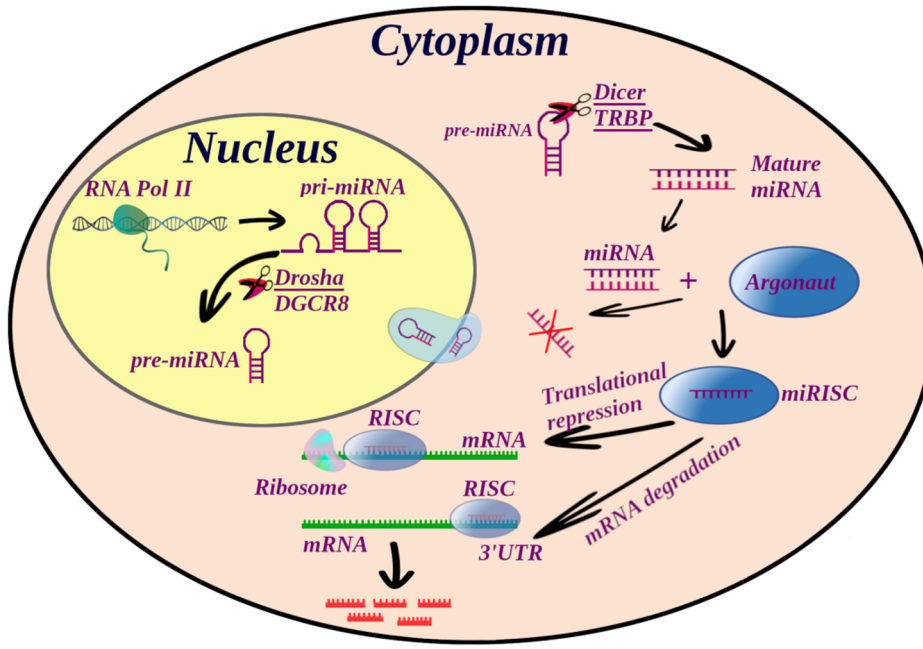

Figure 8. Overview of miRNA synthesis: pri-miRNA is transcribed in the nucleus from the gene, then cleaved to pre-miRNA by Drosha enzyme and transported to the cytoplasm where it is cleaved by Dicer to mature miRNA. The mature miRNA may bind Argonaut and thus form the RISC complex. This complex can terminate protein translation by binding to the open reading frame or lead to mRNA degradation when binding with 3'UTR.

The expression of endometrial miRNAs is cell-specific, thus it probably mediates cross-talk between endometrial cell-types (Hull and Nisenblat, 2013; Montazerian et al., 2018). Altmäe and colleagues found miRNAs, differentially expressed in endometrial transition from ES to MS phase, that regulate genes involved in epithelial and endothelial cell differentiation, implantation, cell communication and migration, and inflammatory responses (Altmäe et al., 2013). Up-regulation of miRNAs in epithelial cells in the MS phase was also demonstrated by others (Kuokkanen et al., 2010; Paul et al., 2018), where some of them are, probably, down-regulating expression of cell-cycle specific genes, thereby suppressing cell proliferation (Kuokkanen et al., 2010). miRNA potential role in creating immune-tolerant environment during the secretory phase and regulation of endometrial decidualization (Paul et al., 2018; Sigurgeirsson et al., 2016), involvement during embryo implantation process and connection with recurrent miscarriage (Feng et al., 2018) and endometriosis has also been shown (Hull and Nisenblat, 2013). Sha and colleagues observed altered miRNA expression in stimulated IVF cycles in comparison with natural cycles, thus

indicating miRNA importance for obtaining endometrial receptivity (Sha et al., 2011). Some miRNAs with altered gene expression had estrogen or progesterone response elements in their promoter region, suggesting direct response of these miRNAs to the ovarian hormones (Sha et al., 2011). In the work of Vilella *et al.*, several miRNAs showed changes in expression during WOI in comparison with the rest of the menstrual cycle; and were actively produced by glandular epithelium, then packed into EVs and secreted into the lumen of the endometrial cavity (Vilella et al., 2015). These small RNAs participate in oocyte maturation and were found in both oocyte and follicular fluid, thus they could mediate communication between the oocyte and somatic cells (Battaglia et al., 2017). miRNAs are also associated with endometrial cancer development, where miRNA expression levels are correlating with expression levels of endometrial cancer-associated genes (Boren et al., 2008). The exposure to harmful environmental factors is playing its role in miRNA regulated infertility and there has been shown, that cigarette smoke exposure induces ovarian dysregulation in mice through several miRNAs (Furlong et al., 2018). In comparison with mRNA miRNAs are more stable and could remain intact longer periods of time, thus have great potential to become robust biomarkers of infertility or some other reproductive disorders (Robles et al., 2019).

Sigurgeirsson *et al.* showed another type of non-coding RNAs – small nucleolar RNAs (snoRNAs), which are involved in posttranscriptional maturation of rRNA and are expressed in MSE, thus may have a potential role in receptivity regulation (Sigurgeirsson et al., 2016). lncRNAs also influence endometrial receptivity, probably through competing for shared miRNAs with other RNA transcripts, and some of them could function as potential biomarkers due to maintaining stability in peripheral blood (Feng et al., 2018). It was shown that lncRNAs are dysregulated in RIF women in comparison to women who conceived after embryo transfer (Fan et al., 2017) and some lncRNAs could potentially target oocyte miRNAs and this way regulate oocyte maturation (Battaglia et al., 2017).

With the development of new methods and technology, there is great potential to find other RNA types influencing the regulation of the menstrual cycle and reproductive health in general. Also, it has been shown, especially in cancer studies, that expression of miRNAs and other non-coding RNAs are regulated through epigenetic mechanisms such as methylation of CpG islands or histone modifications (Strmsek and Kunej, 2015). The epigenetic regulation of gene expression and their regulatory non-coding RNAs indicate the importance of adding new ‘omics’ layer – methylomics, for better understanding of endometrial receptivity mechanisms.

1.7. Methylomics methods

DNA methylation is one of the epigenetic mechanisms for gene expression regulation, thus it is relevant for genomic imprinting, X chromosome inactivation, cell-type-specific expression, gene silencing, suppression of repetitive elements, and alternative splicing (Gao et al., 2012; Jones, 1999, 2012; Lokk et al., 2014; Yamagata et al., 2009).

DNA methylation takes place when methyltransferases transfer the methyl group to 5'-cytosine of CpG (cytosine-guanine pair bound with phosphodiester bond) site, thus accumulating 5'-methylcytosine on DNA strand (Gao et al., 2012; Jones, 1999, 2012; Laird, 2010; Oakeley, 1999; Yamagata et al., 2009). Previously it was shown that in bacterial cells both cytosine and adenine can be methylated (Noyer-Weidner and Trautner, 1993), while in eukaryotes mainly cytosines are methylated (Gupta et al., 2010; Harrison and Parle-McDermott, 2011; Laird, 2010). In humans methylation usually takes place at cytosine bases that are followed by guanines (Bock et al., 2010; Harrison and Parle-McDermott, 2011), thus CpG is the name of methylation site, but in embryonic stem cells also the non-CpG specific methylation can be found (Lister et al., 2009; Ramsahoye et al., 2000). The CpG sites usually aggregate to clusters in gene promoter areas and are called CpG islands (CGI) (Gao et al., 2012; Jones, 1999, 2012; Yamagata et al., 2009). The hypomethylation of such islands usually leads to enhanced gene expression, while hypermethylation results in gene silencing (Gao et al., 2012; Jones, 1999, 2012; Laird, 2010). Still, some CGIs can be found in the coding regions of genes or locations associated with transcriptional start sites, and for some genes in the gene body (the gene region between 1st exon and 3'UTR) regions, where hypermethylation leads to increased expression levels – such phenomenon is called "DNA methylation paradox" (Jones, 1999; Laird, 2010).

There are four methylation/demethylation types: *de novo* methylation, where previously unmethylated cytosines become methylated; methylation maintenance, that takes place during DNA replication; passive demethylation, when maintenance is suppressed during each cycle of DNA replication and DNA strand between methylated and unmethylated states are hemimethylated (there is a methylation mark on one strand and on the other one the methylation mark is absent); and active demethylation, when enzymes decrease methylation levels (Figure 9) (Oakeley, 1999). *DNMT3a* with *DNMT3b* methyltransferases are responsible for *de novo* methylation, while *DNMT1* is responsible for accurately replicating methylation patterns during cell division thus maintaining the correct DNA methylation pattern in mammalian cells (Cedar and Bergman, 2009; Jones, 2012; Wreczycka et al., 2017; Yamagata et al., 2009).

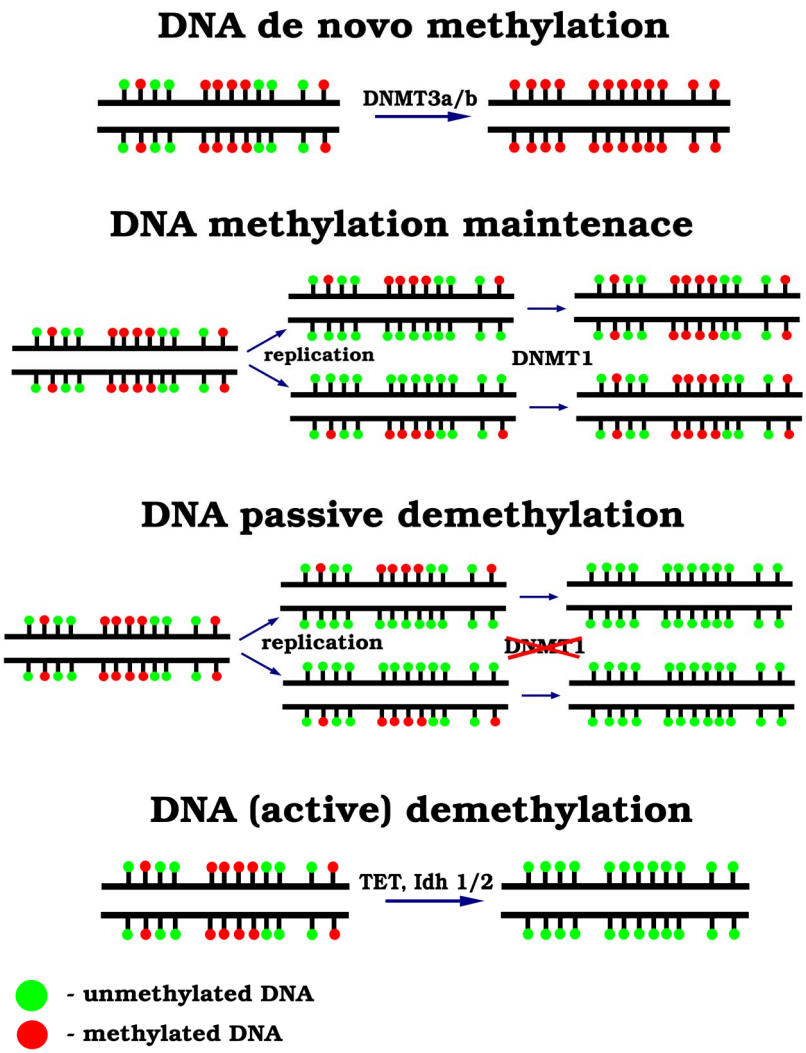

Figure 9. There are four types of methylation/demethylation processes: *de novo* methylation, methylation maintenance, passive demethylation, and active demethylation. The first two and active demethylation are dependent on enzyme activity, while the passive demethylation occurs when methyltransferases are inactive. *De novo* methylation is mainly curated by DNMT3a/b methyltransferases, where methylation marks are directly added to unmethylated cytosines (green circles). Methylation maintenance takes place after DNA replication, where one strand is unmethylated and the other one has methylation patterns from replicated DNA. DNMT1 is the main enzyme that restores the previous methylation pattern in this scenario. If the DNMT1 is inactive, the passive demethylation takes place, where DNA between methylated and unmethylated states is hemimethylated. TET and Idh1/2 enzymes remove methylation marks from methylated cytosines (red circles), thus are responsible for active DNA demethylation.

Methylation processes are tissue-specific and similar tissues cluster together according to methylation patterns; still, as methylation is a dynamic process, it could be influenced by nutritional- (Liu et al., 2003), pathogenic- (Leonard et al., 2012), environmental (Bind et al., 2012) and age-related factors (Madrigano et al., 2012). The tissue-specificity according to Schroeder *et al.* (Schroeder et al., 2013) is associated with partially methylated domains (PMDs), which in some hypomethylated tissues, such as placenta and some cancers, could cover up to 40% of the genome. Genes that are located within PMDs have lower expression levels due to highly methylated promoters and hypomethylated other gene regions, which is true for hypomethylated tissues – placenta in this case, but not for most human tissues that are highly methylated (Schroeder et al., 2013).

1.7.1. Evolution of methods for studying DNA methylation

The methods for studying DNA methylation have evolved (Figure 10) from the simple 5'-methylcytosine amount measurement in the genome to a variety of comparative methods, that were later adapted for arrays and sequencing (Gupta et al., 2010; Harrison and Parle-McDermott, 2011). Methods available before 2000 were gene-specific and non-specific or genome-wide methods. The first ones include sodium-bisulfite reaction, restriction endonucleases (*HpaII/MspI*), ligation-mediated PCR (Maxam-Gilbert sequencing applied), methylation-sensitive single nucleotide primer extension (MS-SNuPE) and combined bisulfite restriction methods like COBRA. The second group included reverse-phase high-performance liquid chromatography (RP-HPLC), thin layer chromatography, *SssI* methyltransferase assay, methylation-sensitive single-strand conformation polymorphism analysis (SSPA) and hybridization-based methods (Fraga and Esteller, 2002; Harrison and Parle-McDermott, 2011; Oakeley, 1999; Shen and Waterland, 2007).

Nowadays methylation arrays and sequencing are based on modified methods such as immunoprecipitation with antibodies or specific methyl-binding proteins, DNA cleavage with methyl-sensitive restrictases and bisulfite treatment followed by PCR amplification and sequencing (Gupta et al., 2010; Harrison and Parle-McDermott, 2011; Laird, 2010; Zuo et al., 2009). While bisulfite treatment-based sequencing is most popular among methylation sequencing methods, there is also a single-molecule real-time sequencing, that allows direct detection of DNA methylation without previous bisulfite treatment (Harrison and Parle-McDermott, 2011; Kurdyukov and Bullock, 2016; Laird, 2010).

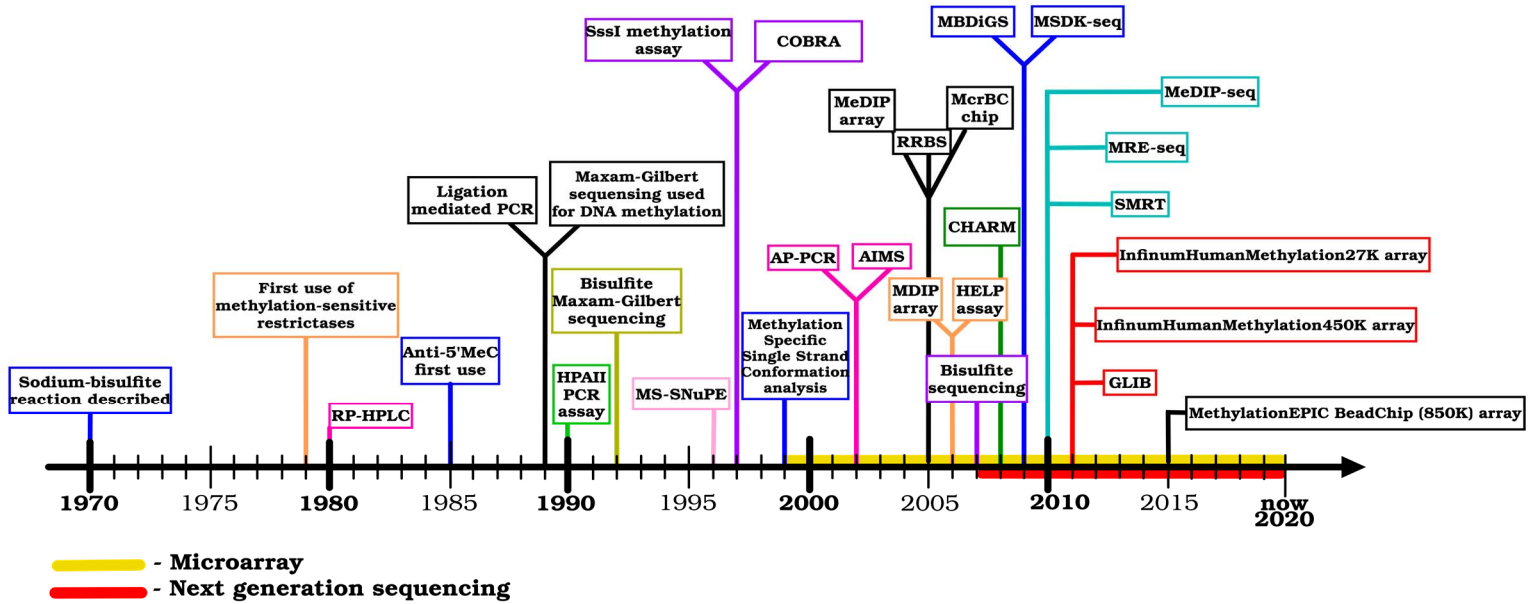

Figure 10. The methylation research methods developed from simple protocols, where the amount of 5'-methylcytosines in the genome was simply measured, to basic comparative methods (the 80s-90s). These basic methods were later adapted for microarray and next-generation sequencing (NGS) technologies during the 2000s, opening opportunities for simultaneous methylation analysis of numerous genomes. Some of them are shown on this time-scale. **RP-HPLC** – reverse-phase high-performance liquid chromatography; **anti-5'-MeC** – anti 5'-methylcytosine; **HPAII** – methylation-sensitive restriction enzyme used in pair with MspI; **MS-SNuPE** – methylation-sensitive single nucleotide primer extension; **SssI** – bacterial methyltransferase; **COBRA** – combined bisulfite restriction analysis; **AP-PCR** – arbitrary primed PCR; **AIMS** – amplification of inter-methylated sites; **MeDIP** – methylated DNA immunoprecipitation; **McrBC** – methylation-sensitive restrictase; **MDIP** – methyl-DNA immunoprecipitation; **HELP** – HpaII (tiny fragment) enrichment by ligation-mediated PCR; **CHARM** – comprehensive high-throughput arrays for relative methylation; **MBDiGS** – methyl-binding domain isolated genome sequencing; **MSDK** – methylation-specific digital karyotyping; **MRE-seq** – methyl-sensitive restriction enzyme sequencing; **SMRT** – submegabase resolution tiling; **GLIB** – glycosylation, periodate oxidation, biotinylation. The figure is mainly based on the publication of Harrison and Parle-McDermott (Harrison and Parle-McDermott, 2011).

1.7.2. Methylation sequencing and microarray data analysis workflows

The broad usage of arrays and NGS on big datasets increased the need for proper analyzing methods and bioinformatic tools for DNA methylation data analysis. The methylation data analyzing methods depend on whether the data is array or sequencing-based. But in both cases, there are common basic steps, such as data normalization, exploratory analysis, and accounting for batch effects.

For the sequenced reads, the first steps are quality control (QC) and adapter trimming, followed by mapping good quality reads to the reference genome, mapped data quality control and methylation calling. The raw read base call QC, adapter removal and end trimming are necessary to prevent false C-T conversion from miscalled bases and increase alignment rates, while the alignment QC helps to detect and eliminate unmethylated Cs at the end of fragments, incomplete bisulfite conversion, DNA degradation and PCR bias (Wreczycka et al., 2017). The last step is the analysis itself, where the most common option is differential methylation analysis on single CpG or region level. There are several tools available for differential methylation analysis that are based on multiple statistical approaches such as Fisher's test, hidden Markov models, linear and logistic regression-based models, and some more complex regression models that use beta-binomial distribution (Wreczycka et al., 2017). Some of the tools have several approaches to choose from, like RnBeads (Assenov et al., 2014) and Bsmooth (Hansen et al., 2012), that use both Fisher's exact test and regression models. The choice of the tool depends on the availability of replicates, so regression-based tools are a good option for data with replicates, while Fisher's test-based ones should be preferred for data without replicates (Wreczycka et al., 2017). The analysis workflow of bisulfite sequencing methylation data is shown on Figure 11 A. Steps for microarray data analysis are as follows: QC, filtering by different parameters (single nucleotide polymorphism (SNP) associated, from X and Y chromosomes, multi-hits, etc.), data normalization, covariate analysis, exploratory analysis, batch effect correction and differential methylation analysis (Figure 11 B), where most of these steps include visualization to help data interpretation (Assenov et al., 2014; Bock, 2012; Morris et al., 2014).

