

BARIŞ YAŞAR

Advanced chromosomal testing tools
for embryo quality and fetal health



**Karolinska
Institutet**



BARIŞ YAŞAR

Advanced chromosomal testing tools
for embryo quality and fetal health



UNIVERSITY OF TARTU
Press

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

The doctoral studies included in this thesis has been conducted as a cooperation between the University of Tartu and Karolinska Institutet within the Marie Skłodowska-Curie Actions European Joint Degree program MATER.

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on March 16, 2026 by the Council of the Institute of Molecular Cell Biology, University of Tartu, Estonia, and the degree of Doctor of Philosophy in Medical Science on February 16, 2026 by the Dissertation Committee of Karolinska Institutet, Sweden.

Supervisors: University of Tartu:

Main supervisor: Professor Ants Kurg, PhD

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

Co-supervisor: Associate Professor Tõnis Org, PhD

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

Centre for Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Tartu, Estonia

Karolinska Institutet:

Main supervisor: Professor Juha Kere, MD, PhD

Department of Medicine Huddinge, Karolinska Institutet, Huddinge, Sweden

Folkhälsan Research Center, Helsinki, Finland

Stem Cells and Metabolism Research Program, University of Helsinki, Helsinki, Finland

Co-supervisor: Docent Shintaro Katayama, PhD

Department of Medicine Huddinge, Karolinska Institutet, Huddinge, Sweden

Folkhälsan Research Center, Helsinki, Finland

Stem Cells and Metabolism Research Program, University of Helsinki, Helsinki, Finland

Reviewer (University of Tartu):

Associate Professor Tambet Tõnissoo, PhD

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

Examination Board (Karolinska Institutet):

Associate Professor Jan-Bernd Stukenborg, PhD

NORDFERTIL Research Lab Uppsala, Department of Organismal Biology, Uppsala

University, Uppsala, Sweden

NORDFERTIL Research Lab Stockholm, Childhood Cancer Research Unit,

Department of Women's and Children's Health, Karolinska Institutet, Solna, Sweden

Professor Catharina Larsson, MD, PhD

Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden

Associate Professor Sergey Rodin, PhD

Cardio-Thoracic Translational Medicine Laboratory, Department of Surgical Sciences,

Uppsala University, Uppsala, Sweden

Opponent: Amélie Bonnet-Garnier, PhD

Université Paris-Saclay, UVSQ, INRAE, BREED, Jouy-en-Josas, France

Ecole Nationale Vétérinaire d'Alfort, BREED, Maisons-Alfort, France

Commencement: Auditorium 105, Riia 23B, Tartu, Estonia on April 29, 2026 at 13:15 (EET) / 12:15 (CET)

The publication of this dissertation is granted by the University of Tartu.

The research presented in this thesis was funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813707

ISSN 1024-6479 (print)

ISBN 978-9908-57-169-0 (print)

ISSN 2806-2140 (pdf)

ISBN 978-9908-57-170-6 (pdf)

The logo for the MATER program, featuring a stylized orange 'M' followed by the word 'MATER' in orange capital letters.