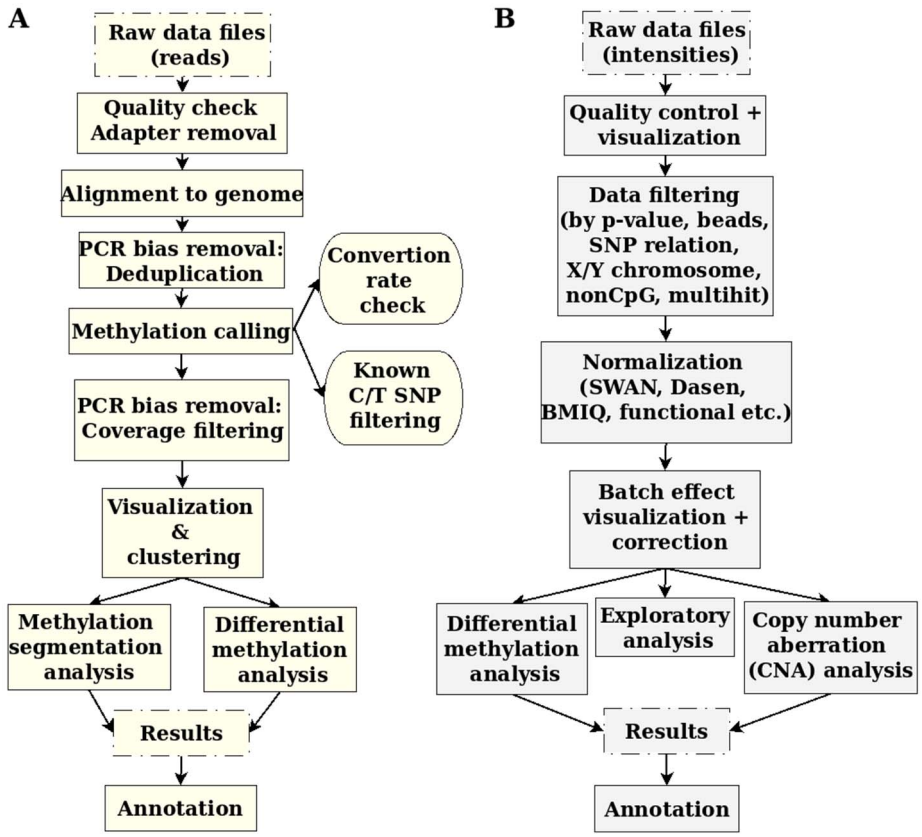

Figure 11. **A.** Analysis workflow of bisulfite sequenced DNA methylation data. The first step includes base quality check, adapter removal, and read end trimming. Next is alignment to the genome with Bismark or some similar tool. Third – the alignment quality check includes removing PCR bias (could be partly fixed by removing multi-mapped reads – deduplication and coverage filtering) and methylation calling (advised checking of conversion rates and filtering of known SNPs). These steps are followed by visualization, analysis, annotation, and interpretation of the results. Scheme based on Wreczycka *et al.* (Wreczycka et al., 2017). **B.** Analysis workflow of methylation array based on InfiniumHumanMethylation450K array example. The filtering and QC steps may be shifted. The QC step helps to distinguish outlier probes, for example with inefficient bisulfite conversion. The filtering step removes CpGs that could affect normalization and further analysis, such as sites located on sex chromosomes. The normalization is necessary to decrease technical noise and eliminate biases. Next is batch effect correction. Different analyses could be done after data preprocessing steps, where exploratory and differential methylation analyses are the most popular analyses. Scheme based on ChAMP (Morris et al., 2014) and RnBeads (Assenov et al., 2014) manuals.

QC of bisulfite array methylation data usually addresses bisulfite conversion efficiency, unspecific probe hybridization, sample sex prediction and in some cases sample mix-ups. The filtering step includes parameters used for probe filtering such as detection P-value, known SNP overlapping probes, cross-hybridization probes, probes with missing values, probes on sex chromosomes, non-CpG probes, sometimes user-defined criteria, and may be done before and/or after normalization, for example in RnBeads filtering is done before and after data normalization. (Assenov et al., 2014; Fortin et al., 2014, 2016; Morris et al., 2014). The array data normalization is necessary to eliminate biases and decrease the level of technical noise. Several within-array (SWAN (Maksimovic et al., 2012), BMIQ (Teschendorff et al., 2013), NOOB (Davis et al., 2019), PBC (Dedeurwaerder et al., 2011), etc.) and between-array (Funnorm (Fortin et al., 2014), Dasen (Pidsley et al., 2013), SQN (Aryee et al., 2014; Touleimat and Tost, 2012)) normalization methods are available (Assenov et al., 2014; Fortin et al., 2014; Liu and Siegmund, 2016; Morris et al., 2014; Wang et al., 2015). There is no golden normalization standard/method for methylation data, but several works compared Infinium 450K array within-array normalization methods and agreed that BMIQ is in general good for Infinium I/II-type probe bias correction (Dedeurwaerder et al., 2014; Liu and Siegmund, 2016; Wang et al., 2018), also there were proposed combinations of several methods such as NOOB+BMIQ (Liu and Siegmund, 2016).

The normalized data could be analyzed in many ways, where the most popular ones are exploratory and differential methylation analyses. The exploratory analysis gives an overview of normalized data including batch effects through principal component analysis, sample clustering and similarity, and global distribution of methylation levels, while differential methylation analysis gives a deeper insight of methylation levels between sample groups (Assenov et al., 2014; Bock, 2012). The methylation is usually analyzed on a single CpG or region level (DMR – differentially methylated region) (Assenov et al., 2014; Bibikova et al., 2011; Bock, 2012; Wreczycka et al., 2017). Single CpGs could be clustered according to region length, or gene functionality regions (gene body – gene coding region excluding 1st exon; 1st Exon; TSS500/TSS200 – 500/200 kb upstream transcription start site; 3' untranslated region (UTR) and 5'UTR), or considering their relation to CpG islands (CpG island (CGI) – a region with high CpG site frequency; Shore – region flanking CGI from the north (N) and south (S); Shelf – 2 kb regions directly adjacent to shores from N and S; and Open Sea – regions more than 4kb away from CGIs) (Figure 12) (Assenov et al., 2014; Bibikova et al., 2011; Wreczycka et al., 2017). DMRs could be predefined through relation to the island or gene functionality; or through user-defined criteria like fixed tiling window length, or fixed number of CpGs per region, or smoothed estimated effect sizes (Wreczycka et al., 2017). There are a lot of statistical models and tests available for DMR detection in array-based data. As was mentioned before, Fisher's exact test is used for samples without replicates, Wilcoxon signed-rank test is good for not normally distributed data, while the paired Student's t-test is used for paired normally distributed samples.

Various regression-based models (linear/logistic regression) and hidden Markov models (Bock, 2012; Wreczycka et al., 2017) are also used.

After obtaining differentially methylated CpGs or regions, these need to be annotated, and in the case of Illumina arrays one can use the Illumina provided annotation file, but also files from different databases like Ensembl, or custom annotation files (Assenov et al., 2014). The methylation analysis may include gene enrichment analysis, differential methylation interaction hotspots, copy number variation (CNV) analysis (Morris et al., 2014) and also could be combined with gene expression data to find novel expression regulatory mechanisms and provide evidence for the methylational findings.

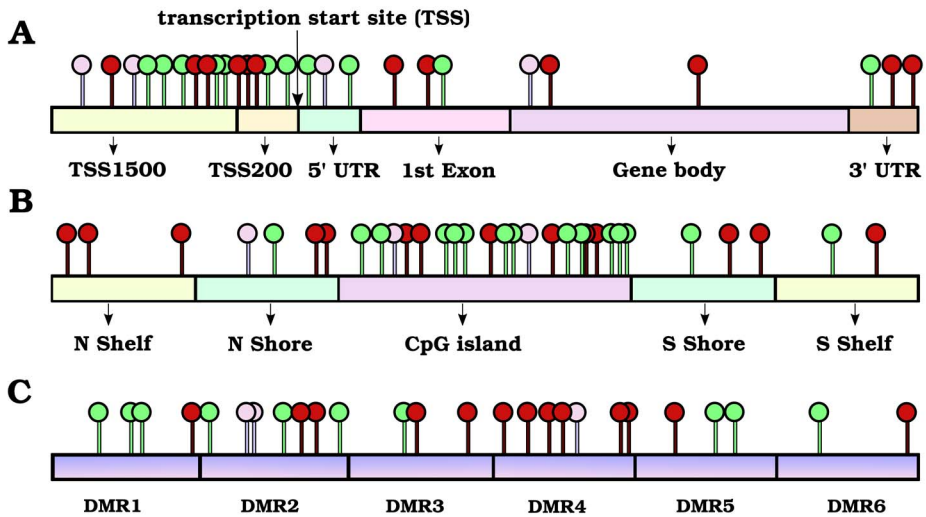

Figure 12. Grouping of CpGs. Red pins – methylated CpGs, green pins – non-methylated CpGs, and pink pins – hemimethylated CpGs. **A.** CpG clustering by gene functional regions. TSS – transcription start site, TSS200 – 200 bp upstream of TSS, TSS1500 – 1500 bp upstream of TSS. **B.** CpG clustering in relation to CpG island. **C.** CpG clustering to differentially methylated regions (DMRs) with fixed tiling window length. **A** and **B** are examples of pre-defined DMRs, while **C** is user-defined (criteria: fixed tiling window length, or fixed number of CpGs per region, or smoothed estimated effect sizes) DMRs. Figure based on Bibikova *et al.* (Bibikova et al., 2011) and Wreczycka *et al.* (Wreczycka et al., 2017).

1.7.3. Factors affecting DNA methylation analysis

In general, factors affecting DNA methylation analysis are similar to factors affecting RNA analysis, such as sample donor's lifestyle and living environment (even maternal environment during pregnancy (Dias and Ressler, 2014; Franklin et al., 2010; Heijmans et al., 2008)); sample collection, storage and preparation, data preprocessing and analysis methods (Perrino et al., 2017). It has been shown that methylation is age-related (Bell et al., 2012; Zhu et al., 2018), thus the donor's age could influence analysis results if not taken into account. Even though DNA is more stable than RNA, it also degrades in time. In the preparation step, the bisulfite conversion could be incomplete, leading to false analysis results (Assenov et al., 2014; Kurdyukov and Bullock, 2016). Also, bisulfite conversion is incapable of discriminating between 5-hydroxymethylcytosine (indicates the active demethylation process mediated by TET enzyme) and 5-methylcytosine, which can raise issues for some DNA regions (Figure 13) (Beck, 2010; Bock, 2012; Kurdyukov and Bullock, 2016; Wreczycka et al., 2017). In the case of sequencing, bisulfite conversion makes amplification of long fragments difficult due to DNA fragmentation and makes the alignment step complicated due to reduced genome complexity. The bisulfite conversion could be avoided using direct methylation detection through the sequencing of unmodified DNA, for example with single-molecule real-time sequencing (Kurdyukov and Bullock, 2016).

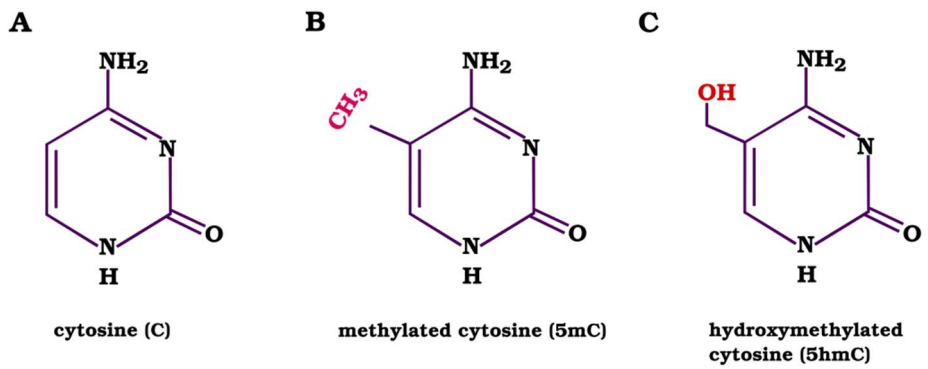

Figure 13. A. Unmethylated cytosine (C), **B.** 5-methylated cytosine (5mC), **C.** 5-hydroxymethylated cytosine (5hmC). Unmethylated cytosine is converted with bisulfite treatment to uracil, which is later converted to thymine. While 5-methylated and 5-hydroxymethylated cytosines are left intact. The distinction between 5mC and 5hmC is necessary, as the 5hmC indicates an active demethylation process, and thus 5hmC is later converted to C.

There are array and sequencing platforms specific issues, that are valid for all analyses regardless of 'omics' data type. Some of the issues are: unspecific probe hybridization could occur using arrays (Bock, 2012; Wreczycka et al., 2017), or lane batch effect using sequencing. Analyses with big sample sets are prone to sample mix-ups, which could be solved using genotyping data or eliminated via exclusion of dubious samples during data preprocessing steps (Assenov et al., 2014; Bock, 2012). Both array and sequencing datasets have various batch effects, that are addressed during QC, normalization, and filtering steps (Assenov et al., 2014; Bock, 2012; Wreczycka et al., 2017). A broad choice of methods is available for data normalization, especially for microarrays (Assenov et al., 2014; Morris et al., 2014; Wang et al., 2015), where one method could behave better than others depending on data. As there is no golden method and comparative works often report contradictory results (Dedeurwaerder et al., 2014; Fortin et al., 2014; Liu and Siegmund, 2016; Wang et al., 2015, 2018), the user should try different methods to see which one fits the best. Similarly, the range of models is available for differential analysis, making the analysis slightly more complicated (Bock, 2012; Wreczycka et al., 2017). Like in RNA analysis, the tissue cell-types and biopsy cellular composition also affect the results. Variations in methylation levels of diverse cell types could cause a dilution effect and influence detected methylation levels (Houseman et al., 2012; Kurdyukov and Bullock, 2016; Perrino et al., 2017; Teschendorff and Zheng, 2017). Fortunately, this issue could be solved, using deconvolution techniques similar to those described for transcriptome analysis (Houseman et al., 2012; Teschendorff and Zheng, 2017).

1.8. Role of DNA methylation in endometrial receptivity context

The majority of methylation research is made in the cancer field, but some studies have also been carried out for studying the female reproductive health. As arrays are the most popular platform for human methylation studies due to their good price/data quantity ratio, the majority of methylation studies in the reproduction field are array-based (Zhang et al., 2014a). For example, Illumina HumanMethylation450K bead chip was used in Saare *et al*. and Maekawa *et al*. studies (Maekawa et al., 2019; Saare et al., 2016), while Houshdaran preferred the IlluminaHumanMethylation27K array (Houshdaran et al., 2014, 2016). The Maekawa and colleagues' study is interesting, as they used three different methods at once: ChIP-seq (immuno-precipitation) to look into histone modifications, 450K bead chip to compare methylation patterns between cell types and bisulfite sequencing to validate methylation patterns of genes of interest (Maekawa et al., 2019). Some earlier studies of endometrial cell cultures used immunohistochemistry to study DNA methyltransferases (Gao et al., 2012; Yamagata et al., 2009) and MeDIP-seq to study endometrial carcinoma types (Zhang et al., 2014a). Thus, the choice of methodology depends on the study

question, where arrays are used to get an overall picture of whole-genome methylation profiles and their differences between physiological states, while other methods help to get a deeper insight of other methylation aspects.

Although the majority of studies have focused on the role of methylation in the context of pathology (Fan et al., 2018; Jiang et al., 2017; Zhang et al., 2014b), it has become clear that methylation is also involved in regulating the normal physiological processes related to female fertility (Gao et al., 2012; Koukoura et al., 2016; Maekawa et al., 2019; Mortlock et al., 2019; Saare et al., 2016; Schroeder et al., 2013; Yamagata et al., 2009). The fact that the menstrual cycle affects endometrial methylation (Houshdaran et al., 2016; Saare et al., 2016) implies that it may also play a role in endometrial receptivity. Indeed, it has been shown that the expression of three methyltransferases (*DNMT1*, *DNMT3a*, and *DNMT3b*) changes depending on the menstrual cycle phase (proliferative, early-secretory and mid-secretory) (Yamagata et al., 2009). In concordance with Yamagata, Gao et al. showed on mice model, that *DNMT1* and *DNMT3a* take part in the regulation of uterine methylation during the peri-implantation period and DNA methylation plays an important role in decidualization of stromal cells in the implantation site during early pregnancy (Gao et al., 2012) although recent studies have begun to question these results (Maekawa et al., 2019).

The overlap between the results of endometrial methylation studies is small. This could be due to differences in experimental, technological and physiological variability, choice of data analysis methods, and/or other factors. Recent studies focus more on tissue cellular composition and tissue-type-specific methylation patterns. Rahmioglu and colleagues looked into tissue-derived and experimental variability of methylation and expression profiles and showed that methylation profiles could be influenced by age, menstrual cycle phase, smoking and tissue (sub-)type (Rahmioglu et al., 2017). Mortlock *et al.*, while studying methylation changes in endometrium and blood during the menstrual cycle, observed methylation differences only in the endometrium, underlining the importance of using process-/disease-relevant tissues (Mortlock et al., 2019). In a review article Saare and colleagues speculated about endometrial and endometrioma tissue heterogeneity, pointing out that there are many cell types with cell-specific methylation profiles, that can influence analysis results (Saare et al., 2018). Besides, the biopsy does not only contain different cell types, but the proportions of cells also vary from biopsy to biopsy, resulting in a bias towards profiles of cells with larger proportions (Saare et al., 2018).

From the studies mentioned above, it is clear that methylation plays its role in endometrial receptivity and female fertility in general. As the endometrium is so dynamic and complex a tissue, and as methylation patterns are tissue- or even cell-type-specific, new approaches for fertility studies, that could manage dissection of whole tissue into cell types and at the same time would not lose the sight of interactions between various cells, are needed. The methylation alone is only a piece of the puzzle and is incapable of showing the whole

picture, but when put together with the other pieces, it could bring us a step closer to the solution of endometrial receptivity conundrum.

1.9. Integration of 'omics' layers

Living organisms are very complex structures, where everything is connected, and every process has an impact on other processes. Some of these effects are slight and unnoticeable on the whole system level, while others are so drastic that they could lead to disease or even death. The basic scheme of DNA becoming a protein usually goes along the lines of DNA being transcribed to RNA and RNA translated to protein. But actual processes are a lot more complicated. For instance, if we zoom in on RNA transcription, RNA splicing for production of mature coding RNA comes into play, and to make things even more complicated, alternative splicing can result in several mRNAs with different functions, which are all transcribed from the same DNA sequence. And then there are non-coding RNAs with a variety of functions, playing by their own rules, regulating mRNA transcription and degradation, while at the same time influencing each other. Similar multi-layered regulation can also be observed on DNA and protein level

The easiest way to solve complicated problems is to dissect them into parts, solve these parts separately and then put everything together. In the context of the current thesis, the 'parts' in question are different 'omics' data layers. Admittedly, it would be an overstatement to say that we have completely 'solved' the biological questions on single 'omics' levels, but we do have enough insight to start putting them together to get an overview of the whole functionality of an organism, as all processes occurring in the organism are connected and depend on each other (Das, 2014; Perrino et al., 2017; Xu et al., 2017). To avoid biases and other problems, it is better to obtain different data layers from the same tissue, organ or organism, if possible. Putting 'omics' data layers together helps track connections between them and get a bigger picture of processes of interest (Perrino et al., 2017; Xu et al., 2017). For example, DNA methylation is usually correlated with gene expression, as this helps to get a better understanding of differences in methylation, that are actually changing biological pathways and have an impact on current processes that take place in the tissue, organ or organism (Das, 2014; Perrino et al., 2017). Also correlating gene methylation with transcription level helps to identify novel transcriptional regulatory mechanisms (Das, 2014; Wan et al., 2015). The same is valid for the combination of several transcriptomic layers as coding and non-coding RNAs (Perrino et al., 2017). At the same time, these correlations provide additional findings, which cannot be obtained from studying gene expression and methylation patterns separately (Perrino et al., 2017; Wan et al., 2015; Xu et al., 2017). The same rules are also valid for genome mutations, that may affect downstream processes, and by adding transcriptome studies could help track down mutations with real impact (Xu et al., 2017). When the transcriptome is integrated with the proteome, it

points out how many and which mRNAs are actually translated to proteins (Perrino et al., 2017), because many mRNAs, especially with “broken code”, are degraded before translation. So, for better understanding of how complex phenotypes are affected by gene expression, there is a need for integration of other ‘omics’ layers such as genomics, epigenomics, proteomics and metabolomics (Perrino et al., 2017; Xu et al., 2017). Still, the integration of more than two layers simultaneously is challenging with current bioinformatics tools (Xu et al., 2017). In recent years, several works have integrated different data layers in the context of female fertility studies, combining genome methylation with gene expression (Houshdaran et al., 2014; Maekawa et al., 2019; Rahmioglu et al., 2017), coding RNAs with non-coding RNAs (Furlong et al., 2018; Sigurgeirsson et al., 2016), or genotype with methylation (Mortlock et al., 2019). Hopefully, these studies mark the beginning of a new and more elaborate approach to studying female reproductive health.

2. AIMS OF THE STUDY

This study aimed to gain a better insight into endometrial biology in transition between early-secretory and mid-secretory phases, coinciding with the window of implantation, by combining an endometrial gene expression dataset with additional 'omics' data layers (endometrial cell population-specific transcriptome datasets, DNA methylation, and non-coding RNA (miRNA) transcription datasets).

The specific objectives of this thesis were as follows:

1. To evaluate the effects of whole tissue computational deconvolution on differential gene expression analysis outcome on the example of human endometrium.
2. To explore the correlation between miRNAs and gene expression datasets to evaluate interactions of potential endometrial receptivity mRNA marker genes and their associated miRNAs that were predicted based on three different databases.
3. To observe genome-wide changes in methylation during the two studied menstrual cycle phases and assess the correlation between methylation and gene expression patterns.

3. RESULTS AND DISCUSSION

3.1. Effect of cellular composition on endometrial gene expression

The endometrium is a very dynamic and heterogeneous tissue mainly studied using biopsies or *in vitro* cultured cells. While the biopsy studies have not taken into account cellular composition (possible reason of moderate overlap between the previous studies), the cell culture studies lack a more general view on tissue complexity (Hayman et al., 2006; Strell et al., 2019). Thus, an alternative is to use whole tissue with cellular composition adjustment (Shen-Orr and Gaujoux, 2013). Computational deconvolution is a good method to account for the cellular composition of biopsies, that has been previously used for gene expression profiling in blood, tumors, and lymphoid tissues (Chen et al., 2017b; Li et al., 2017; Pan et al., 2017; Schelker et al., 2017). However, such an approach has never been used for analyzing the endometrial transcriptome.

The goals of Ref. I were to use computational deconvolution to adjust for the effect of cellular composition on transcriptomic profile of endometrial biopsies, to compare gene expression profiles from analyses with and without cellular composition adjustment, and to point out highly potential receptivity biomarker candidate genes.

3.1.1. Description of cohort, materials, and methods

The study was approved by the Research Ethics Committee of the University of Tartu, Estonia (No 221/M-31) and Ethical Clinical Research Committee of IVI Clinic, Valencia, Spain (No 1201-C-094-CS). Informed consent was signed by all women who entered the study.

The biopsies were obtained for early-secretory (ES or LH+2) and mid-secretory (MS or LH+8) endometrium from fertile women, with at least one child, regular menstrual cycle (25–35 days) and normal body mass index (24.1±4.8 kg/m^2). 20 participants were from Estonia and 15 from Spain, thus 35 women and 70 samples in total. Samples were sequenced in 3 different sequencing facilities: 20 at Estonian Genome Center Core Facility (EGCUT, Tartu, Estonia), 20 at the Institute for Molecular Medicine Finland (FIMM, Helsinki, Finland) and 30 at Lifesequencing S.L. (Valencia, Spain). The paired-end RNA reads were trimmed and adapters removed with Trimmomatic-0.32 (Bolger et al., 2014), quality filtered with FASTQ quality filter tool from FASTX-Toolkit v.0.0.14 and mapped using TopHat2 (Kim et al., 2013) to human genome version 19 (hg19) (Editorial, 2010; Guo et al., 2017). The read counts per transcript were obtained with the HTSeq-count script (Anders et al., 2015), using Ensembl human release 75 (Zerbino et al., 2018) annotation file. The read counts were used, as they behave better than FPKMs and are sufficient for differential expression analysis. Further differential analysis was performed for

each sample group separately, where samples were grouped according to sequencing center, and in two separate ways: with and without intermediate deconvolution step and cell proportion adjustment (Figure 14).

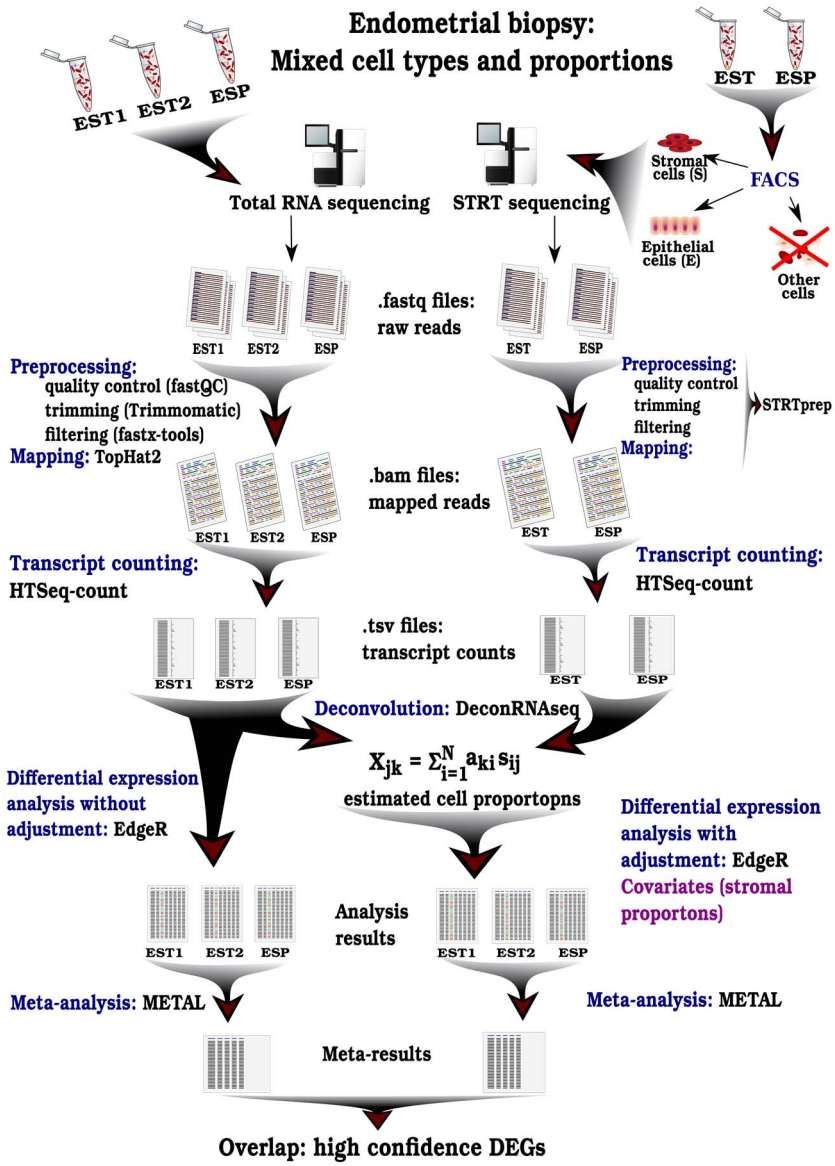

Figure 14. Workflow from biopsy to significantly differentially expressed genes (DEGs). Endometrial biopsies were obtained from Estonian and Spanish women, and total RNA of whole tissues was sequenced in three different facilities, hence the grouping according to origin and sequencing facility: Estonian (EST1), Finnish (EST2) and Spanish (ESP). The single-cell samples were STRT sequenced after FACS in the

same facility, hence grouping according to the origin only (Estonian: EST and Spanish: ESP) was conducted. The whole-tissue reads were preprocessed and mapped using different tools: FastQC (Andrews, 2010), Trimmomatic (Bolger et al., 2014), FASTX-tools (Hannon lab, 2009), and TopHat2 (Kim et al., 2013) separately, and single-cell reads were pre-processed and mapped using STRTprep pipeline (Krjutškov et al., 2016). In both cases human genome version 19 (Editorial, 2010; Guo et al., 2017) was used for mapping. Transcript counts were obtained with HTSeq-count (Anders et al., 2015) script using Ensembl human annotation file release 75 (Zerbino et al., 2018) for both whole-tissue and single-cell reads. The single-cell and whole-tissue count matrices were used to computationally estimate cell proportions of whole-tissue samples using DeconRNASeq software (Gong and Szustakowski, 2013). DeconRNASeq estimates the cell-type proportions over samples by fitting a non-negative least squares equation for each transcript. Further, it calculates the theoretical proportions of epithelial and stromal cells across the whole transcriptome. The expression level X_{jk} of gene j in a sample k is the average of expected expression level s (stromal) across the cell types s_{ij}, weighted by the respective cell-type proportions a_{ki} ($i = 1 \ldots N$, N: the total number of cell types) (Adapted from: (Gong and Szustakowski, 2013)). Then the two types of differential expression (DE) analysis using EdgeR (Robinson et al., 2010) software were performed per each whole-tissue sample group: with cell proportion adjustment and without cell proportion adjustment. Average stroma proportions, calculated from all whole-tissue samples, were applied as covariates in DE analysis with cell proportion adjustment. The DE results for mid-secretory (receptive) endometrium of groups were meta-analyzed according to adjusted and non-adjusted analysis type with a METAL tool (Willer et al., 2010). Meta-analyzed results were compared and genes with the same direction of regulation in all groups and with Bonferroni p-value $< 2.79 \times 10^{-6}$ that overlapped between two meta-analyses were indicated as high confidence DEGs.

For the deconvolution step, the R package DeconRNASeq v.1.10.0 (Gong and Szustakowski, 2013) was used. There are plenty of deconvolution packages for blood, that usually already contain reference genes for blood cells, but there is no such package available for endometrium. DeconRNASeq lacks any cell-type references, thus works on many tissues, also it is free, locally installable and easy to use, therefore well suited for the task at hand. To estimate cell proportions in the whole tissue, the package requires whole tissue and single-cell transcription levels – in this case, CPMs (counts per million reads) were used. For cell-type-specific expression profiles for endometrial stromal and epithelial cells, we used data from the SARM project (Altmäe et al., 2017; Teder et al., 2018), which included endometrial samples from Estonia and Spain for similar time-points (early secretory and mid-secretory). These cell populations had been sorted by FACS. For analysis, we used the STRTprep pipeline (Krjutškov et al., 2016), up to the transcript counting step, where counting was made with HTSeq-count script using Ensembl human release 75. Separate counting was done, as there was another annotation file used in the STRTprep pipeline and it is important to identically annotate both datasets. The DeconRNASeq requires one expression profile per cell type, but the SARM dataset provided us with 19

stromal and 17 epithelial samples for the ES phase and 24 samples per cell type for the MS phase that matched with bulk-tissue samples. For the elimination of interindividual differences, the median expression levels per cell type were used for each secretory phase for downstream deconvolution.

To evaluate the accuracy of computationally estimated proportions in endometrial biopsies, we also analyzed the histological images of 9 samples per each cycle day (LH+2/LH+8), estimating proportions, which the two cell types occupied in each histology image, and calculating the Pearson correlation between the two calculated proportion sets.

The differential analysis results were further meta-analyzed together using the sample size weighted method (based on Stouffer's Z-score) implemented into the METAL tool (Willer et al., 2010). METAL is flexible and can be easily adapted for gene expression studies (Peters et al., 2015) despite its general purpose – GWAS (genome-wide association study) meta-analysis. Genes that overlapped between two meta-analyses were considered as most likely potential receptivity biomarker candidates and were compared with already known gene lists from previous studies, for example, the ERA genes. Some of these genes were also validated by M. Suhorutsenko with quantitative real-time PCR.

3.1.2. Differentially expressed genes with and without cell proportion adjustment

The number of expressed transcripts varied between 22,000 and 30,500 per sample. As the PCA plot (Figure 15) showed clustering according to the sequencing facility, we carried out three differential expression analyses with and without cell proportion adjustment for each sample group according to the sequencing center, followed by meta-analysis. This approach was used to retain as many samples as possible in the analysis and at the same time avoid false results caused by batch effects. The meta-analysis of differential expression analysis results provided 3,591 (1,800 up-regulated and 1,791 down-regulated) significantly deferentially expressed transcripts in MSE without cell-fraction adjustment and 1,211 (679 up-regulated and 532 down-regulated) with cell fraction adjustment.

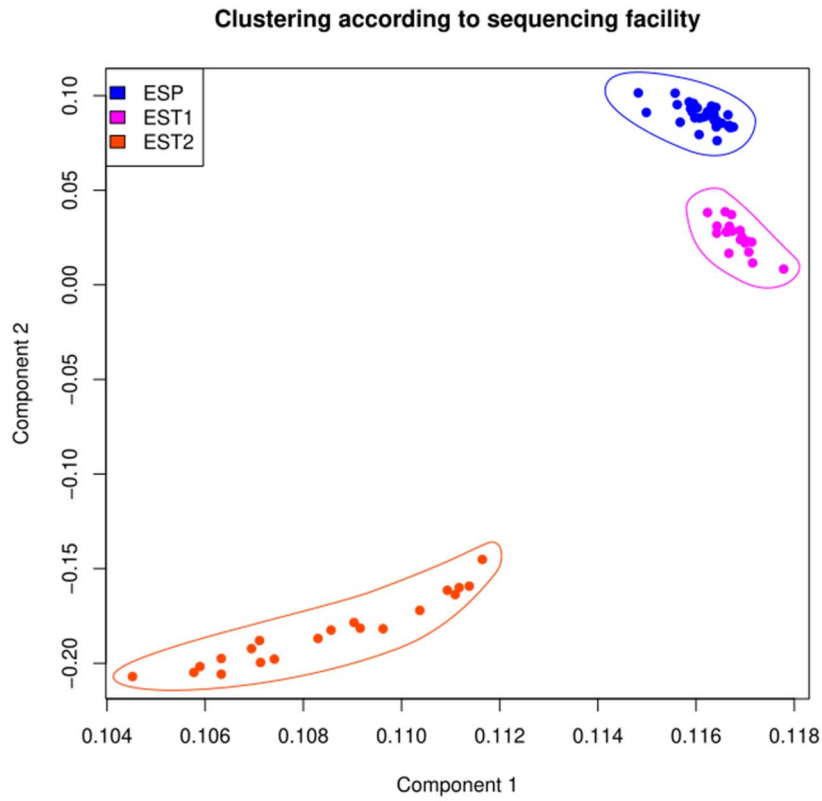

Figure 15. The plot shows sample clustering according to the sequencing facility: EGCUT (EST1, pink), FIMM (EST2, orange) and Spain (ESP, blue). Unpublished data.

The overlap between the two meta-analyses resulted in 573 up-regulated and 373 down-regulated transcripts (Bonferroni p-value < 2.79×10^{-6}) (Figure 16). The genes that lost their significance after cell-fraction adjustment are probably reflecting the differences in cellular composition, but not the differences in expression. From the other side, the genes that gained their significance after cell-fraction adjustment were masked by the difference in cell proportions that in the end summed up to similar expression between ESE and MSE in the analysis without cell-type proportion adjustment. We concentrated our attention on the overlap between two meta-analyses to eliminate possible false-positive genes and to be extra confident in our results. The genes common for the two analyses and previously reported in other receptivity studies were considered as highly potential receptivity biomarker candidates, in total 171 genes. Interestingly, out of the 238 ERA test genes (Díaz-Gimeno et al., 2011), 112 do not show significant expression differences after cell fraction adjustment, thus indicating that these markers could reflect the uneven cellular content of the samples or the menstrual-cycle-specific physiological changes in the tissue cellular composition.

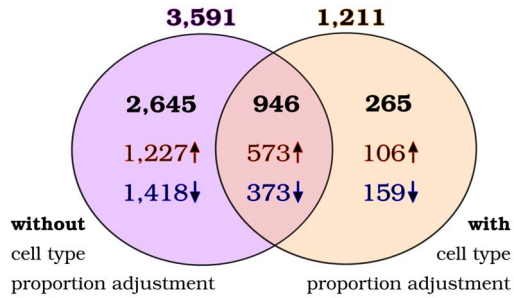

Figure 16. Venn diagram shows the number of significant DEGs after meta-analysis. The violet circle indicates results **without** and yellow circle indicates results **with** cell proportion adjustment, respectively. Numbers outside circles indicate the total number of DEGs per analysis. Bold numbers in circles indicate unique (left and right) DEGs per analysis and overlapping (middle) between analysis. Arrows indicate up-regulated (↑) and down-regulated (↓) DEGs. Adapted from Ref. I.

As an example, the top 20, by average logFC between two meta-analyses, significantly (Bonferroni p-value < 2.79×10^{-6}) up- and down-regulated genes are shown in Table 1. The 7 out of 10 positively (*PAEP, C4BPA, CXCL14, SLC15A1, AOX1*, and *TSPAN8*) and 3 out of 10 negatively (*GREM2, MMP26,* and *SLC15A2*) expressed genes are present in the ERA test and are well known endometrial receptivity genes. The *PAK7(PAK5), SLC26A4-AS1, SULT1C2P, PKHD1L1, EPPIN, ADCYAP1R1, GJB6, HAP1, C2CD4A*, and *IRX3* are missing from the ERA list. These genes are associated with signaling pathways (*PAK7, HAP1*) (Cotteret et al., 2003; Shimojo, 2008), responses to different stimuli (*ADCYAP1R1, GJB6*) (UniProtKB, 2020a, 2020b), apoptosis regulation (*PAK7, GJB6, ADCYAP1R1*) (Cotteret et al., 2003; UniProtKB, 2020a, 2020b), immune responses (*PKHD1L1, EPPIN, HAP1*) (Hogan et al., 2003; Shimojo, 2008; UniProtKB, 2020c), cell processes such as proliferation, differentiation, growth, migration and communication (*PAK7, ADCYAP1R1, GJB6, IRX3*) (Cotteret et al., 2003; Scarlett et al., 2015; UniProtKB, 2020a, 2020b), and vesicle function or transport (*C2CD4A, HAP1*) (Shimojo, 2008; Warton et al., 2004). The *SLC26A4-AS1* is antisense RNA of the *SLC26A4* gene, and *SULT1C2P* is a pseudogene. The processes associated with the genes mentioned above also take place during the transition from ESE to MSE phases and embryo implantation, indicating the high biomarker potential of these genes. No or slight changes in logFC after cellular adjustment could indicate that changes in gene expression were not influenced by cellular composition (Table 1), while significant changes indicate that gene expression levels were influenced by cellular composition.

Table 1. Top 20 (by average LogFC) significantly (Bonferroni p-value < $2.79x10^{-6}$) up- and down-regulated genes overlapping between two meta-analyses, where the first 10 genes with negative average logFC are separated with a double line from the 10 genes with positive average logFC. "Y" indicates for yes, "N" indicates for no, "average LogFC" shows average log fold change between two analyses. Unpublished table.

Gene	P-value without cellular adjustment	P-value with cellular adjustment	LogFC without cellular adjustment	LogFC with cellular adjustment	Average LogFC	Present in ERA test
PAK7(PAK5)	3.9e-121	5.789e-21	-5.1713	-4.9104	-5.0409	N
SLC26A4-AS1	5.75e-91	6.681e-14	-5.0651	-4.1699	-4.6175	N
SULT1C2P1	3.256e-35	4.707e-17	-3.6112	-5.0962	-4.3537	N
GREM2	1.55e-92	5.097e-26	-3.9054	-4.7435	-4.3244	Y
PKHD1L1	3.53e-35	8.002e-17	-3.4298	-4.8355	-4.1327	N
EPPIN	1.41e-39	2.393e-17	-3.5919	-4.5516	-4.0718	N
ADCYAP1R1	2.377e-75	1.381e-12	-4.2652	-3.3092	-3.7872	N
MMP26	1.013e-07	3.765e-10	-2.1720	-5.3495	-3.7607	Y
GJB6	4.418e-28	5.388e-17	-2.8037	-4.5132	-3.6585	N
SLC15A2	8.857e-56	5.832e-21	-3.1632	-4.1328	-3.6480	Y
PAEP	2.44e-227	1.078e-31	8.8231	7.8935	8.3583	Y
C4BPA	3.02e-161	7.651e-30	7.6210	6.2626	6.9418	Y
GPX3	2.24e-184	7.726e-38	7.0251	6.4316	6.7284	Y
CXCL14	3.5e-93	1.801e-18	7.0401	6.2944	6.6673	Y
HAP1	2.959e-72	4.854e-17	6.3741	6.6619	6.5180	N
SLC15A1	6e-180	3.676e-48	6.8048	6.0657	6.4353	Y
C2CD4A	3.82e-160	2.559e-30	7.5038	5.3121	6.4080	N
IRX3	4.51e-125	1.445e-25	6.0159	5.3905	5.7032	N
AOX1	1.54e-177	2.478e-38	5.3500	4.9856	5.1678	Y
TSPAN8	4.88e-89	1.232e-15	5.5485	4.7050	5.1267	Y

3.1.3. Validation of cell proportions

To understand the reliability level of computational deconvolution, additional evaluation of cell fraction using histological methods was made. The epithelial fraction estimates from 18 paired samples (9 ESE and 9 MSE) with detailed cycle day information were correlated by my colleague and co-author T. Laisk

with epithelial fractions obtained with the deconvolution method. The deconvoluted fractions were also compared with stromal and epithelial proportions from histological images. The comparison was made for the eight ESE and five MSE samples (for some individuals, more than one image was available for evaluation), where proportions were calculated by measuring cell-type occupied areas in histological images, and then the cell type area in pixels was divided by whole pixel count. The histological day of ESE and MSE samples correlated well ($r = 0.88$ (95% CI 0.71–0.96), $P = 1.1 \times 10^{-6}$) with epithelial cells' proportion from computational deconvolution (ESE ~35% and MSE ~54%), and thus, confirmed epithelial cell fraction dominance in MSE, where, on average, >50% of epithelial cells were seen by the time of WOI in histology preparations (Figure 17). These results show that computational deconvolution is a reliable method for stromal and epithelial (the most represented cell types in endometrium) cell fraction estimation in whole tissue samples, but it still needs confirmation for less represented, but not less important endometrial cell types.

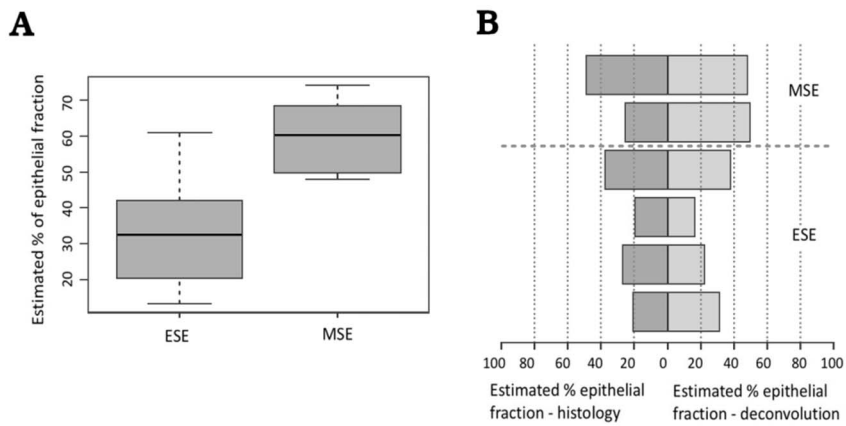

Figure 17. Histological and deconvoluted epithelial cells' fractions. **A.** Proportions of epithelial cells' fractions estimated by computational deconvolution (ESE – early-secretory and MSE – mid-secretory endometrium). **B.** Epithelial fractions estimated by histological evaluation (dark-gray) and computational deconvolution (light gray). In five out of six samples the estimated fractions within samples showed a similarity between the two methods. Adapted from Ref. I.