Copyright: Barış Yaşar, 2026

University of Tartu Press

www.tyk.ee



To my parents

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS	10
ABBREVIATIONS	11
1 INTRODUCTION.....	13
2 REVIEW OF THE LITERATURE.....	14
2.1 Foundations of reproductive biology.....	14
2.2 Preimplantation embryo development.....	15
2.3 Assisted reproductive technologies	17
2.4 Bovine genome and its evolution	18
2.5 RNA sequencing methods to study early embryo development.....	20
2.6 Embryonic genome activation and transcriptional studies in mammals	25
2.6.1 Mouse	26
2.6.2 Human	27
2.6.3 Bovine	29
2.6.4 Expression of repetitive DNA sequences in preimplantation embryos	32
2.7 Epigenetic regulation in preimplantation embryos.....	33
2.7.1 Mouse	34
2.7.2 Drosophila and Zebrafish	35
2.7.3 Human	35
2.7.4 Bovine	37
2.8 Roles of PRD-like homeobox genes in preimplantation embryo development	38
2.8.1 Homeobox genes	38
2.8.2 Classification of homeobox genes.....	39
2.8.3 Classification of PRD and PRD-like class	40
2.8.4 PRD-like class transcription factors from an evolutionary perspective.....	40
2.8.5 <i>DUX</i> family genes	41
2.8.5.1 History of <i>DUX4</i>	43
2.8.5.2 <i>DUX4</i> mRNAs and proteins in early human embryos	46
2.8.5.3 <i>DUX4</i> and repeat elements.....	46
2.8.5.4 <i>DUX4</i> and nonsense-mediated mRNA decay	47
2.8.5.5 Toxicity of <i>DUX4</i>	48
2.8.5.6 The effect of <i>DUX4</i> on translation	49
2.8.5.7 C-terminal domain of <i>DUX4</i>	50
2.8.5.8 Further characterization of <i>DUX4</i>	51
2.8.5.9 Human <i>DUX4</i> homologue mouse <i>Dux</i>	51
2.8.5.10 The role of <i>Dux</i> and its timely repression in mouse embryos	53

2.9	Roles of PRDL TFs in early embryo development	55
2.9.1	Identification of PRDL TFs in early embryo development	55
2.9.2	Molecular cloning of PRDL TF homeobox genes.....	56
2.9.3	Characterization of PRDL TF homeobox genes.....	57
2.9.3.1	<i>LEUTX</i>	57
2.9.3.2	<i>TPRX</i> genes	58
2.9.3.3	<i>DUXA</i> and <i>DUXB</i>	59
2.9.3.4	<i>ARGFX</i>	60
2.10	Cell culture models for studying early embryo development.....	61
3	AIMS	63
4	MATERIALS AND METHODS	64
4.1	Bovine IVF procedures and embryo culture (Study I, II, III, and IV) ..	64
4.2	Ethical considerations (Study I, II, III, and IV).....	64
4.3	Gene prediction (Study I and III)	64
4.4	Designing PCR primers (Study I and III).....	65
4.5	Preparation of cDNA library for molecular cloning (Study I and III)	65
4.6	PCR for molecular cloning (Study I and III).....	67
4.7	Generation of the <i>TPRX3</i> overexpression plasmid (Study I)	68
4.8	<i>TPRX3</i> overexpression in bovine fibroblasts for RNA-seq (Study I)...	68
4.9	Reanalysis of whole-genome sequencing datasets (Study I).....	69
4.10	Preparation of cDNA library for RNA-seq (Study I, II and IV).....	69
4.11	RNA-seq analysis (Study I, II, and IV)	70
4.12	Analysis of <i>DUXC</i> genomic organization across cattle breeds (Study III).....	72
4.13	Reanalysis of publicly available RNA-seq datasets (Study III)	73
5	RESULTS AND DISCUSSION	74
5.1	Study I: Molecular cloning of bovine IVF embryo-derived PRD-like homeobox genes	74
5.1.1	Experimental validation of the expression of seven PRD-like gene during early bovine development.....	74
5.1.2	Identification of SNVs in PRD-like homeobox transcription factor genes.....	76
5.1.3	Functional domains found in PRD-like proteins	76
5.1.4	Ectopic expression of <i>TPRX3</i> confirms its transcription activator role.....	77
5.2	Study II: Identification of embryonic genome activation regulators through transcriptome profiling of bovine embryo developmental stages	78
5.2.1	5'-end mRNA profiling of bovine oocytes and preimplantation embryos	78

5.2.2	Changes of poly(A)-tailed RNAs and the percentages of aligned reads assigned to each genomic feature	79
5.2.3	Differential expression of TFEs in consecutive bovine developmental stages.....	80
5.2.4	Temporal expression of PRD-like homeobox genes	82
5.2.5	Identification of embryonic stage-specific gene expression.....	83
5.2.6	Putative regulators of bovine embryonic genome activation ...	84
5.3	Study III: Characterization of bovine <i>DUXC</i>	84
5.3.1	<i>DUXC</i> is a putative binder of the promoters of the genes activated at the 16-cell stage.....	84
5.3.2	Molecular cDNA cloning of <i>DUXC</i> from the 8-cell stage bovine embryos	85
5.3.3	Clearance of <i>DUXC</i> mRNA depends on minor bovine EGA...	86
5.3.4	Genomic organization of bovine <i>DUXC</i> locus	87
5.3.5	Identifying the repeat unit origin of <i>DUXC</i> mRNA	88
5.3.6	Genomic organization of the <i>DUXC</i> locus across eight bovine breeds.....	90
5.4	Study IV: Effect of oxygen levels on bovine early embryo development	90
5.4.1	Blastocyst formation rates of embryos grown under normoxia, hypoxia and ultrahypoxia	90
5.4.2	Transcriptome profiling of embryos grown under normoxia, hypoxia and ultrahypoxia	91
6	CONCLUSIONS.....	93
	SUMMARY IN ESTONIAN	94
	REFERENCES.....	96
	ACKNOWLEDGEMENTS	125
	PUBLICATIONS	127
	CURRICULUM VITAE	222
	ELULOOKIRJELDUS.....	224

LIST OF ORIGINAL PUBLICATIONS

- I. Yaşar, B.**, Boskovic, N., Ivask, M., Weltner, J., Jouhilahti, E. M., Vill, P., Skoog, T., Jaakma, Ü., Kere, J., Bürglin, T. R., Katayama, S., Org, T., & Kurg, A. (2024). Molecular cloning of PRD-like homeobox genes expressed in bovine oocytes and early IVF embryos. *BMC Genomics*, 25(1), 1048. <https://doi.org/10.1186/S12864-024-10969-W> PMID: 39506635.
- II. Yaşar, B.***, Boskovic, N.*, Org, T., Kere, J., Kurg, A., & Katayama, S. (2025). RNA sequencing of mRNA 5'-ends reveals regulators of bovine embryonic genome activation. *BMC Genomics*, 26(1), 910. <https://doi.org/10.1186/S12864-025-12110-X> PMID: 41087861.
- III. Yaşar, B.**, Org, T., Ivask, M., Gunaydin, G. Y., Boskovic, N., Jaakma, Ü., Kere, J., Kurg, A., & Katayama, S. (2026). Structure and genomic organization of the human DUX4 homologue bovine DUXC. *BioRxiv*, 2026.03.05.709741. <https://doi.org/10.64898/2026.03.05.709741>
- IV. Boskovic, N.**, Ivask, M., Yazgeldi Gunaydin, G., **Yaşar, B.**, Katayama, S., Salumets, A., Org, T., Kurg, A., Lundin, K., Tuuri, T., Daub, C. O., & Kere, J. (2025). Oxygen level alters energy metabolism in bovine pre-implantation embryos. *Scientific Reports*, 15(1), 11327. <https://doi.org/10.1038/s41598-025-95990-z> PMID: 40175462.

*These authors contributed equally. The publications above have been reproduced with permission under Creative Commons license [CC BY 4.0].

Author's contributions

Study I: Methodology, Conceptualization, Investigation, Validation, Data curation, Formal analysis, Visualization, Project administration, Writing – original draft, Writing – review & editing.

Study II: Methodology, Software, Conceptualization, Data curation, Formal analysis, Visualization, Project administration, Writing – original draft, Writing – review & editing.

Study III: Methodology, Software, Conceptualization, Investigation, Validation, Data curation, Formal analysis, Visualization, Project administration, Writing – original draft, Writing – review & editing.

Study IV: Assisted with preprocessing of RNA sequencing data and reviewed the manuscript.