Overall this study gives an overview of changes in the global transcriptome between early-secretory and mid-secretory phases in epithelial and stromal cell proportions in endometrium and points out the effects of cellular composition on analysis outcome, thus also emphasizes the importance of cellular adjustment of whole tissue studies during differential gene expression analysis. This work is a good example of expression meta-analysis in the case of split datasets and it also shows how a combination of bulk-tissue and single-cell RNA-seq datasets could improve analysis results.

3.2. Correlation between potential endometrial receptivity marker genes and their associated miRNAs

As mentioned earlier, the overlap between the results of many endometrial receptivity transcriptome studies has been modest for different reasons. To compile a list of genes that show most robust results across various studies, Altmäe *et al.* undertook a meta-analysis of previous studies, using the robust rank aggregation method (RRA) (Kolde et al., 2012), and as a result, established a signature of highly putative endometrial receptivity biomarkers. As a part of this study, we also aimed to gain insight into the expression regulation of these genes by analyzing their possible regulatory miRNAs *in silico* and validating the results in two independent datasets.

3.2.1. Description of cohort, materials, and methods

The lists of marker genes and their associated miRNAs were obtained according to methods in the Ref. II. The validation of *in silico* found marker genes was performed on two datasets – the whole endometrial tissue biopsies obtained from 20 healthy fertile women (NOTED project) and endometrial cell-type-specific data from 16 healthy fertile women (SARM project). Both datasets had paired samples for ES (LH+2) and MS (LH+8) endometrium. For whole tissue the miRNA and mRNA data were available. Whole tissue miRNA raw reads were trimmed and adapters removed with Trimmomatic-0.32 (Bolger et al., 2014), while quality filtering was prepared with FASTX-Toolkit v.0.0.13 (Hannon lab, 2009). Then the reads were mapped to hg19 (Editorial, 2010; Guo et al., 2017), annotated and counted based on miRBase v.21 (Kozomara and Griffiths-Jones, 2014; Kozomara et al., 2019) annotation files with miRDeep2.0.0.5 tool (Friedländer et al., 2012). The whole tissue mRNA was preprocessed using Trimmomatic, FASTX-tool, and FastQC (quality control) (Andrews, 2010) and mapped with TopHat2 to hg19. While the cell-specific mRNA was preprocessed and mapped to hg19 with STRTprep pipeline v.3.0.0 (Krjutškov et al., 2016). Both the whole tissue and cell-type-specific dataset transcripts were counted with HTSeq-count based on Ensembl human release 72 (Zerbino et al., 2018). The edgeR (Robinson et al., 2010) software was used to obtain differentially expressed genes and miRNAs between two groups. The *DDX52, DYNLT3, C1R* and A*POD* gene expression was also validated with qPCR method.

3.2.2. Potential receptivity marker genes and their associated regulating miRNAs

The 52 up-regulated and 5 down-regulated significant meta-signature genes were obtained with RRA analysis from nine published studies. The enrichment analysis revealed that a significant proportion of these genes is involved in responding to external stimuli, wounding, and different immune responses, while some proportion of genes was connected with exosomes and extracellular compartment. This is understandable since the balance between maternal immune tolerance and anti-infectious protective mechanisms should be established in the receptive uterus to provide an implantation-favorable environment for the embryo (Giudice, 2004; Haller-Kikkatalo et al., 2014).

The validation on 20 whole-tissue (endometrium) and 32 cell-type-specific (16 epithelial and 16 stromal) samples resulted in 52 out of 57 (48 up-regulated and 4 down-regulated) differentially expressed meta-signature genes for whole-tissue samples and 43 out of 57 (39 up-regulated and 4 down-regulated) genes for cell-type-specific samples. 16 genes showed epithelial-specific up-regulation, while four genes were up-regulated and one down-regulated only in stromal cells. The overlap between the two validation sets counted 35 up- and 4 down-regulated genes.

The involvement of miRNAs has been previously shown for endometrial receptivity (Altmäe et al., 2014), implantation, and pregnancy development (Luense et al., 2009). The TargetScan (Agarwal et al., 2015), DIANA (Paraskevopoulou et al., 2013) and miRanda (Betel et al., 2010) used for miRNA target prediction resulted in 818 miRNAs and 1,403 potential unique binding sites for 43 meta-signature genes overlapping between all three algorithms. The potential binding sites were additionally filtered by overlap with AGO-CLIP datasets containing experimentally determined Argonaut protein binding sites. 395 sites showed overlap, indicating interactions between 30 meta-signature genes and 348 miRNAs. The expressions of these miRNAs were additionally checked in the miRNA-seq dataset from ESE and MSE biopsies, resulting in significantly correlated by logFC (ESE vs MSE) 17 miRNAs and their corresponding 9 genes (Figure 18). The correlations were also made for ESE and MSE phases separately, resulting in 6 genes and 29 their corresponding miRNAs for ESE, and 8 genes and 49 corresponding miRNAs for MSE (9 genes and 65 miRNAs in total) (Figure 19).

A most common scenario is miRNA down-regulation of gene expression (negative correlation), but figures 18 and 19 show that most of the miRNA-gene pairs are positively correlated, pointing to the up-regulation of gene expression. There are many possibilities of how this situation can occur: 1) miRNAs can up-regulate gene expression through targeting gene promoters or enhancers in cell nucleus (Catalanotto et al., 2016; Yang et al., 2016); 2) length of 3'UTRs of transcripts can vary depending on the cell cycle and embryonic development, and miRNA binding sites are omitted in shorter 3'UTRs, thus preventing miRNA binding (Chen et al., 2020; Yang et al., 2016); 3) lncRNAs could act as

miRNA sponges, competing with coding genes for miRNAs (Paraskevopoulou and Hatzigeorgiou, 2016; Yang et al., 2016); 4) RNA-binding proteins could bind to miRNA recognition sites, thus preventing miRNA binding (Vasudevan, 2012; Yang et al., 2016); 5) miRNAs can indirectly regulate gene expression through other genes (Vasudevan, 2012; Yang et al., 2016); 6) miRNAs could act as a decoy of repressive proteins (Vasudevan, 2012); and 7) there could be different binding-site affinity, where partial binding could up-regulate gene expression (Vasudevan, 2012). As this study is limited to only 57 genes, the processes listed above could be the possible reason why the gene expression is up-regulated. Still, it is important to keep in mind that correlation does not necessarily point to the causal relationship between two states and could be just a mathematical artifact instead. Thus correlation results should be interpreted cautiously.

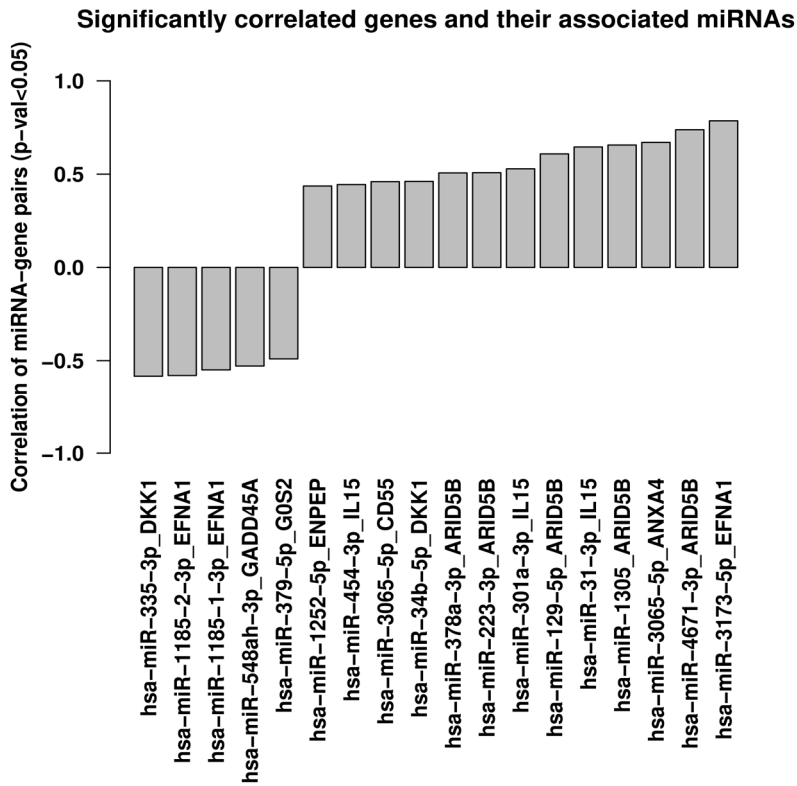

Figure 18. Significantly correlated (p-value < 0.05) expression differences (logFC) between two endometrial phases (ES and MS) of 9 meta-signature genes to expression differences (logFC) between ESE and MSE of 17 corresponding miRNAs. The x-axis shows miRNA-gene pairs and the y-axis shows correlation values. Unpublished data.

Figure 19. Significantly correlated (p-value < 0.05) expression (in CPMs) of 9 meta-signature genes to expression (in CPMs) of their associated miRNAs for ESE (LH2) and MSE (LH8) separately. X-axis shows miRNA-gene pairs and y-axis shows correlation values. Black names (above) match black columns and gray names (below) match gray columns. Unpublished data.

While taking into account only the genes overlapped between whole-tissue and cell-specific datasets, five genes and their corresponding 10 miRNAs remained important (Figure 20). These genes are *EFNA1* (ephrin-A1), receptivity marker that has a possible role in embryo-maternal communication (Fujiwara et al., 2002); *SPP1* (osteopontin), a secreted extracellular matrix protein that plays role in implantation process (Berneau et al., 2019); *DKK1* (Dickkopf WNT signaling pathway inhibitor 1), plays role in embryonic development (Huang et al., 2018); *ARID5B* (AT-rich interactive domain 5B), plays role in cell growth and differentiation of B-lymphocyte progenitors, and is WNT-signaling pathway inhibitor (Lahoud, 2001); and *ANXA4* (annexin A4), a possible regulator of ion and water transport in the endometrial epithelium (Ponnampalam and Rogers, 2006).

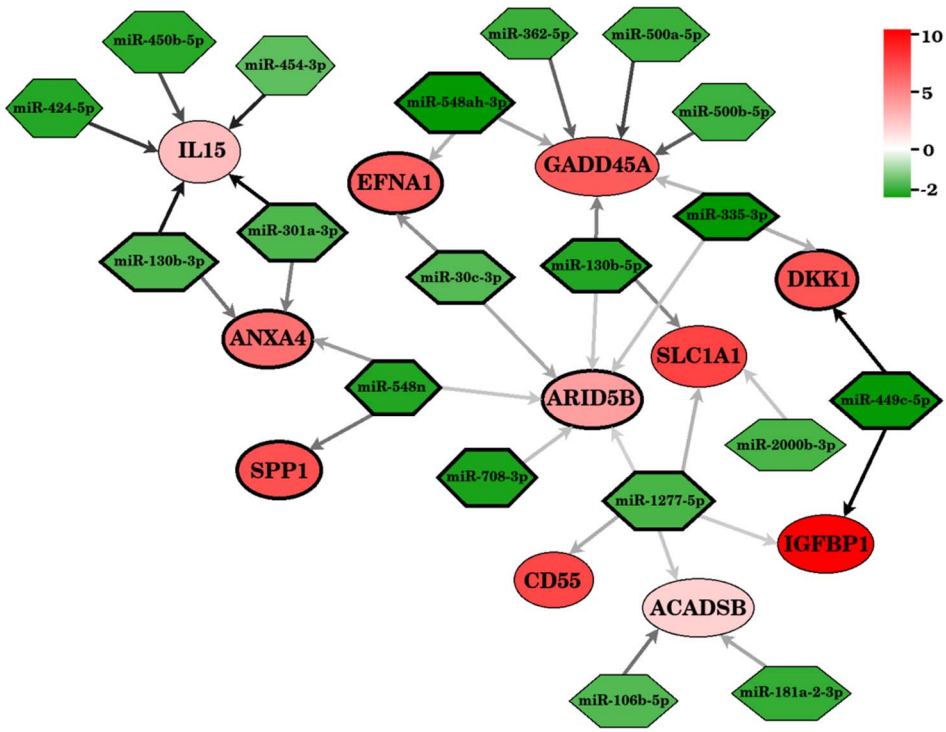

Figure 20. *In silico* predicted mRNA-miRNA interactions, where up-regulated mRNAs are in red circles and down-regulated miRNAs in green rectangles. In both datasets validated mRNA-miRNA pairs have bold borders. The color intensities show strength of up/down-regulation (FDR < 0.05) where arrow intensities indicate probability interactions (darker have higher probability) based on TargetScan ++ score. Adapted from Ref. II.

My role in the Ref. II publication was to put together mRNA and miRNA data layers in validation datasets for *in silico* predicted mRNA-miRNA pairs. The combination of two layers helps a) to find stronger evidence for already known marker genes for endometrial receptivity; b) to better understand the regulation of gene expression in endometrial tissue during the mid-secretory phase, when the endometrium is usually most receptive to embryo; c) to point out some possible causes of why endometrium is not receptive to the embryo and d) to find new potential receptivity biomarkers among the interacting mRNA and miRNA pairs.

3.3. Methylation pattern differences between early-secretory and mid-secretory endometrium and its correlation with gene expression

Recent studies have shown that methylation patterns might change during the menstrual cycle and correlate with changes in gene expression (Houshdaran et al., 2014, 2016). However, there is still a lack of knowledge about global methylation changes in endometrium during the transition from early secretory to mid-secretory state, and of the role methylation plays in the regulation of gene expression during that period. To understand how changes in endometrial DNA methylation patterns can affect endometrial receptivity and endometrial gene expression, a more thorough understanding of normal endometrial methylome is needed.

The aim of Ref. III study was to characterize the methylation patterns in early-secretory and mid-secretory phases using genome-wide technologies, and using integration with RNA-seq data, to find genes with expression changes possibly affected by changes in methylation.

3.3.1. Description of cohort, materials, and methods

The study was approved by the Ethics Review Committee of the University of Tartu, Estonia (permission no 221/M-31). Informed consent was signed by all women before tissue collection and all methods were carried out following relevant guidelines and regulations.

For this study the endometrial biopsies from early-secretory (LH+2) and mid-secretory endometrium (LH+8) were obtained from 17 healthy fertile women, thus in total 34 paired samples. All participants were non-smokers, had at least one born child and had not taken any hormonal medications at least three months before the biopsy. The Infinium HumanMethylation 450K BeadChip (Illumina, San Diego, CA, USA) was used for extracted DNA hybridization. RnBeads v.1.1.8 (Assenov et al., 2014) package was used for methylation data quality control and filtering. The RnBeads was chosen as it contains all the necessities for array-based methylation analysis, starting from

data filtering and ending with differential methylation analysis while producing graphical output during each step. This package also allows user to choose from several data normalization methods, has paired sampling possibility and exploratory analysis, making it convenient and easy to use. After all preprocessing steps, 437,022 out of 485,577 probes remained for further analysis. The probes were annotated according to Illumina annotation file and contained TSS200, TSS500, 5'UTR, 3'UTR, 1st exon, and gene body (gene coding region excluding 1st exon) regions. The CGI locations as Island, S and N shore, S and N shelf and Open sea (isolated CpGs in the genome) were also used. The differential methylation analysis on the site level was prepared by RnBeads, seqlm (Kolde et al., 2016), and Wilcoxon signed-rank test using m-values as input calculated with lumi (Du et al., 2008) package. Three distinct approaches were used to eliminate possible false-positive results and the results provided by them were comparable on a single CpG site level. The analysis was age-adjusted, as there are known effects of age on DNA methylation (Bell et al., 2012). The FDR (false discovery rate) < 0.05 was set as significance criteria and CpGs overlapping between all three analysis were considered most likely truly differentially methylated sites, that were later used in the correlation analysis.

The whole tissue RNA-seq dataset was used the same as in the Ref. I, except the number of samples for correlation analysis. Only 7 samples (14 paired samples) overlapped between methylation and expression datasets and were used for correlation. Briefly about expression data: about 1 μg of endometrial total RNA, extracted from the biopsies, was sequenced using paired-end Illumina TruSeq technology (Illumina) at Estonian Genome Center. The raw reads were trimmed and adapters removed using Trimmomatic-0.32 (Bolger et al., 2014), quality filtered with FASTQ quality filter tool from FASTX-Toolkit v.0.0.14, and mapped using TopHat2 (Kim et al., 2013) to human genome version 19 (hg19). The counts per million for further correlation analysis were obtained with edgeR (Robinson et al., 2010) tool.

For cis-correlation between significantly methylated CpG β values > 0.1 and significantly differentially expressed gene CPMs were used. Correlations were obtained with Spearman's rank test and permutation p-values were calculated to evaluate the significance. The enrichment analysis was performed using g:Profiler (Reimand et al., 2007, 2011) and PANTHER v11.1 (Mi et al., 2016).

3.3.2. Differential methylation in mid-secretory vs early-secretory endometrium

The study showed that for both early-secretory and mid-secretory phase about 19% and 33% CpGs were hyper- or hypomethylated, respectively. The pairwise comparison with Kolmogorov-Smirnov test of genomic locations relative to CpG islands and functional regions showed relatively low CpG methylation levels in islands compared to shelves in both receptivity phases (Figure 21 A), while TSS1500 CpG sites were slightly more methylated than TSS200 sites in the promoter region (Figure 21 B). The promoter regions on average were slightly more hypomethylated than gene body regions, indicating that active genomic regions are likely to be hypomethylated to provide access to transcription factors (Lokk et al., 2014). Overall, early-secretory and mid-secretory phases showed similar methylation patterns with no great magnitude changes (Figure 22).

The three approaches (RnBeads, Wilcoxon, and Seqlm) used for differential methylation analysis resulted in 22,272 overlapping CpG sites associated with 5,979 genes differentially methylated in mid-secretory phase (Figure 23) and were considered as the most likely set of truly differentially methylated sites and their associated genes. These sites were used in further site-level analyses and included both increased and decreased methylation levels in mid-secretory phase samples. Figure 24 shows the top 10 sites with the largest methylation differences between two endometrial phases. One of the top genes is *ZMIZ1*, that is significant in both site- and region-level analysis, is transcription factor regulator, and also regulates androgen receptor, Smad3/4 and p53 signaling (previously associated with endometrial receptivity (Altmäe et al., 2013; Revel et al., 2011)).

The DMR analysis, where DMR region is defined as 3 CpGs within 500 bp window, resulted in 2,026 significant regions in mid-secretory endometrium. Some of the CpG sites were located in genes previously associated with endometrial receptivity and embryo implantation, like *PAEP, GPX3, ARID5B,* and *ANXA4*. The most differentially methylated annotated sites and regions were located in gene bodies, a small fraction was located in promoter and other gene regions, while the majority of the sites and DMRs could not be annotated. By relation to the CpG islands, most of the sites were located in ‘Open sea’ while the location in the CpG islands was underrepresented.

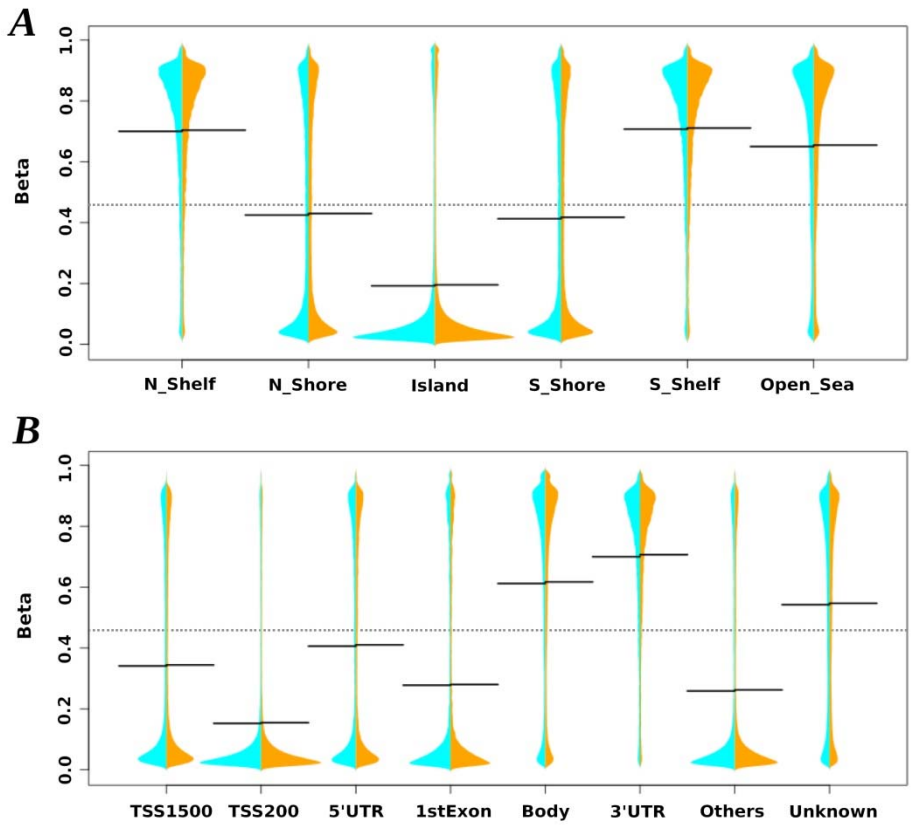

Figure 21. Methylation levels (Beta), obtained with Infinium HumanMethylation 450K BeadChip, in early-secretory (cyan-colored left side,) and mid-secretory (orange-colored right side) endometrium: **A.** by the location related to CpG island, **B.** by functional regions. The x-axis represents the location/region and the y-axis shows methylation level as β-value (0–1). The bean-plot width represents data distribution, the black line shows mean methylation level in the group and the dashed line shows the overall average methylation level. "Others" and "Unknown" indicate CpGs annotated to multiple locations and with unknown annotations, respectively. Adapted from Ref. III.

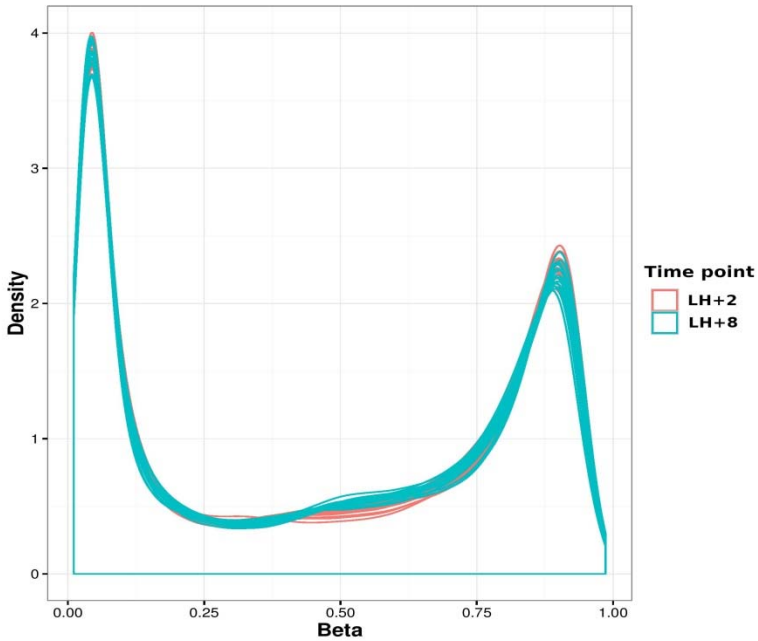

Figure 22. DNA methylation level and beta-density plot of endometrial genes between two time-points, obtained with Infinium HumanMethylation 450K BeadChip. The plot shows only minor methylation differences between early-secretory (LH+2, red lines) and mid-secretory (LH+8, blue lines) endometrium samples from 17 women (34 samples in total). Adapted from Ref. III.

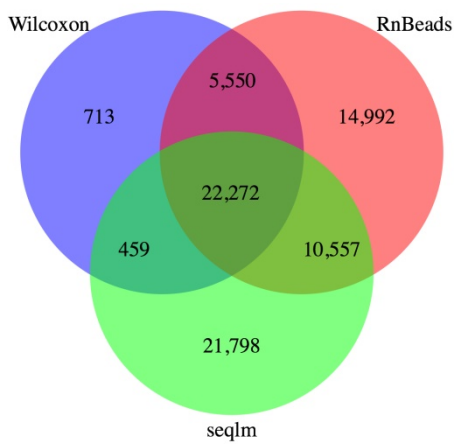

Figure 23. Venn diagram shows the overlap between three used analysis methods: Wilcoxon (blue), RnBeads (red) and seqlm (green). The numbers in the circles indicate the amount of overlapping and unique differentially methylated CpG sites.

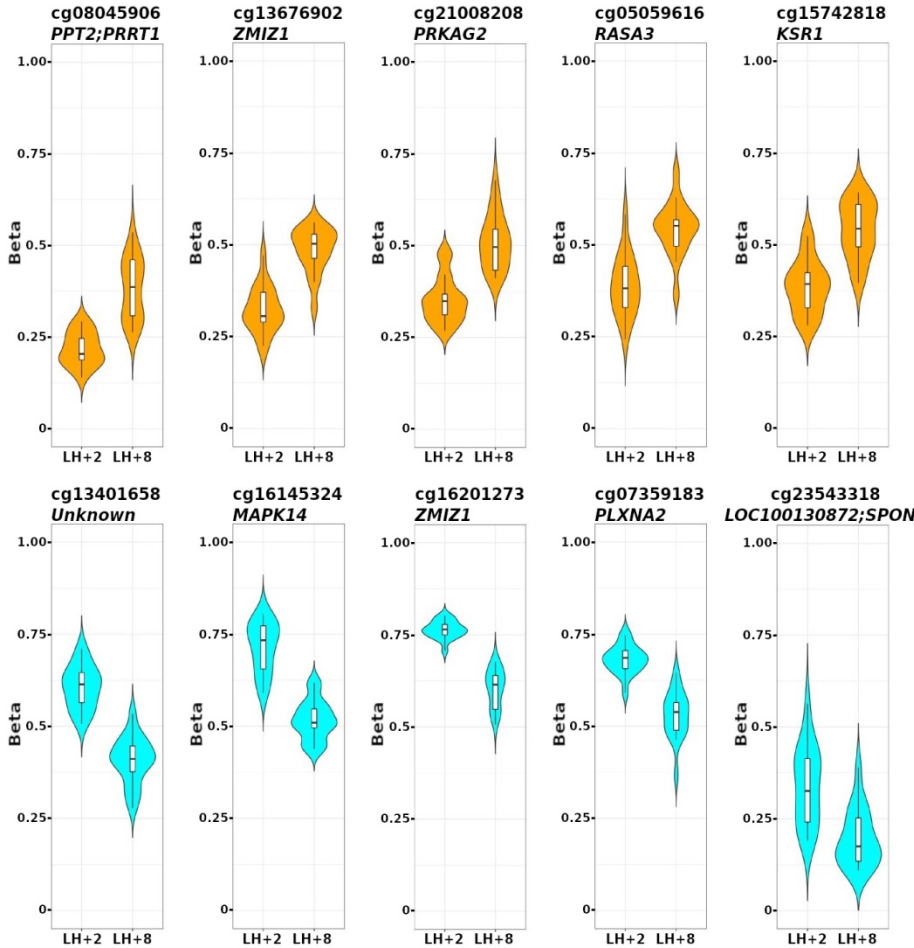

Figure 24. Top 10 differentially methylated CpGs between early-secretory (LH+2) and mid-secretory (LH+8) endometrium with their associated genes. The x-axis shows the secretory phase, y-axis shows methylation level (Beta), where colors indicate sites that have higher (orange) and lower (cyan) methylation levels in the mid-secretory phase. Adapted from Ref. III.

3.3.3. Correlation between methylation and gene expression levels, enrichment and pathway analyses

The correlation analysis was conducted between the RNA-seq dataset, where gene expression was evaluated in CPMs, and Illumina-annotated significantly methylated CpGs with absolute delta-β value > 0.1. Thus, all 'Open sea' sites were excluded from the analysis. In the end, a total of 464 genes and 531 corresponding CpGs were used for analysis, resulting in 536 gene-CpG pairs, including some sites that were annotated to more than one gene. The analysis

resulted in 169 significantly correlated gene-CpG pairs (permutation p-value <0.05), where positive correlations were more prevalent in gene body regions, while negative ones were mainly seen in 5'UTR and 1st exon regions (Table 2), which is consistent with "DNA methylation paradox" when hypermethylation of CGIs in gene body region leads to increased gene expression levels (Jones, 1999).

Table 2. The number of correlated CpG-gene pairs per region. Adapted from Ref. III.

Region	CpG-gene pairs per region (n)	Significantly correlated CpGs-gene pairs n *(% of total correlated)*	Positively correlated CpG-gene pairs n *(% of significantly correlated)*	Negatively correlated CpG-gene pairs n *(% of significantly correlated)*
TSS1500	38	9 *(23.7%)*	6 *(66.7%)*	3 *(33.3%)*
TSS200	16	4 *(25.0%)*	2 *(50%)*	2 *(50%)*
5' UTR	73	28 *(38.4%)*	11 *(39.3%)*	17 *(60.7%)*
1st Exon	18	4 *(22.2%)*	1 *(25%)*	3 *(75%)*
Body	353	109 *(30.9%)*	62 *(56.9%)*	47 *(43.1%)*
3'UTR	48	15 *(31.3%)*	8 *(53.3%)*	7 *(46.7%)*
Total	**536**	**169** ***(31.5%)***	**90** ***(53.3%)***	**79** ***(46.7)***

The gene ontology and pathway analyses were prepared on the site and region levels by V. Modhukur with g:Profiler and PANTHER tools, where for site-level 22,272 CpGs were used, corresponding to 1,464 hypomethylated and 5,196 hypermethylated genes, that shared 681 genes because of multiple annotations; while region-level analysis included 1,206 hypermethylated and 275 hypomethylated genes in mid-secretory phase.

The g:Profiler site level analysis showed that genes with increased methylation levels were associated with extracellular matrix organization, cellular signaling, and development, while genes with decreased methylation associated with immune response regulation, cell activation, and adhesion. The region-level analysis showed associations with the extracellular matrix and cellular adhesion, which is similar to site-level analyses. The functional analysis of significantly correlated gene-CpG pairs resulted in genes related to extracellular matrix organization (*TGFB3, COL42A, ADAMTS1*) and immune response for positively correlated genes (*IL1RL1, FYN, BCL3*), and no enrichment was seen for negatively correlated genes.

The analysis done using the PANTHER tool was not able to detect enrichment on region level, but the site-level enrichment resulted in 16 pathways, for example, angiogenesis, integrin signaling, Wnt signaling and GnRH receptor, and chemokine/cytokine signaling mediated inflammation pathways. The pathway analysis of significantly correlated genes showed enrichment in the integrin signaling pathway. Some of these pathways like extracellular matrix remodeling, immune response, and integrin signaling are already shown to play an important role in endometrial receptivity and functionality, through modifying maternal immunity and tissue remodeling (Ruiz-Alonso et al., 2012).

This study provides insight into methylation pattern changes between early-secretory and mid-secretory endometrium, additionally connecting these changes with changes in gene expression. Putting together two data layers from the same samples helps to gain a better understanding of gene regulation during the transition from early- to mid-secretory phase and at the same time eliminates interindividual differences. The limitations of the study are relatively limited sample size and whole tissue biopsy without cellular decomposition, thus the methylation profile for distinct cell populations and validation on bigger datasets are highly recommended.

CONCLUSIONS

The understanding of processes and mechanisms underlying endometrial receptivity helps to improve the detection of pathological states related to female infertility and failure of infertility treatment. Using and integrating several 'omics' layers could highlight possible gene expression regulatory pathways and interactions between data layers, narrow down lists of biomarker candidates, help to find new and better biomarkers, and give a better understanding of underlying processes in female reproduction, putting separate pieces of the puzzle together.

The main conclusions drawn from the current thesis are as follows:

- Variations in cell fractions could drastically influence the results of differential expression analysis, as the expression profiles of overrepresented cell types could mask the expression values of underrepresented ones. Thus, it is important to deconvolute transcriptomic data of whole tissue biopsy samples and adjust the analysis by cell fractions, as biopsies have uneven amounts of different cell types, particularly in the case of highly heterogeneous endometrial tissue. We showed that computational deconvolution provides trustworthy results for endometrial tissue transcriptomic analysis and could be a big improvement in the identification of robust endometrial receptivity biomarkers.
- Endometrial receptivity is governed by the precise interplay between a set of miRNAs and their target genes/transcripts as revealed by the interactions predicted based on three public databases and our experimental results. The correlations found between miRNAs and their target mRNAs allow for a better understanding of the expression regulation of endometrial receptivity biomarkers.
- The differences in methylation of genes between the early-secretory and receptive/mid-secretory endometrial phases are minor. However, the correlation analysis between the epigenetic markers of certain genes and their expression changes required for the endometrial receptivity helps to identify the endometrial receptivity biomarkers, which are regulated by epigenetic mechanisms. These correlations between the two 'omics' data layers could give some hints of possible regulatory pathways and mechanisms for obtaining endometrial receptivity, as well as help to select the most robust and reliable biomarkers for clinical purposes.

This thesis shows that integration of 'omics' data layers is helping to better understand the processes underlying endometrial receptivity, and is instrumental for gathering novel information, supporting already existing knowledge and helping to avoid false-positive results, which is currently considered as one of the major drawbacks of 'omics' approaches. The work also encourages further studies to integrate even more data layers, to get the best possible understanding of the endometrial functioning.

SUMMARY IN ESTONIAN

Geeniekspressiooni andmete integreerimine teiste 'oomika' andmetega kirjeldamaks endomeetriumi retseptiivsuse bioloogilisi mehhanisme

Rasestumiseks on vaja regulaarset ning korrektselt reguleeritud menstruaaltsüklit. Menstruaaltsükkel on jagatud erinevateks faasideks. Rasestumiseks on eriti olulised endomeetriumi ehk emaka sisemise limaskesta proliferatsiooni (paksenemine) faas ja sellele järgnev sekretoorne faas. Sekretoorne faas omakorda jaguneb kolmeks alamfaasiks: varajaseks (pre-retseptiivne), keskmiseks (retseptiivne) ja hiliseks (post-retseptiivne). Varajases alamfaasis endomeetrium on jõudnud optimaalse paksuseni, kuid ei ole veel embrüo vastuvõtmiseks valmis. Keskmises alamfaasis endomeetrium on valmis embrüot vastu võtma, siis kui hilises alamfaasis embrüo pesastumine ei ole enam võimalik. Menstruaaltsükkel algab emaka ettevalmistamisega tulevase embrüo vastuvõtuks. Sellel ajal endomeetrium pakseneb ning umbes tsükli 14. päevaks (tavalise 28 päevase tsükli korral) jõuab optimaalse paksuseni. Samal päeval toimub ka munaraku ovulatsioon. Alates *ca* 20. päevast algab 48-tunnine periood ehk implantatsiooni aken, mille ajal endomeetrium on kõige vastuvõtlikum (retseptiivne) embrüo suhtes. Just implantatsiooni akna ajal toimub embrüo pesastumine emakasse. Mõlemad, nii endomeetriumi küpsemine kui ka embrüo pesastumine nõuavad täpselt reguleeritud koostööd erinevate faktorite (signaalmolekulid, valgud, RNA molekulid jne), rakkude ja kudede vahel. Seda kõike on võimalik uurida erinevatel bioloogilistel tasemetel, näiteks genoomi struktuursed muutused (genoomika), DNA modifikatsioonid sh metülatsioon (epigenoomika), geenide avaldumine ja DNA transkriptsioon (transkriptoomika), valkude avaldumine ja nende struktuur (proteoomika), metaboliidid (metaboloomika). Teadusharusid, mis neid bioloogilisi tasemeid uurivad, kollektiivselt kutsutakse "oomika" teadusteks. Igal "oomika" harul on omad uurimismeetodid, kuid enamik nendest meetoditest põhineb siiski sekveneerimisel (DNA järjestuse määramine) ja mikrokiibi analüüsil. Sekveneerimise ja mikrokiibi andmeid analüüsitakse erinevalt, kasutades selleks spetsiifilisi programmide pakette. Andmeanalüüsi programmi või paketi peab kasutaja valima lähtudes enda andmetest, nende päritolust ja kvaliteedist. Mõnel juhul on valikut teha keeruline, ning siis tuleb pakette ükshaaval läbi proovida, et leida nende hulgast parim lähenemisviis. Heaks näiteks programmide rohkusest ja sobiva programmi valikust on metülatsiooni andmete normaliseerimine. Kuna üks "oomika" haru uurib ainult ühte bioloogilise info taset, võimaldab see näha ainult terviku ühte osa. Tervikpildi nägemiseks on aga vajalik erinevate "oomika" tasemete kombineerimine.

Antud töös olen kombineerinud geenide ekspressiooni (mRNA) andmeid endomeetriumi koe tasemel, koos mRNA andmetega endomeetriumi rakutüübi tasemel; mittekodeerivate mikroRNA-de (miRNA) andmetega ja endomeetriumi DNA metülatsiooni andmetega, ning analüüsinud geenide metülatsiooni,

miRNA-de ja mRNA-de tasemete omavahelisi korrelatsioone. Läbiviidud uuringute eesmärgiks oli paremini aru saada, mille poolest erineb pre-retseptiivne endomeetrium (varajane sekretoorne faas) retseptiivsest endomeetriumist (kesk-sekretoorne faas), ehk millised muutused on vajalikud, et endomeetrium muutuks embrüole vastuvõtlikuks.

Endomeetriumi kude koosneb erinevatest rakutüüpidest, lisaks muutub koe rakuline koostis tsükli vältel pidevalt. Varasemalt on näidatud, et geenide avaldumine on eri rakutüüpides erinev, ning rakutüüpide proportsioonid koeproovides võivad mõjutada analüüsi tulemusi, st, et enamesindatud rakkude geenide avaldumise muster varjab vähem esindatud rakutüüpide geenide avaldumise mustrit. Järelikult kajastub analüüsi tulemustes ennekõike enimlevinud rakutüübi avaldumise muster, mitte üldine genoomi avaldumise muster. Meie uurimisobjektideks olid endomeetriumi koeproovid ning eelnevalt puudus informatsioon rakkude proportsioonide kohta uuritavas koes ja erinevatele rakutüüpidele iseloomulik geenide aktiivsuse muster. Endomeetriumi koeproovide analüüsil on nimetatud probleem eriti terav, kuna kude on heterogeenne, sisaldades mitmeid erinevaid rakutüüpe, mis kõik täidavad kindlaid ülesandeid emaka limaskesta funktsioneerimisel. Koe heterogeensuse probleemi lahendamiseks kasutasime endomeetriumi kahe peamise rakutüübi (strooma ja epiteeli rakkude) mRNA andmeid selleks, et *in silico* lähenemise abil hinnata rakulist koosseisu meie koeproovides. Analüüsi tulemusena tuvastasime potentsiaalseid markergeene, mis võimaldavad arvestada koe rakulist koostist ning tuvastada biomarkereid, mille ekspressiooni muutus ei sõltu erinevate rakupopulatsioonide osakaalust. Antud uuringutega näitasime, et juhul, kui puudub otsene informatsioon koe rakulise koostise kohta, on siiski võimalik usaldusväärselt kasutada *in silico* lähenemist, mis võitab arvesse koe rakulise heterogeensuse.

Endomeetriumi mRNA ja miRNA ekspressiooni andmeid kombineerisin selleks, et hinnata andmebaasidel põhinevaid miRNA-mRNA vahelisi interaktsioone, mille korral miRNA reguleerib oma märklaudgeenide transkriptsiooni taset. Oma uuringus kasutasin kolleegide poolt publikatsioonide ja andmebaaside alusel koostatud nimekirjasid geenidest ja neid potentsiaalselt reguleerivatest miRNA-dest, mille interaktsioonid tõenäoliselt omavad olulist rolli endomeetriumi retseptiivsuse kujunemises. Järgneva analüüsi käigus korreleerisin antud geenide ja miRNA-de ekspressiooni väärtusi ning antud analüüsi tulemusena tuvastasin 17 miRNA-d ja 9 nendega seotud märklaudgeeni, mis on seotud endomeetriumi retseptiivsusega.

Endomeetriumi DNA metülatsiooni ja geenide ekspressiooni andmestike ühildamine aitas tuvastada, kas muutused geenide metülatsiooni mustrites kajastuvad geenide avaldumises. Üldjuhul põhjustab geenide promootorpiirkondade metüleeritus nende geenide ekspressiooni langust, samas kui hüpometülatsioon viib ekspressiooni tõusuni. Lisaks on tuvastatud ka ebatüüpiline epigeneetilise regulatsiooni mehhanism nn – "metülatsiooni paradoks", mille puhul põhjustab hüpermetülatsioon geeni kodeerivas piirkonnas või transkriptsiooni alguspunktis antud geeni ekspressiooni tõusu. Käesolevas töös kasutasime metülatsiooni muutuste tuvastamiseks kolme meetodit, selleks et eemaldada võimalikult palju

valepositiivseid tulemusi. Kolme meetodi vahel kattunud metülatsiooni analüüsi tulemusi võrdlesime geeniekspressiooni väärtustega. Korrelatsioonanalüüsi tulemusena leidsime 169 geeni ja vastavas geenis asetseva metülatsioonisaidi paari. Need geenid, mis korreleerusid positiivselt (geeni metülatsioonis ja ekspressiooni muutused toimuvad samas suunas) olid seotud rakkude välise maatriksi struktuuri ja immuunvastuse protsessidega. Varem on samuti näidatud nimetatud bioloogiliste protsesside seost endomeetriumi retseptiivsusega. Siiski jäid metülatsiooni muutused üleminekul pre-retseptiivsest faasist retseptiivse endomeetriumi faasi suhteliselt tagasihoidlikuks, mistõttu võib arvata, et epigeneetiline genoomi regulatsioon mängib endomeetriumi retseptiivsuse kujunemises küllaltki väikest rolli. Kokkuvõtvalt näitavad antud töö tulemused, et "oomika" andmekihtide kombineerimine aitab paremini mõista endomeetriumi retseptiivsusega seotud bioloogilisi protsesse, mis ei oleks võimalik, kui kasutatakse ainult ühe andmekihi infot.