ABBREVIATIONS

2CLC	2-Cell-Like Cell
3PN	Three Pronuclei
8CLC	8-Cell-Like Cell
9aaTAD	9 amino acid Transactivation Domain
<i>ARGFX</i>	<i>Arginine-fifty Homeobox</i>
ART	Assisted Reproductive Technology
ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput-seq
BAM file	Binary alignment map file
BCL file	Binary base call file
BFF	Bovine Fetal Fibroblast
CDS	Coding Sequence
ChIP-seq	Chromatin Immunoprecipitation followed by Sequencing
CTD	C-Terminal Domain
dsRNA	double-stranded RNA
<i>DUX</i>	<i>Double Homeobox</i>
EGA	Embryonic Genome Activation
EMSA	Electrophoretic Mobility Shift Assay
ERV	Endogenous Retroviruses
ESCs	Embryonic Stem Cells
ETCHbox	Eutherian Totipotent Cell Homeobox
FISH	Fluorescence <i>In Situ</i> Hybridization
FSHD	Facioscapulohumeral Muscular Dystrophy
GV oocyte	Germinal Vesicle oocyte
H3K27ac	histone H3 K27 acetylation
H3K27me3	histone H3 K27 tri-methylation
H3K4me3	histone H3 K4 tri-methylation
H3K9me3	histone H3 K9 tri-methylation
HERVK	Human ERV K
hESCs	human Embryonic Stem Cells
hPSCs	human Pluripotent Stem Cells
iBM	induced Blastomere-like
ICM	Inner Cell Mass
ICSI	Intracytoplasmic Sperm Injection
iPSCs	induced Pluripotent Stem Cells
IVF	<i>In Vitro</i> Fertilization
IVM	<i>In Vitro</i> Maturation
IVT	<i>In Vitro</i> Transcription
KO	Knockout
<i>LEUTX</i>	<i>Leucine Twenty Homeobox</i>
LINE	Long Interspersed Element
LTR	Long Terminal Repeat

MaLRs	Mammalian apparent LTR-Retrotransposons
MERV L	murine ERV L
mESCs	mouse Embryonic Stem Cells
MII oocyte	Metaphase II-arrested oocyte
MZT	Maternal-to-zygotic Transition
ncH3K4me3	non-canonical H3K4me3
NMD	Nonsense-Mediated mRNA Decay
<i>NOBOX</i>	<i>Newborn Ovary Homeobox</i>
ORF	Open Reading Frame
PCA	Principal Component Analysis
Pol II	RNA Polymerase II
PRDL	PRD-like
RBP	RNA-Binding Protein
RNA-FISH	RNA Fluorescence <i>In Situ</i> Hybridization
RNA-seq	RNA sequencing
RT-PCR	Reverse Transcription Polymerase Chain Reaction
scRNA-seq	Single-cell RNA-seq
SNPs	Single-Nucleotide Polymorphisms
SNV	Single-Nucleotide Variant
SINE	Short Interspersed Element
siRNA	small interfering RNA
STRT-Seq	Single-cell Tagged Reverse Transcription
TF	Transcription Factor
TFE	Transcript Far 5'-End
<i>TPRX</i>	<i>Tetra-peptide Repeat Homeobox</i>
TSS	Transcription Start Site
UMI	Unique Molecular Identifiers
UMAP	Uniform Manifold Approximation and Projection
WGBS	Whole Genome Bisulfite Sequencing
WGS	Whole-Genome Sequencing
ZGA	Zygotic Genome Activation

1 INTRODUCTION

Mammalian embryonic development is a very complex biological process that requires precise temporal regulation of gene expression. This highly coordinated process gives rise to hundreds of distinct cell types and trillions of cells with specialized roles in the human body. This cycle of cell generation is sustained through reproduction, which involves the joining of two differentiated germ cells, the spermatozoon and the oocyte, leading to the formation of a zygote, the starting point of embryo development.

The first week of embryonic development is critically important, as extensive cellular reprogramming occurs during this period. Germ cells are first reprogrammed into totipotent cells and subsequently toward pluripotency, accompanied by drastic changes in both the transcriptome and epigenome. These changes include embryonic genome activation (EGA) and maternal-to-zygotic transition (MZT) during which transcripts expressed from the embryonic genome gradually become predominant, whereas maternal transcripts are degraded as embryo develops. Understanding the mechanisms that govern these early developmental processes could help improve clinical procedures in assisted reproductive technology (ART). Although ART has led to the birth of close to 13 million infants worldwide, its overall efficiency remains limited. Meanwhile, around the world, approximately one in six individuals is affected by infertility. In addition, many countries face a fertility crisis. The fertility rate across Europe is well below 2.1 births per woman, the level required to maintain a steady population size. On the other hand, *in vitro* fertilization (IVF) is also used to produce livestock, and improving the efficiency of such procedures, which are similar to those used in humans, is also of interest. Given the prevalence of infertility, declining population growth, and agricultural implications, improving IVF success rates has both clinical and economic significance. A more comprehensive understanding of early embryo development is therefore crucial, as it may provide insights needed to enhance embryo viability *in vitro* and increase the efficiency of ART.