REFERENCES

Abbas, A.R., Wolslegel, K., Seshasayee, D., Modrusan, Z., and Clark, H.F. (2009). Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus. PLoS ONE *4*, e6098.

Achache, H., and Revel, A. (2006). Endometrial receptivity markers, the journey to successful embryo implantation. Human Reproduction Update *12*, 731–746.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. ELife *4*, e05005.

Allen, F.W. (1941). The Biochemistry of the Nucleic Acids, Purines, and Pyrimidines. Annual Review of Biochemistry *10*, 221–244.

Altmae, S., Martinez-Conejero, J.A., Salumets, A., Simon, C., Horcajadas, J.A., and Stavreus-Evers, A. (2010). Endometrial gene expression analysis at the time of embryo implantation in women with unexplained infertility. Molecular Human Reproduction *16*, 178–187.

Altmäe, S., Martinez-Conejero, J.A., Esteban, F.J., Ruiz-Alonso, M., Stavreus-Evers, A., Horcajadas, J.A., and Salumets, A. (2013). MicroRNAs miR-30b, miR-30d, and miR-494 Regulate Human Endometrial Receptivity. Reproductive Sciences *20*, 308–317.

Altmäe, S., Esteban, F.J., Stavreus-Evers, A., Simón, C., Giudice, L., Lessey, B.A., Horcajadas, J.A., Macklon, N.S., D'Hooghe, T., Campoy, C., et al. (2014). Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium. Human Reproduction Update *20*, 12–28.

Altmäe, S., Koel, M., Võsa, U., Adler, P., Suhorutšenko, M., Laisk-Podar, T., Kukushkina, V., Saare, M., Velthut-Meikas, A., Krjutškov, K., et al. (2017). Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers. Scientific Reports *7*.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data, URL: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Acessed: 2019-05-16).

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nature Methods *11*, 1138–1140.

Avila Cobos, F., Vandesompele, J., Mestdagh, P., and De Preter, K. (2018). Computational deconvolution of transcriptomics data from mixed cell populations. Bioinformatics *34*, 1969–1979.

Bates, G.W., and Bowling, M. (2013). Physiology of the female reproductive axis. Periodontology 2000 *61*, 89–102.

Battaglia, R., Vento, M.E., Borzì, P., Ragusa, M., Barbagallo, D., Arena, D., Purrello, M., and Di Pietro, C. (2017). Non-coding RNAs in the Ovarian Follicle. Front. Genet. *8*, 57.

Beck, S. (2010). Taking the measure of the methylome. Nat Biotechnol *28*, 1026–1028.
Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., et al. (2012). Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. PLoS Genetics *8*, e1002629.
Berneau, S.C., Ruane, P.T., Brison, D.R., Kimber, S.J., Westwood, M., and Aplin, J.D. (2019). Characterisation of Osteopontin in an In Vitro Model of Embryo Implantation. Cells *8*, 432.
Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biology *11*, R90.
Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.
Bind, M.-A., Baccarelli, A., Zanobetti, A., Tarantini, L., Suh, H., Vokonas, P., and Schwartz, J. (2012). Air Pollution and Markers of Coagulation, Inflammation, and Endothelial Function: Associations and Epigene-environment Interactions in an Elderly Cohort. Epidemiology *23*, 332–340.
Bock, C. (2012). Analysing and interpreting DNA methylation data. Nat Rev Genet *13*, 705–719.
Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G., and Meissner, A. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol *28*, 1106–1114.
Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
Boren, T., Xiong, Y., Hakam, A., Wenham, R., Apte, S., Wei, Z., Kamath, S., Chen, D.-T., Dressman, H., and Lancaster, J.M. (2008). MicroRNAs and their target messenger RNAs associated with endometrial carcinogenesis. Gynecologic Oncology *110*, 206–215.
Borthwick, J.M., Charnock-Jones, D.S., Tom, B.D., Hull, M.L., Teirney, R., Phillips, S.C., and Smith, S.K. (2003). Determination of the transcript pro®le of human endometrium. Molecular Human Reproduction *9*, 19–33.
Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525–527.
Carson, D.D., Lagow, E., Thahiah, A., Al-Shami, R., Farach-Carson, M.C., Vernon, M., Yuan, L., Fritz, M.A., and Lessey, B. (2002). Changes in gene expression during the early to mid-luteal (receptive phase) transition in human endometrium detected by high-density microarray screening. Molecular Human Reproduction *8*, 871–879.
Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. IJMS *17*, 1712.
Cavagna, M., and Mantese, J.C. (2003). Biomarkers of Endometrial Receptivity—A Review. Placenta *24*, S39–S47.
Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics *10*, 295–304.
Chen, B.J., Byrne, F.L., Takenaka, K., Modesitt, S.C., Olzomer, E.M., Mills, J.D., Farrell, R., Hoehn, K.L., and Janitz, M. (2017a). Transcriptome landscape of long intergenic non-coding RNAs in endometrial cancer. Gynecologic Oncology *147*, 654–662.

Chen, J., Wang, L., Wang, F., Liu, J., and Bai, Z. (2020). Genomic Identification of RNA Editing Through Integrating Omics Datasets and the Clinical Relevance in Hepatocellular Carcinoma. Front. Oncol. *10*, 37.

Chen, Z., Huang, A., Sun, J., Jiang, T., Qin, F.X.-F., and Wu, A. (2017b). Inference of immune cell composition on the expression profiles of mouse tissue. Scientific Reports *7*, 40508.

Costa-Silva, J., Domingues, D., and Lopes, F.M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. PLoS ONE *12*, e0190152.

Cotteret, S., Jaffer, Z.M., Beeser, A., and Chernoff, J. (2003). p21-Activated Kinase 5 (Pak5) Localizes to Mitochondria and Inhibits Apoptosis by Phosphorylating BAD. Molecular and Cellular Biology *23*, 5526–5539.

Cunningham, F., Leveno, K., Bloom, S., Dashe, J., Hoffman, B., Casey, B., and Spong, C. (2018). Implantation and Placental Development. In Williams Obstetrics, (New York, NY: McGraw-Hill), p.

Da Vinci, L. (1510). Studies of the Foetus in the Womb, URL: https://commons.wikimedia.org/wiki/File:Leonardo_da_Vinci_-_Studies_of_the_foetus_in_the_womb.jpg (Accessed: 2019-04-23).

Damario, M.A. (2014). General Aspects of Fertility and Infertility. In Human Fertility, Z. Rosenwaks, and P.M. Wassarman, eds. (New York, NY: Springer New York), pp. 3–23.

Das, S.K. (2014). Integrating Transcriptome and Epigenome: Putting Together the Pieces of the Type 2 Diabetes Pathogenesis Puzzle. Diabetes *63*, 2901–2903.

Davis, S., Du, P., Bilke, S., Triche, T.J., and Bootwalla, M. (2019). methylumi: Handle Illumina methylation data. R package version 2.30.0.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771–784.

Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G., and Fuks, F. (2014). A comprehensive overview of Infinium HumanMethylation450 data processing. Briefings in Bioinformatics *15*, 929–941.

Di Pietro, C., Cicinelli, E., Guglielmino, M.R., Ragusa, M., Farina, M., Palumbo, M.A., and Cianci, A. (2013). Altered Transcriptional Regulation of Cytokines, Growth Factors, and Apoptotic Proteins in the Endometrium of Infertile Women with Chronic Endometritis. American Journal of Reproductive Immunology *69*, 509–517.

Dias, B.G., and Ressler, K.J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci *17*, 89–96.

Díaz-Gimeno, P., Horcajadas, J.A., Martínez-Conejero, J.A., Esteban, F.J., Alamá, P., Pellicer, A., and Simón, C. (2011). A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. Fertility and Sterility *95*, 50-60.e15.

Díaz-Gimeno, P., Ruiz-Alonso, M., Blesa, D., Bosch, N., Martínez-Conejero, J.A., Alamá, P., Garrido, N., Pellicer, A., and Simón, C. (2013). The accuracy and reproducibility of the endometrial receptivity array is superior to histology as a diagnostic method for endometrial receptivity. Fertility and Sterility *99*, 508–517.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547–1548.

Editorial (2010). E pluribus unum. Nat Methods *7*, 331–331.
Egashira, M., and Hirota, Y. (2013). Uterine receptivity and embryo-uterine interactions in embryo implantation: lessons from mice. Reproductive Medicine and Biology *12*, 127–132.
Enciso, M., Carrascosa, J.P., Sarasa, J., Martínez-Ortiz, P.A., Munné, S., Horcajadas, J.A., and Aizpurua, J. (2018). Development of a new comprehensive and reliable endometrial receptivity map (ER Map/ER Grade) based on RT-qPCR gene expression analysis. Human Reproduction *33*, 220–228.
Evans, C., Hardin, J., and Stoebel, D.M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Briefings in Bioinformatics *19*, 776–792.
Fan, L., Han, H., Guan, J., Zhang, X., Cui, Q., Shen, H., and Shi, C. (2017). Aberrantly expressed long noncoding RNAs in recurrent implantation failure: A microarray related study. Systems Biology in Reproductive Medicine *63*, 269–278.
Fan, Y., Wang, Y., Fu, S., Yang, L., Lin, S., Fan, Q., and Wen, Q. (2018). The diagnostic role of DNA methylation in sporadic endometrial cancer: a systematic review and meta-analysis. Oncotarget *9*.
Feng, C., Shen, J.-M., Lv, P.-P., Jin, M., Wang, L.-Q., Rao, J.-P., and Feng, L. (2018). Construction of implantation failure related lncRNA-mRNA network and identification of lncRNA biomarkers for predicting endometrial receptivity. Int. J. Biol. Sci. *14*, 1361–1377.
Figueira, P.G.M., Abrão, M.S., Krikun, G., and Taylor, H. (2011). Stem cells in endometrium and their role in the pathogenesis of endometriosis: Stem cells, the endometrium, and endometriosis. Annals of the New York Academy of Sciences *1221*, 10–17.
Findlen, P. (1998). A history of the male and female genitalia, URL: https://web.stanford.edu/class/history13/earlysciencelab/body/femalebodypages/genitalia.html (Acessed: 09.05.2019).
Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B.W., Hudson, T.J., Fertig, E.J., Greenwood, C.M., and Hansen, K.D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. 17.
Fortin, J.-P., Triche, T.J., and Hansen, K.D. (2016). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics btw691.
Fraga, M.F., and Esteller, M. (2002). DNA Methylation: A Profile of Methods and Applications. *33*, 16.
Franklin, T.B., Russig, H., Weiss, I.C., Gräff, J., Linder, N., Michalon, A., Vizi, S., and Mansuy, I.M. (2010). Epigenetic Transmission of the Impact of Early Stress Across Generations. Biological Psychiatry *68*, 408–415.
Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Research *40*, 37–52.
Fujiwara, H., Yoshioka, S., Tatsumi, K., Kosaka, K., Satoh, Y., Nishioka, Y., Egawa, M., Higuchi, T., and Fujii, S. (2002). Human Endometrial Epithelial Cells Express Ephrin A1: Possible Interaction between Human Blastocysts and Endometrium via Eph-Ephrin System. The Journal of Clinical Endocrinology & Metabolism *87*, 5801–5807.

Furlong, H.C., Stämpfli, M.R., Gannon, A.M., and Foster, W.G. (2018). Identification of microRNAs as potential markers of ovarian toxicity: miRNAs Markers for Ovarian Toxicity. J Appl Toxicol *38*, 744–752.

Gao, F., Ma, X., Rusie, A., Hemingway, J., Ostmann, A.B., Chung, D., and Das, S.K. (2012). Epigenetic Changes Through DNA Methylation Contribute to Uterine Stromal Cell Decidualization. Endocrinology *153*, 6078–6090.

Giacomini, E., Vago, R., Sanchez, A.M., Podini, P., Zarovni, N., Murdica, V., Rizzo, R., Bortolotti, D., Candiani, M., and Viganò, P. (2017). Secretome of in vitro cultured human embryos contains extracellular vesicles that are uptaken by the maternal side. Scientific Reports *7*.

Giudice, L.C. (2004). Microarray Expression Profiling Reveals Candidate Genes for Human Uterine Receptivity. Am J Pharmacogenomics 14.

Gómez, E., Ruíz-Alonso, M., Miravet, J., and Simón, C. (2015). Human Endometrial Transcriptomics: Implications for Embryonic Implantation. Cold Spring Harbor Perspectives in Medicine *5*, a022996.

Gong, T., and Szustakowski, J.D. (2013). DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics *29*, 1083–1085.

Gong, T., Hartmann, N., Kohane, I.S., Brinkmann, V., Staedtler, F., Letzkus, M., Bongiovanni, S., and Szustakowski, J.D. (2011). Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples. PLoS ONE *6*, e27156.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol *29*, 644–652.

Graziottin, A., and Gambini, D. (2015). Anatomy and physiology of genital organs – women. In Handbook of Clinical Neurology, (Elsevier), pp. 39–60.

Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D.C., and Shyr, Y. (2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. Genomics *109*, 83–90.

Gupta, R., Nagarajan*, A., and Wajapeyee, N. (2010). Advances in genome-wide DNA methylation analysis. Biotech. *49*, iii–xi.

Haller-Kikkatalo, K., Altmäe, S., Tagoma, A., Uibo, R., and Salumets, A. (2014). Autoimmune Activation toward Embryo Implantation is Rare in Immune-Privileged Human Endometrium. Seminars in Reproductive Medicine *32*, 376–384.

Hannon lab (2009). FASTX-Toolkit, URL: http://cancan.cshl.edu/labmembers/gordon/fastx_toolkit/ (Accessed: 2019-05-16).

Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol *13*, R83.

Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics *11*, 422.

Harrison, A., and Parle-McDermott, A. (2011). DNA Methylation: A Timeline of Methods and Applications. Front. Gene. *2*.

Hayman, D.M., Blumberg, T.J., Scott, C.C., and Athanasiou, K.A. (2006). The Effects of Isolation on Chondrocyte Gene Expression. Tissue Engineering *12*, 2573–2581.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences *105*, 17046–17049.

Hertwig, O. (1876). Beiträge zur Kenntniss der Bildung, Befruchtung und Theilung des thierischen Eies. Morphol. Jahrb. *1*, 347–434.

Himelstein-Braw, R., Byskov, A.G., Peters, H., and Faber, M. (1976). Follicular atresia in the infant human ovary. J Reprod Fertil *46*, 55.

Hogan, M.C., Griffin, M.D., Rossetti, S., Torres, V.E., Ward, C.J., and Harris, P.C. (2003). PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. Human Molecular Genetics *12*, 685–698.

Homer, H., Rice, G.E., and Salomon, C. (2017). Review: Embryo- and endometrium-derived exosomes and their potential role in assisted reproductive treatments–liquid biopsies for endometrial receptivity. Placenta *54*, 89–94.

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.

Houshdaran, S., Zelenko, Z., Irwin, J.C., and Giudice, L.C. (2014). Human Endometrial DNA Methylome Is Cycle-Dependent and Is Associated With Gene Expression Regulation. Molecular Endocrinology *28*, 1118–1135.

Houshdaran, S., Nezhat, C.R., Vo, K.C., Zelenko, Z., Irwin, J.C., and Giudice, L.C. (2016). Aberrant Endometrial DNA Methylome and Associated Gene Expression in Women with Endometriosis. Biology of Reproduction *95*, 93–93.

Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis: RNA-Seq. WIREs RNA *8*, e1364.

Hu, P., Zhang, W., Xin, H., and Deng, G. (2016). Single Cell Isolation and Analysis. Front. Cell Dev. Biol. *4*.

Hu, S., Yao, G., Wang, Y., Xu, H., Ji, X., He, Y., Zhu, Q., Chen, Z., and Sun, Y. (2014). Transcriptomic Changes During the Pre-Receptive to Receptive Transition in Human Endometrium Detected by RNA-Seq. The Journal of Clinical Endocrinology & Metabolism *99*, E2744–E2753.

Huang, Y., Liu, L., and Liu, A. (2018). Dickkopf-1: Current knowledge and related diseases. Life Sciences *209*, 249–254.

Hull, M.L., and Nisenblat, V. (2013). Tissue and circulating microRNA influence reproductive function in endometrial disease. Reproductive BioMedicine Online *27*, 515–529.

Hüttenhofer, A., Schattner, P., and Polacek, N. (2005). Non-coding RNAs: hope or hype? Trends in Genetics *21*, 289–297.

Irwin, J.C., Kirk, D., King, R.J.B., Quigley, M.M., and Gwatkin, R.B.L. (1989). Hormonal regulation of human endometrial stromal cells in culture: an in vitro model for decidualization. Fertility and Sterility *52*, 761–768.

Jiang, S.-W., Li, J., Podratz, K., and Dowdy, S. (2017). Application of DNA methylation biomarkers for endometrial cancer management. 18.

Jones, P.A. (1999). The DNA methylation paradox. Trends in Genetics *15*, 34–37.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484–492.

Kao, L.C., Tulac, S., Lobo, S., Imani, B., Yang, J.P., Germeyer, A., Osteen, K., Taylor, R.N., Lessey, B.A., and Giudice, L.C. (2002). Global Gene Profiling in Human Endometrium during the Window of Implantation. 20.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology *14*, R36.

King, A., Wellings, V., Gardner, L., and Loke, Y.W. (1989). Immunocytochemical characterization of the unusual large granular lymphocytes in human endometrium throughout the menstrual cycle. Human Immunology *24*, 195–205.

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics *28*, 573–580.

Kolde, R., Märtens, K., Lokk, K., Laur, S., and Vilo, J. (2016). seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics *32*, 2604–2610.

Koot, Y.E.M., van Hooff, S.R., Boomsma, C.M., van Leenen, D., Groot Koerkamp, M.J.A., Goddijn, M., Eijkemans, M.J.C., Fauser, B.C.J.M., Holstege, F.C.P., and Macklon, N.S. (2016). An endometrial gene expression signature accurately predicts recurrent implantation failure after IVF. Scientific Reports *6*.

Koukoura, O., Sifakis, S., and Spandidos, D.A. (2016). DNA methylation in endometriosis (Review). Molecular Medicine Reports *13*, 2939–2948.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucl. Acids Res. *42*, D68–D73.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Research *47*, D155–D162.

Krjutškov, K., Katayama, S., Saare, M., Vera-Rodriguez, M., Lubenets, D., Samuel, K., Laisk-Podar, T., Teder, H., Einarsdottir, E., Salumets, A., et al. (2016). Single-cell transcriptome analysis of endometrial tissue. Human Reproduction *31*, 844–853.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long Noncoding RNAs: Past, Present, and Future. Genetics *193*, 651–669.

Kuokkanen, S., Chen, B., Ojalvo, L., Benard, L., Santoro, N., and Pollard, J.W. (2010). Genomic Profiling of MicroRNAs and Messenger RNAs Reveals Hormonal Regulation in MicroRNA Expression in Human Endometrium1. Biology of Reproduction *82*, 791–801.

Kurdyukov, S., and Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. Biology *5*, 3.

Kyono, K., Hashimoto, T., Nagai, Y., and Sakuraba, Y. (2018). Analysis of endometrial microbiota by 16S ribosomal RNA gene sequencing among infertile patients: a single-center pilot study. Reproductive Medicine and Biology *17*, 297–306.

Lahoud, M.H. (2001). Gene Targeting of Desrt, a Novel ARID Class DNA-Binding Protein, Causes Growth Retardation and Abnormal Development of Reproductive Organs. Genome Research *11*, 1327–1334.

Laird, P.W. (2010). Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet *11*, 191–203.

Laqueur, T. (2003). Making sex: body and gender from the Greeks to Freud (Cambridge, Mass.: Harvard University Press).

Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K., and Ritchie, M.E. (2018). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res *5*.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet *11*, 733–739.

Leonard, S.M., Wei, W., Collins, S.I., Pereira, M., Diyaf, A., Constandinou-Williams, C., Young, L.S., Roberts, S., and Woodman, C.B. (2012). Oncogenic human papillomavirus imposes an instructive pattern of DNA methylation changes which parallel the natural history of cervical HPV infection in young women. Carcinogenesis *33*, 1286–1293.

Li, B., Liu, J.S., and Liu, X.S. (2017). Revisit linear regression-based deconvolution methods for tumor gene expression data. Genome Biology *18*, 127.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 16.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, J., and Siegmund, K.D. (2016). An evaluation of processing methods for HumanMethylation450 BeadChip data. BMC Genomics *17*, 469.

Liu, L., Wylie, R.C., Andrews, L.G., and Tollefsbol, T.O. (2003). Aging, cancer and nutrition: the DNA methylation connection. Mechanisms of Ageing and Development *124*, 989–998.

Liu, W., Yuan, K., and Ye, D. (2008a). Reducing microarray data via nonnegative matrix factorization for visualization and clustering analysis. Journal of Biomedical Informatics *41*, 602–606.

Liu, Y., Lee, K.-F., Ng, E.H.Y., Yeung, W.S.B., and Ho, P.-C. (2008b). Gene expression profiling of human peri-implantation endometria between natural and stimulated cycles. Fertility and Sterility *90*, 2152–2164.

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T.K., Vilo, J., Salumets, A., et al. (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biology *15*.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. PLoS Comput Biol *13*, e1005457.

Lucas, E.S., Dyer, N.P., Murakami, K., Hou Lee, Y., Chan, Y.-W., Grimaldi, G., Muter, J., Brighton, P.J., Moore, J.D., Patel, G., et al. (2016). Loss of Endometrial Plasticity in Recurrent Pregnancy Loss: MSC Deficiency in Recurrent Miscarriage. STEM CELLS *34*, 346–356.

Luense, L.J., Carletti, M.Z., and Christenson, L.K. (2009). Role of Dicer in female fertility. Trends in Endocrinology & Metabolism *20*, 265–272.

Machtinger, R., Laurent, L.C., and Baccarelli, A.A. (2015). Extracellular vesicles: roles in gamete maturation, fertilization and embryo implantation. Human Reproduction Update dmv055.

Madrigano, J., Baccarelli, A.A., Mittleman, M.A., Sparrow, D., Vokonas, P.S., Tarantini, L., and Schwartz, J. (2012). Aging and epigenetics: Longitudinal changes in gene-specific DNA methylation. Epigenetics *7*, 63–70.

Maekawa, R., Tamura, I., Shinagawa, M., Mihara, Y., Sato, S., Okada, M., Taketani, T., Tamura, H., and Sugino, N. (2019). Genome-wide DNA methylation analysis

revealed stable DNA methylation status during decidualization in human endometrial stromal cells. BMC Genomics *20*, 324.
Magner, L.N. (2002). A history of the life sciences (New York: M. Dekker).
Makieva, S., Giacomini, E., Ottolina, J., Sanchez, A., Papaleo, E., and Viganò, P. (2018). Inside the Endometrial Cell Signaling Subway: Mind the Gap(s). International Journal of Molecular Sciences *19*, 2477.
Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol *13*, R44.
Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research *18*, 1509–1517.
Martín, J.C., Jasper, M.J., Valbuena, D., Meseguer, M., Remohí, J., Pellicer, A., and Simón, C. (2000). Increased Adhesiveness in Cultured Endometrial-Derived Cells Is Related to the Absence of Moesin Expression1. Biology of Reproduction *63*, 1370–1376.
McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research *40*, 4288–4297.
McIntyre, L.M., Lopiano, K.K., Morse, A.M., Amin, V., Oberg, A.L., Young, L.J., and Nuzhdin, S.V. (2011). RNA-seq: technical variability and sampling. BMC Genomics *12*.
Mendel, G. (1865). Versuche uber pflanzen-hybriden. Vorgelegt in den Sitzungen.
Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Research *44*, D336–D342.
Miescher, F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen. Med.-Chem. Unters. 441–460.
Mihm, M., Gangooly, S., and Muttukrishna, S. (2011). The normal menstrual cycle in women. Animal Reproduction Science *124*, 229–236.
Min Jou, W., Haegeman, G., Ysebaert, M., and Feirs, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature *273*, 82–88.
Mirkin, S., Arslan, M., Churikov, D., Corica, A., Diaz, J.I., Williams, S., Bocca, S., and Oehninger, S. (2005). In search of candidate genes critically expressed in the human endometrium during the window of implantation. Human Reproduction *20*, 2104–2117.
Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413–420.
Montazerian, M., Yasari, F., and Aghaalikhani, N. (2018). Ovarian extracellular MicroRNAs as the potential non-invasive biomarkers: An update. Biomedicine & Pharmacotherapy *106*, 1633–1640.
Moreno, I., Codoñer, F.M., Vilella, F., Valbuena, D., Martinez-Blanch, J.F., Jimenez-Almazán, J., Alonso, R., Alamá, P., Remohí, J., Pellicer, A., et al. (2016). Evidence that the endometrial microbiota has an effect on implantation success or failure. American Journal of Obstetrics and Gynecology *215*, 684–703.
Morris, T.J., Butcher, L.M., Feber, A., Teschendorff, A.E., Chakravarthy, A.R., Wojdacz, T.K., and Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics *30*, 428–430.

Mortlock, S., Restuadi, R., Levien, R., Girling, J.E., Holdsworth-Carson, S.J., Healey, M., Zhu, Z., Qi, T., Wu, Y., Lukowski, S.W., et al. (2019). Genetic regulation of methylation in human endometrium and blood and gene targets for reproductive diseases. 19.

Murphy, C.R. (2004). Uterine receptivity and the plasma membrane transformation. Cell Research *14*, 259–267.

Nair, A.R., and Taylor, H.S. (2010). The Mechanism of Menstruation. In Amenorrhea, N.F. Santoro, and G. Neal-Perry, eds. (Totowa, NJ: Humana Press), pp. 21–34.

Nazarov, P.V., Muller, A., Kaoma, T., Nicot, N., Maximo, C., Birembaut, P., Tran, N.L., Dittmar, G., and Vallar, L. (2017). RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. BMC Genomics *18*, 443.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat Methods *12*, 453–457.

Noyer-Weidner, M., and Trautner, T.A. (1993). Methylation of DNA in Prokaryotes. Jost JP., Saluz HP. (Eds) DNA Methylation, EXS *64*.

Noyes, R.W., Hertig, A.T., and Rock, J. (1950). Dating the Endometrial Biopsy.

Noyes, R.W., Hertig, A.T., and Rock, J. (1975). Dating the endometrial biopsy. American Journal of Obstetrics and Gynecology *122*, 262–263.

Oakeley, E. (1999). DNA methylation analysis a review of current methodologies. Pharmacology & Therapeutics *84*, 389–400.

OpenStax (2016). Anatomy and Physiology, chapter 27 Anatomy and Physiology of the Female Reproductive System, OpenStax CNX, URL: https://opentextbc.ca/anatomyandphysiology/chapter/27-2-anatomy-and-physiology-of-the-female-reproductive-system/ (Accessed: 2019-05-13).

Palomino, W.A., Tayade, C., Argandoña, F., Devoto, L., Young, S.L., and Lessey, B.A. (2018). The endometria of women with endometriosis exhibit dysfunctional expression of complement regulatory proteins during the mid secretory phase. Journal of Reproductive Immunology *125*, 1–7.

Pan, Y., Bush, E.C., Toonen, J.A., Ma, Y., Solga, A.C., Sims, P.A., and Gutmann, D.H. (2017). Whole tumor RNA-sequencing and deconvolution reveal a clinically-prognostic PTEN/PI3K-regulated glioma transcriptional signature. Oncotarget *8*.

Paraskevopoulou, M.D., and Hatzigeorgiou, A.G. (2016). Analyzing MiRNA–LncRNA Interactions. In Long Non-Coding RNAs, Y. Feng, and L. Zhang, eds. (New York, NY: Springer New York), pp. 271–286.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A.G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Research *41*, W169–W173.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods *14*, 417–419.

Pauerstein, C.J., Eddy, C.A., Croxatto, H.D., Hess, R., Siler-Khodr, T.M., and Croxatto, H.B. (1978). Temporal relationships of estrogen, progesterone, and luteinizing hormone levels to ovulation in women and infrahuman primates. American Journal of Obstetrics and Gynecology *130*, 876–886.

Paul, A.B.M., Sadek, S.T., and Mahesan, A.M. (2018). The role of microRNAs in human embryo implantation: a review. Journal of Assisted Reproduction and Genetics.

Paul, A.B.M., Sadek, S.T., and Mahesan, A.M. (2019). The role of microRNAs in human embryo implantation: a review. J Assist Reprod Genet *36*, 179–187.

Perrino, C., Barabási, A.-L., Condorelli, G., Davidson, S.M., De Windt, L., Dimmeler, S., Engel, F.B., Hausenloy, D.J., Hill, J.A., Van Laake, L.W., et al. (2017). Epigenomic and transcriptomic approaches in the post-genomic era: path to novel targets for diagnosis and therapy of the ischaemic heart? Position Paper of the European Society of Cardiology Working Group on Cellular Biology of the Heart. Cardiovascular Research *113*, 725–736.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat Commun *6*, 8570.

Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L.C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics *14*, 293.

Ponnampalam, A.P., and Rogers, P.A.W. (2006). Cyclic changes and hormonal regulation of annexin IV mRNA and protein in human endometrium. MHR: Basic Science of Reproductive Medicine *12*, 661–669.

Pratt, J.P., Allen, E., Newell, Q.U., and Bland, L.J. (1929). Human ova from the uterine tubes. Journal of the American Medical Association *93*, 834–837.

Qiao, W., Quon, G., Csaszar, E., Yu, M., Morris, Q., and Zandstra, P.W. (2012). PERT: A Method for Expression Deconvolution of Human Blood Samples from Varied Microenvironmental and Developmental Conditions. PLoS Comput Biol *8*, e1002838.

Rahmioglu, N., Drong, A.W., Lockstone, H., Tapmeier, T., Hellner, K., Saare, M., Laisk-Podar, T., Dew, C., Tough, E., Nicholson, G., et al. (2017). Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. Epigenetics *12*, 897–908.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences *97*, 5237–5242.

Reed, B.G., and Carr, B.R. (2015). The Normal Menstrual Cycle and the Control of Ovulation. 20.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler – a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Research *35*, W193–W200.

Reimand, J., Arak, T., and Vilo, J. (2011). g:Profiler – a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Research *39*, W307–W315.

Revel, A., Achache, H., Stevens, J., Smith, Y., and Reich, R. (2011). MicroRNAs are associated with human embryo implantation defects. Human Reproduction *26*, 2830–2840.

Richardson, S.J., Senikas, V., and NELSONf, J.F. (1987). Follicular Depletion During the Menopausal Transition: Evidence for Accelerated Loss and Ultimate Exhaustion. *65*, 7.

Riesewijk, A. (2003). Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology. Molecular Human Reproduction *9*, 253–264.

Riesewijk, A., Martin, J., van Os, R., Horcajadas, J.A., Polman, J., Pellicer, A., Mosselman, S., and Simon, C. (2003). Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology. Molecular Human Reproduction *9*, 253–264.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl. Acids Res. *43*, e47–e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Robles, V., Valcarce, D.G., and Riesco, M.F. (2019). Non-coding RNA regulation in reproduction: Their potential use as biomarkers. Non-Coding RNA Research *4*, 54–62.

Rosner, J., and Sarao, M.S. (2019). Physiology, Female Reproduction. In StatPearls, (Treasure Island (FL): StatPearls Publishing), p.

Ruiz-Alonso, M., Blesa, D., and Simón, C. (2012). The genomics of the human endometrium. Biochimica et Biophysica Acta (BBA) – Molecular Basis of Disease *1822*, 1931–1942.

Ruiz-Alonso, M., Blesa, D., Díaz-Gimeno, P., Gómez, E., Fernández-Sánchez, M., Carranza, F., Carrera, J., Vilella, F., Pellicer, A., and Simón, C. (2013). The endometrial receptivity array for diagnosis and personalized embryo transfer as a treatment for patients with repeated implantation failure. Fertility and Sterility *100*, 818–824.

Saare, M., Modhukur, V., Suhorutshenko, M., Rajashekar, B., Rekker, K., Sõritsa, D., Karro, H., Soplepmann, P., Sõritsa, A., Lindgren, C.M., et al. (2016). The influence of menstrual cycle and endometriosis on endometrial methylome. Clinical Epigenetics *8*.

Saare, M., Krigul, K.L., Laisk-Podar, T., Ponandai-Srinivasan, S., Rahmioglu, N., Lalit Kumar, P.G., Zondervan, K., Salumets, A., and Peters, M. (2018). DNA methylation alterations—potential cause of endometriosis pathogenesis or a reflection of tissue heterogeneity?†. Biology of Reproduction *99*, 273–282.

Sampson, J.A. (1927). Metastatic or embolic endometriosis, due to the menstrual dissemination of endometrial tissue into the venous circulation. The American journal of pathology *3*, 93–110.

Sanchez, A.M., Viganò, P., Somigliana, E., Panina-Bordignon, P., Vercellini, P., and Candiani, M. (2014). The distinguishing cellular and molecular features of the endometriotic ovarian cyst: from pathophysiology to the potential endometrioma-mediated damage to the ovary. Human Reproduction Update *20*, 217–230.

Scarlett, K., Pattabiraman, V., Barnett, P., Liu, D., and Anderson, L.M. (2015). The Proangiogenic Effect of Iroquois Homeobox Transcription Factor Irx3 in Human Microvascular Endothelial Cells. J. Biol. Chem. *290*, 6303–6315.

Schelker, M., Feau, S., Du, J., Ranu, N., Klipp, E., MacBeath, G., Schoeberl, B., and Raue, A. (2017). Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. Nature Communications *8*, 2032.

Schroeder, D.I., Blair, J.D., Lott, P., Yu, H.O.K., Hong, D., Crary, F., Ashwood, P., Walker, C., Korf, I., Robinson, W.P., et al. (2013). The human placenta methylome. Proceedings of the National Academy of Sciences *110*, 6037–6042.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*, 1086–1092.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA *22*, 839–851.

Schurmann, C., Heim, K., Schillert, A., Blankenberg, S., Carstensen, M., Dörr, M., Endlich, K., Felix, S.B., Gieger, C., Grallert, H., et al. (2012). Analyzing Illumina Gene Expression Microarray Data from Different Tissues: Methodological Aspects of Data Analysis in the MetaXpress Consortium. PLoS ONE *7*, e50938.

Seitz, H. (2009). Redefining MicroRNA Targets. Current Biology *19*, 870–873.

Seo, M., Kim, K., Yoon, J., Jeong, J.Y., Lee, H.-J., Cho, S., and Kim, H. (2016). RNA-seq analysis for detecting quantitative trait-associated genes. Scientific Reports *6*.

SEQC/MAQC-III Consortium (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nature Biotechnology *32*, 903–914.

Sha, A.-G., Liu, J.-L., Jiang, X.-M., Ren, J.-Z., Ma, C.-H., Lei, W., Su, R.-W., and Yang, Z.-M. (2011). Genome-wide identification of micro-ribonucleic acids associated with human endometrial receptivity in natural and stimulated cycles by deep sequencing. Fertility and Sterility *96*, 150-155.e5.

Shen, L., and Waterland, R.A. (2007). Methods of DNA methylation analysis: Current Opinion in Clinical Nutrition and Metabolic Care *10*, 576–581.

Sheng, Q., Vickers, K., Zhao, S., Wang, J., Samuels, D.C., Koues, O., Shyr, Y., and Guo, Y. (2016). Multi-perspective quality control of Illumina RNA sequencing data analysis. Briefings in Functional Genomics elw035.

Shen-Orr, S.S., and Gaujoux, R. (2013). Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Current Opinion in Immunology *25*, 571–578.

Sherman, B.M., and Korenman, S.G. (1975). Hormonal Characteristics of the Human Menstrual Cycle throughout Reproductive Life. The Journal of Clinical Investigation *55*, 8.

Shimojo, M. (2008). Huntingtin Regulates RE1-silencing Transcription Factor/Neuron-restrictive Silencer Factor (REST/NRSF) Nuclear Trafficking Indirectly through a Complex with REST/NRSF-interacting LIM Domain Protein (RILP) and Dynactin p150 ^{Glued}. J. Biol. Chem. *283*, 34880–34886.

Sigurgeirsson, B., Åmark, H., Jemt, A., Ujvari, D., Westgren, M., Lundeberg, J., and Gidlöf, S. (2016). Comprehensive RNA sequencing of healthy human endometrium at two time points of the menstrual cycle†. Biology of Reproduction.

Singh, M., Chaudhry, P., and Asselin, E. (2011). Bridging endometrial receptivity and implantation: network of hormones, cytokines, and growth factors. Journal of Endocrinology *210*, 5–14.

Strell, C., Hilscher, M.M., Laxman, N., Svedlund, J., Wu, C., Yokota, C., and Nilsson, M. (2019). Placing RNA in context and space – methods for spatially resolved transcriptomics. The FEBS Journal *286*, 1468–1481.

Strmsek, Z., and Kunej, T. (2015). MicroRNA Silencing by DNA Methylation in Human Cancer: a Literature Analysis. NcRNA *1*, 44–52.

Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Aneichyk, T. (2018). Comprehensive evaluation of computational cell-type quantification methods for immuno-oncology (Bioinformatics).

Tapia, A., Gangi, L.M., Zegers-Hochschild, F., Balmaceda, J., Pommer, R., Trejo, L., Pacheco, I.M., Salvatierra, A.M., Henriquez, S., Quezada, M., et al. (2007). Differences in the endometrial transcript profile during the receptive period between women who were refractory to implantation and those who achieved pregnancy. Human Reproduction *23*, 340–351.

Teder, H., Koel, M., Paluoja, P., Jatsenko, T., Rekker, K., Laisk-Podar, T., Kukuškina, V., Velthut-Meikas, A., Fjodorova, O., Peters, M., et al. (2018). TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting. Npj Genomic Medicine *3*, 34.

Tervise Arengu Instituut (2015). In vitro fertilization (IVF) 2013-2014, URL: https://www.tai.ee/en/health-data/health-statistics-and-health-research-database/latest-updates/3869-in-vitro-fertilization-2013-2014 (Accessed: 2019-04-13).

Tervise Arengu Instituut (2016). IVF procedures 2015, URL: https://www.tai.ee/en/health-data/health-statistics-and-health-research-database/latest-updates/3966-ivf-procedures-2015 (Accessed: 2019-04-13).

Teschendorff, A.E., and Zheng, S.C. (2017). Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. Epigenomics *9*, 757–768.

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics *29*, 189–196.

Tesfaye, D., Gebremedhn, S., Salilew-Wondim, D., Hailay, T., Hoelker, M., Grosse-Brinkhaus, C., and Schellander, K. (2018). MicroRNAs: tiny molecules with a significant role in mammalian follicular and oocyte development. Reproduction *155*, R121–R135.

Thie, M., Harrach-Ruprecht, B., Sauer, H., Fuchs, P., Albers, A., and Denker, H.-W. (1995). Cell adhesion to the apical pole of epithelium: a function of ce 11 polarity. European Journal of Cell Biology *66*, 180–191.

Touleimat, N., and Tost, J. (2012). Complete pipeline for Infinium ® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics *4*, 325–341.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511–515.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562–578.

Trzybulska, D., Vergadi, E., and Tsatsanis, C. (2018). miRNA and other non-coding RNAs as promising diagnostic markers. eJIFCC *29*, 221–226.

Tuerk, A., Wiktorin, G., and Güler, S. (2017). Mixture models reveal multiple positional bias types in RNA-Seq data and lead to accurate transcript concentration estimates. PLoS Comput Biol *13*, e1005515.

UniProtKB (2020a). ADCYAP1R1,UniProtKB – P41586 (PACR_HUMAN), URL: https://www.uniprot.org/uniprot/P41586#function (Acessed: 6.03.2020).

UniProtKB (2020b). GJB6, UniProtKB – O95452 (CXB6_HUMAN), URL: https://www.uniprot.org/uniprot/O95452 (Accessed: 6.03.2020).

UniProtKB (2020c). EPPIN, UniProtKB – O95925 (EPPI_HUMAN), URL: https://www.uniprot.org/uniprot/O95925#function (Accessed: 6.03.2020).

Vallejo, G., Maschi, D., Mestre-Citrinovitz, A.C., Aiba, K., Maronna, R., Yohai, V., Ko, M.S.H., Beato, M., and Saragüeta, P. (2010). Changes in global gene expression during in vitro decidualization of rat endometrial stromal cells. Journal of Cellular Physiology *222*, 127–137.

Vasudevan, S. (2012). Posttranscriptional Upregulation by MicroRNAs: Posttranscriptional Upregulation by MicroRNAs. WIREs RNA *3*, 311–330.

Vilella, F., Moreno-Moya, J.M., Balaguer, N., Grasso, A., Herrero, M., Martinez, S., Marcilla, A., and Simon, C. (2015). Hsa-miR-30d, secreted by the human endometrium, is taken up by the pre-implantation embryo and might modify its transcriptome. Development *142*, 3210–3221.

Von Baer, K.E. (1827). De ovi mammalium et hominis genesi. Vossil.

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L., and Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics *16*, 49.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics *26*, 136–138.

Wang, N., Hoffman, E.P., Chen, L., Chen, L., Zhang, Z., Liu, C., Yu, G., Herrington, D.M., Clarke, R., and Wang, Y. (2016). Mathematical modelling of transcriptional heterogeneity identifies novel markers and subpopulations in complex tissues. Sci Rep *6*, 18909.

Wang, T., Guan, W., Lin, J., Boutaoui, N., Canino, G., Luo, J., Celedón, J.C., and Chen, W. (2015). A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. Epigenetics *10*, 662–669.

Wang, Z., Wu, X., and Wang, Y. (2018). A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. BMC Bioinformatics *19*, 115.

Warton, K., Foster, N.C., Gold, W.A., and Stanley, K.K. (2004). A novel gene family induced by acute inflammation in endothelial cells. Gene *342*, 85–95.

Watson, J.D., and Crick, F.H.C. (1953). The Structure Of DNA. Cold Spring Harbor Symposia on Quantitative Biology *18*, 123–131.

Wilcox, A.J., Baird, D.D., and Weinberg, C.R. (1999). Time of Implantation of the Conceptus and Loss of Pregnancy. N Engl J Med 4.

Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

Williams, A.G., Thomas, S., Wyman, S.K., and Holloway, A.K. (2014). RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis: RNA-seq Data: Experimental Design and Analysis. In Current Protocols in Human Genetics, J.L. Haines, B.R. Korf, C.C. Morton, C.E. Seidman, J.G. Seidman, and

D.R. Smith, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 11.13.1–11.13.20.

Wreczycka, K., Gosdschan, A., Yusuf, D., Grüning, B., Assenov, Y., and Akalin, A. (2017). Strategies for analyzing bisulfite sequencing data. Journal of Biotechnology *261*, 105–115.

Wu, R. (1994). Development of the primer-extension approach: a key role in DNA sequencing. Trends in Biochemical Sciences *19*, 429–433.

Xu, C., Zhang, J.-G., Lin, D., Zhang, L., Shen, H., and Deng, H.-W. (2017). A Systemic Analysis of Transcriptomic and Epigenomic Data To Reveal Regulation Patterns for Complex Disease. G3 *7*, 2271–2279.

Yamagata, Y., Asada, H., Tamura, I., Lee, L., Maekawa, R., Taniguchi, K., Taketani, T., Matsuoka, A., Tamura, H., and Sugino, N. (2009). DNA methyltransferase expression in the human endometrium: down-regulation by progesterone and estrogen. Human Reproduction *24*, 1126–1132.

Yang, J.X., Rastetter, R.H., and Wilhelm, D. (2016). Non-coding RNAs: An Introduction. In Non-Coding RNA and the Reproductive System, D. Wilhelm, and P. Bernard, eds. (Dordrecht: Springer Netherlands), pp. 13–32.

Zaitsev, K., Bambouskova, M., Swain, A., and Artyomov, M.N. (2019). Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. Nat Commun *10*, 2209.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Research *46*, D754–D761.

Zhang, B., Xing, X., Li, J., Lowdon, R.F., Zhou, Y., Lin, N., Zhang, B., Sundaram, V., Chiappinelli, K.B., Hagemann, I.S., et al. (2014a). Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. 21.

Zhang, D., Sun, C., Ma, C., Dai, H., and Zhang, W. (2012b). Data Mining of Spatial–Temporal Expression of Genes in the Human Endometrium During the Window of Implantation. Reprod Sci *19*, 1085–1098.

Zhang, R., Yu, C., Wu, R., Zhang, L., Zhu, L., Xu, A., and Wang, C. (2016). RNA-Seq-Based Transcriptome Analysis of Changes in Gene Expression Linked to Human Pregnancy Outcome After In Vitro Fertilization–Embryo Transfer. Reproductive Sciences *23*, 134–145.

Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., et al. (2014b). A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. PLoS ONE *9*, e103207.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE *9*, e78644.

Zhou, J., Dsupin, B.A., Giudice, L.C., and Bondy, C.A. (1994). Insulin-like growth factor system gene expression in human endometrium during the menstrual cycle. Journal of Clinical Endocrinology and Metabolism *79*, 1723–1734.

Zhu, T., Zheng, S.C., Paul, D.S., Horvath, S., and Teschendorff, A.E. (2018). Cell and tissue type independent age-associated DNA methylation changes are not rare but common. Aging *10*, 3541–3557.

Zuo, T., Tycko, B., Liu, T.-M., Lin, H.-J.L., and Huang, T.H.-M. (2009). Methods in DNA methylation profiling. Epigenomics *1*, 331–345.

ACKNOWLEDGMENTS

I would like to thank my dear supervisor Dr. Reedik Mägi. Reedik, thank you for taking me under your wing and supporting all these years. I would not reach this point in my studies without your care, understanding, and help. I cannot express with words how grateful and happy I am to work under your guidance. You are the best supervisor I could wish for.

My sincere gratitude goes to my supervisor Prof. Andres Salumets. Andres, thank you for going on with my desire to work with RNA-seq data, giving me the Ph.D. position and introducing to me the world of female reproductive studies.

I am very grateful to my supervisor Prof. Andres Metspalu. Andres, thank you for welcoming me in to Estonian Genome Center family, for giving me opportunities to better my knowledge abroad, for the feeling of safety, and your support. You are like a guardian angel of Estonian Genome Center.

I would like to thank my dear supervisor Dr. Triin Laisk. Triin, thank you for your tremendous help with this thesis. I enjoy working with you on the projects and you are my lifesaver when it comes to reading my mind and converting my fuzzy explanations to human-understandable language. Thank you, and I hope we continue working together.

I would like to thank my colleagues for the warm and welcoming working atmosphere, help and support, especially Silva Kasela, for help and ideas regarding data integration and R.

I would like to express my gratitude to Prof. Maido Remm for reviewing my thesis and suggesting corrections, for giving me MSc psition and welcoming me in to bioinformatics team despite my total lack of programming knowledge. Thank you, I am very happy I was working under your guidance.

My sincere gratitude goes to my previous supervisors: Lauris Kaplinski, during MSc studies, for teaching me Perl and showing that the black terminal window is not scary at all; Viia Kõiv and Andres Mäe, during BSc studies, for introducing me the magic world of microbiology.

I would like to thank my friends for their support. I am happy you are always there for me, even if we are hundreds of kilometers apart. I wish, we see each other more often.

I am grateful to my parents, Larisa Chernokozova and Roman Chernokozov. My dear mum and dad, thank you for giving me the opportunity to study what I wanted and for your unlimited support and love all these years (Мам, пап, спасибо вам за то, что предоставили мне возможность учиться там, где я хотела и за вашу нескончаемую поддержку и любовь все эти годы. Люблю вас!).

I would like to thank my dear husband Anton Kukushkin. Thank you for installing my first ubuntu; for showing me that programming is interesting and makes (work)life easier; for giving me the confidence in learning programming and trying myself in bioinformatics; for our cat; for discussing scientific

questions, my work-related tasks and problems, and many other interesting things in the kitchen during tea; for your support, love, and understanding. I am very lucky to have you by my side. Also, I thank our cat for her warm and stress-releasing purr-purrs.

Finally, I would like to thank all these people I met during university studies and worked with, for sharing their knowledge and experience, so I learned a lot of interesting things.

PUBLICATIONS

CURRICULUM VITAE

Name: Viktorija Kukushkina
Date of birth: March 31, 1988
Contact: University of Tartu, Institute of Genomics, Riia 23b, 51010, Tartu, Estonia
E-mail: viktorija.kukuskina@ut.ee

Education:

2015–present	PhD studies in Molecular and cell biology; Faculty of Science and Technology, University of Tartu, Estonia
2011–2013	Master's studies in Gene technology; Faculty of Science and Technology, University of Tartu, Estonia
2007–2011	Bachelor's studies in Biology; Faculty of Science and Technology, University of Tartu, Estonia
1995–2007	Võru Vene Gymnasium (honored with gold medal)

Professional employment:

2018–…	University of Tartu, Institute of Genomics, Specialist
2016–2018	University of Tartu, Estonian Genome Center, Specialist
2015–2016	Competence Centre on Health Technologies, Specialist
2014–2015	University of Tartu, Faculty of Medicine, Institute of Biomedicine, Project specialist
2011–2015	Estonian Biocentre, Programmer
2009–2011	Lab Assistant, department of Genetic technology, Institute of Molecular and Cell biology, University of Tartu, Estonia

Fields of research:
CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics

Supervised dissertations:
Ustina Põnova, Master's Degree, 2017, (sup) Tarmo Annilo; Viktorija Kukuškina, Meditsiiniliselt oluliste geenivariantide tõlgendamine RNA ekspressiooniandmete abil (The interpretation of clinically relevant genetic variants through RNA-Seq expression data analysis), University of Tartu, Estonian Genome Center

Publications:

2019

Puurand T., **Kukuškina V.,** Pajuste F-D., Remm M. (2019). AluMine: alignment-free method for the discovery of polymorphic Alu element insertions. Mobile DNA , 10, 31−31.10.1186/s13100-019-0174-3.

Linner RK., Biroli P., Kong E., Meddens FW., Wedow R., Fontana MA., Lebreton M., Tino SP., Abdellaoui A., Hammerschlag AR., Nivard MG., Okbay A., Rietveld CA., …, **Kukushkina V.,** et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics, 51 (2), 245−261.10.1038/s41588-018-0309-3.

Timmers PR., Mounier N., Lall K., Fischer K., Ning Z., Feng X., Bretherick AD., Clark DW., Agbessi M., Ahsan H., Alves I., Andiappan A., Awadalla P., Battle A., Bonder MJ., …, **Kukushkina V**., et al. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. eLife, 8.10.7554/eLife.39856.

Porcu E., Rueger S., Lepik K., Agbessi M., Ahsan H., Alves I., Andiappan A., Arindrarto W., Awadalla P., Battle A., Beutner F., Bonder MJ., Boomsma D., Christiansen M., …, **Kukushkina V**., et al. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nature Communications, 10, ARTN 3300.10.1038/s41467-019-10936-0.

2018

Suhorutshenko M., **Kukushkina V**., Velthut-Meikas A., Altmäe S., Peters M., Mägi R., Krjutškov K., Koel M., Codoñer F., Martinez-Blanch J., Vilella F., Simón C., Salumets A., Laisk T. (2018). Endometrial receptivity revisited: endometrial transcriptome adjusted for tissue cellular heterogeneity. Human Reproduction, 33 (11), 2074−2086.10.1093/humrep/dey301.

Laisk T., **Kukuškina V**., Palmer D., Laber S., Chen C-Y., Ferreira T., Rahmioglu N., Zondervan K., Becker C., Smoller JW., Lippincott M., Salumets A., Granne I., Seminara S., Neale B., Mägi R., Lindgren CM. (2018). Large scale meta-analysis highlights the hypothalamic-pituitary-gonadal (HPG) axis in the genetic regulation of menstrual cycle length. Human Molecular Genetics, 27 (24), 4323−4332.10.1093/hmg/ddy317.

Teder H., Koel M., Paluoja P., Jatsenko T., Rekker K., Laisk-Podar T., **Kukuškina V**., Velthut-Meikas A., Fjodorova O., Peters M., Kere J., Salumets A., Palta P., Krjutškov K. (2018). TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting. npj Genomic Medicine, 3, ARTN 34.10.1038/s41525-018-0072-5.

Pervjakova N., **Kukushkina V**., Haller T., Kasela S., Joensuu A., Kristiansson K., Annilo T., Perola M., Salomaa V., Jousilahti P., Metspalu A., Magi R. (2018). Genome-wide analysis of nuclear magnetic resonance metabolites revealed parent-of-origin effect on triglycerides in medium very low-density lipoprotein in PTPRD gene. Biomarkers in medicine, 12 (5), 439−446. 10.2217/bmm-2018-0020.

2017

Rekker K., Saare M., Eriste E., Tasa T., **Kukuškina V**., Roost AM., Anderson K., Samuel K., Karro H., Salumets A., Peters M., (2017). High-throughput mRNA sequencing of stromal cells from endometriomas and endometrium. Reproduction, 154 (1), 93−100.10.1530/REP-17-0092.

Lepik K., Annilo T., **Kukuškina V**., eQTLGen Consortium, Kisand K., Kutalik Z., Peterson P., Peterson H. (2017). C-reactive protein upregulates the whole blood expression of CD59 – an integrative analysis. PLoS Computational Biology, 13 (9), e1005766−e1005766.10.1371/journal.pcbi.1005766.

Kukushkina V., Modhukur V., Suhorutšenko M., Peters M., Mägi R., Rahmioglu N., Velthut-Meikas A., Altmäe S., Esteban FJ., Vilo J., Zondervan K., Salumets A., Laisk-Podar T. (2017). DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Scientific Reports, 7 (1), 3916−3916.10.1038/s41598-017-03682-0.

Altmäe S., Koel M., Võsa U., Adler P., Suhorutšenko M., Laisk-Podar T., **Kukushkina V**., Saare M., Velthut-Meikas A., Krjutškov K., Aghajanova L., Lalitkumar PG., Gemzell-Danielsson K., Giudice L., Simón C., Salumets A. (2017). Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers. Scientific Reports, 7 (1), 10077−10077.10.1038/s41598-017-10098-3.

Suhorutshenko M., **Kukushkina V**., Velthut-Meikas A., Altmae S., Maire P., Magi R., Codoner F., Vilella F., Simon C., Salumets A., Laisk-Podar T. (2017). Transcriptomic profile of early and mid-secretory endometrium revealed by RNA-seq: a paired sample cohort study. FEBS Journal, 284, 134−134.

Awards and stipends:

2019 Scholarship of the Graduate School in Biomedicine and Biotechnology stipend, Tartu, Estonia (Estonian Society of Human Genetics conference, Estonia)

2019 Scholarship of the Graduate School in Biomedicine and Biotechnology stipend, Tartu, Estonia (European Society of Human Genetics conference, Sweeden)

2018 Scholarship of the Graduate School in Biomedicine and Biotechnology stipend, Tartu, Estonia (GeneForum, Estonia)

2017 Scholarship of the Graduate School in Biomedicine and Biotechnology stipend, Tartu, Estonia (GeneForum, Estonia)

2016 Kirstjan Jaak Scholarship for foreign visits stipend, Ministry of Education and Research, Estonia (Collaboration with Lude Franke group, UMCG, Netherlands)

2015 Kirstjan Jaak Scholarship for foreign visits stipend, Ministry of Education and Research, Estonia (NGS Course: RNA-seq data analysis, LUMC, Netherlands)

ELULOOKIRJELDUS

Nimi: Viktorija Kukuškina
Sünniaeg: 31. märts 1988
Aadress: Genoomika Instituut, Tartu Ülikool, Riia 23b, 51010, Tartu, Eesti
E-post: viktorija.kukuskina@ut.ee

Haridus:

2015–praegu	Doktoriõpe, molekulaar- ja rakubioloogia, molekulaar- ja rakubioloogia instituut, loodus ja tehnoloogia teaduskond, Tartu Ülikool, Tartu, Eesti
2011–2013	Magistriõpe, geeni tehnoloogia, molekulaar- ja rakubioloogia instituut, loodus- ja tehnoloogia teaduskond, Tartu Ülikool, Tartu, Eesti
2007–2011	Bakalaureuseõpe, bioloogia, molekulaar- ja rakubioloogia instituut, loodus- ja tehnoloogia teaduskond, Tatru Ülikool, Tartu, Eesti
1995–2007	Võru Vene Gümnaasium (kuldmedal)

Teenuskäik:

2018–…	Tartu Ülikool, Tartu Ülikooli genoomika instituut, spetsialist
2016–2018	Tartu Ülikool, Tartu Ülikooli Eesti Geenivaramu, spetsialist
2015–2016	Tervisetehnoloogiate Arenduskeskus AS (endine Reproduktiivmeditsiini TAK AS) , spetsialist
2014–2015	Tartu Ülikool, Arstiteaduskond, Biomeditsiini instituut, projekti spetsialist
2011–2015	Eesti Biokeskus, Programmeerija
2009–2011	Labori assistent, Geneetika õppetool, Molekulaar- ja rakubioloogia instituut, Tartu Ülikool, Tartu, Eesti

Teadustöö põhisuunad:
CERS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Juhendatud väitekirjad:
– Ustina Põnova, magistrikraad, 2017, (juh) Tarmo Annilo;
  Viktorija Kukuškina, Meditsiiniliselt oluliste geenivariantide tõlgendamine RNA ekspressiooniandmete abil, Tartu Ülikool, Tartu Ülikooli Eesti Geenivaramu.

Teaduspublikatsioonid:

2019

Puurand T., **Kukuškina V.,** Pajuste F-D., Remm M. (2019). AluMine: alignment-free method for the discovery of polymorphic Alu element insertions. Mobile DNA , 10, 31−31.10.1186/s13100-019-0174-3.

Linner RK., Biroli P., Kong E., Meddens FW., Wedow R., Fontana MA., Lebreton M., Tino SP., Abdellaoui A., Hammerschlag AR., Nivard MG., Okbay A., Rietveld CA., …, **Kukushkina V.,** et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics, 51 (2), 245−261.10.1038/s41588-018-0309-3.

Timmers PR., Mounier N., Lall K., Fischer K., Ning Z., Feng X., Bretherick AD., Clark DW., Agbessi M., Ahsan H., Alves I., Andiappan A., Awadalla P., Battle A., Bonder MJ., …, **Kukushkina V**., et al. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. eLife, 8.10.7554/eLife.39856.

Porcu E., Rueger S., Lepik K., Agbessi M., Ahsan H., Alves I., Andiappan A., Arindrarto W., Awadalla P., Battle A., Beutner F., Bonder MJ., Boomsma D., Christiansen M., …, **Kukushkina V**., et al. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nature Communications, 10, ARTN 3300.10.1038/s41467-019-10936-0.

2018

Suhorutshenko M., **Kukushkina V**., Velthut-Meikas A., Altmäe S., Peters M., Mägi R., Krjutškov K., Koel M., Codoñer F., Martinez-Blanch J., Vilella F., Simón C., Salumets A., Laisk T. (2018). Endometrial receptivity revisited: endometrial transcriptome adjusted for tissue cellular heterogeneity. Human Reproduction, 33 (11), 2074−2086.10.1093/humrep/dey301.

Laisk T., **Kukuškina V**., Palmer D., Laber S., Chen C-Y., Ferreira T., Rahmioglu N., Zondervan K., Becker C., Smoller JW., Lippincott M., Salumets A., Granne I., Seminara S., Neale B., Mägi R., Lindgren CM. (2018). Large scale meta-analysis highlights the hypothalamic-pituitary-gonadal (HPG) axis in the genetic regulation of menstrual cycle length. Human Molecular Genetics, 27 (24), 4323−4332.10.1093/hmg/ddy317.

Teder H., Koel M., Paluoja P., Jatsenko T., Rekker K., Laisk-Podar T., **Kukuškina V**., Velthut-Meikas A., Fjodorova O., Peters M., Kere J., Salumets A., Palta P., Krjutškov K. (2018). TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting. npj Genomic Medicine, 3, ARTN 34.10.1038/s41525-018-0072-5.

Pervjakova N., **Kukushkina V**., Haller T., Kasela S., Joensuu A., Kristiansson K., Annilo T., Perola M., Salomaa V., Jousilahti P., Metspalu A., Magi R. (2018). Genome-wide analysis of nuclear magnetic resonance metabolites revealed parent-of-origin effect on triglycerides in medium very low-density

lipoprotein in PTPRD gene. Biomarkers in medicine, 12 (5), 439−446. 10.2217/bmm-2018-0020.

2017

Rekker K., Saare M., Eriste E., Tasa T., **Kukuškina V**., Roost AM., Anderson K., Samuel K., Karro H., Salumets A., Peters M., (2017). High-throughput mRNA sequencing of stromal cells from endometriomas and endometrium. Reproduction, 154 (1), 93−100.10.1530/REP-17-0092.

Lepik K., Annilo T., **Kukuškina V**., eQTLGen Consortium, Kisand K., Kutalik Z., Peterson P., Peterson H. (2017). C-reactive protein upregulates the whole blood expression of CD59 – an integrative analysis. PLoS Computational Biology, 13 (9), e1005766−e1005766.10.1371/journal.pcbi.1005766.

Kukushkina V., Modhukur V., Suhorutšenko M., Peters M., Mägi R., Rahmioglu N., Velthut-Meikas A., Altmäe S., Esteban FJ., Vilo J., Zondervan K., Salumets A., Laisk-Podar T. (2017). DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Scientific Reports, 7 (1), 3916−3916.10.1038/s41598-017-03682-0.

Altmäe S., Koel M., Võsa U., Adler P., Suhorutšenko M., Laisk-Podar T., **Kukushkina V**., Saare M., Velthut-Meikas A., Krjutškov K., Aghajanova L., Lalitkumar PG., Gemzell-Danielsson K., Giudice L., Simón C., Salumets A. (2017). Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers. Scientific Reports, 7 (1), 10077−10077.10.1038/s41598-017-10098-3.

Suhorutshenko M., **Kukushkina V**., Velthut-Meikas A., Altmae S., Maire P., Magi R., Codoner F., Vilella F., Simon C., Salumets A., Laisk-Podar T. (2017). Transcriptomic profile of early and mid-secretory endometrium revealed by RNA-seq: a paired sample cohort study. FEBS Journal, 284, 134−134.

Stipendiumid:

2019 Biomeditsiini ja Biotehnoloogia doktorikooli stipendium, Tartu, Eesti (Eesti InimeseGeneetika Ühingu konverents, Eesti)

2019 Biomeditsiini ja Biotehnoloogia doktorikooli stipendium, Tartu, Eesti (European Society of Human Genetics konverents, Rootsi)

2018 Biomeditsiini ja Biotehnoloogia doktorikooli stipendium, Tartu, Eesti (GeeniFoorum, Eesti)

2017 Biomeditsiini ja Biotehnoloogia doktorikooli stipendium, Tartu, Eesti (GeeniFoorum, Eesti)

2016 Kirstjan Jaagu välissõidu stipendium, teadusministeerium, Eesti (Koostöö Lude Franke grupiga, UMCG, Holland)

2015 Kirstjan Jaagu välissõidu stipendium, teadusministeerium, Eesti (NGS Course: RNA-seq data analysis, LUMC, Holland)

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.

62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.
63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.

83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.
84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli**. The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.

104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.
106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.

125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.
126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.

146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.

165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.

185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.

205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.

224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.
225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Usküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.

245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.
246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreonemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.

263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.
265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p

282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.
284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.
285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.

301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.
303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.
304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.

321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.
324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.
325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik.** Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae.* Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak.** Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak.** Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar.** Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo.** Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova.** Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare.** The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.

340. **Märt Roosaare.** *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova.** The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas.** Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.
343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.
344. **Teivi Laurimäe.** Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko.** Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand.** Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha.* Tartu, 2018, 148 p.
347. **Andres Ainelo.** Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja.** Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin.** Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver.** Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly.** The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.

360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun.** Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.
362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.
364. **Anne-Mai Ilumäe.** Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte.** Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts.** The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.