This thesis presents four distinct studies using bovine as a model organism, each focusing on the molecular players involved in a critical event that occurs during the first week of preimplantation embryo development: bovine EGA. A transient burst of PRD-like homeobox gene expression is observed during human EGA. The first study provides experimental evidence for the expression of the bovine counterparts of these genes during bovine EGA, highlighting the similarity between the two species during early development. The second study identifies putative bovine EGA inducers through 5'-end-focused transcriptomic analysis of preimplantation stage embryos. The third study investigates the functional bovine homologue of the so-called master regulator of human EGA, DUX4, characterizing the *DUXC* gene structure and its locus across several cattle breeds. Lastly, the fourth study examines the impact of oxygen levels on IVF bovine embryos and their relation to bovine EGA.

2 REVIEW OF THE LITERATURE

2.1 Foundations of reproductive biology

The origins of reproductive biology date back to the first century. Soranus of Ephesus (98–138 AD) was a physician born in the city of Ephesus in Ancient Greece, located in modern-day Turkey, and was the first to describe the anatomy of ovaries (Turgut & Kutlubay, 2023). Andreas Vesalius (1514–1564), who was a Professor at the University of Padua in modern-day Italy, provided a more detailed description of the ovary and ovarian follicles centuries later in 1500s (Sokkary & Dietrich, 2018). Regnier de Graaf (1641–1673), a 17th-century Dutch physician, described ovarian follicles, which are named after him and also referred to as Graafian follicles (Thiery, 2009). An amateur Dutch scientist named Antonie van Leeuwenhoek (1632–1723) discovered the spermatozoa in human semen (Howards, 1997). Baltic-German embryologist Karl Ernst von Baer (1792–1876), who studied medicine in Tartu, made significant contributions to reproductive medicine during his career. He discovered mammalian oocyte and described the preimplantation embryo development until the blastocyst stage (Brauckmann, 2015). Karl Ernst von Baer is considered one of the most famous scientists in Estonia and an important figure for medical and biology students at the University of Tartu. In his honor, a statue was erected in 1886 and still stands today in Toomemägi, a large park in the center of Tartu (Figure 1). The contributions of the physicians and scientists briefly mentioned here, along with many others whose work is not detailed, set the foundations of human embryology and paved the way for modern reproductive biology.



Figure 1. Statue of Karl Ernst von Baer in Tartu. Photo taken by Barış Yaşar.

2.2 Preimplantation embryo development

Mammalian embryo development is initiated with fertilization, which encompasses the fusion of the oocyte by the sperm. Prior to this event, germ cells are produced through gametogenesis in both male and female reproductive organs. While the production of gametes in humans starts during puberty in males, in females, this biological process starts already before birth. Because of the visible nuclear membrane present in fully grown immature oocytes, oocytes at this stage are called germinal vesicle (GV) oocytes. GV oocytes are arrested in meiosis (prophase I) until puberty (R. Li & Albertini, 2013), and through meiotic maturation at reproductive age, meiosis is resumed (Figure 2). This is followed by a second meiotic arrest at metaphase II (MII), which defines matured oocytes at this stage as MII oocytes (Duncan et al., 2020). The completion of meiosis does not occur until fertilization. In males, spermatogonia undergo mitosis to form primary spermatocytes. Through meiosis, secondary spermatocytes are produced first, followed by spermatids, which subsequently differentiate into haploid sperm cells (Figure 2). Fertilization occurs within 24 h of ovulation upon the entry of sperm into mature oocyte in the Fallopian tubes, where spermatozoa can survive at least 48 h. Following fertilization, the diploid zygote undergoes a series of asynchronous cleavage divisions. This leads to a progressive increase in cell number, passing through various multicellular stages, often characterized by 2-, 4-, 8-, and 16-cell stages, until the embryo reaches the morula and finally the blastocyst stage. Until the morula stage, each cell of the embryo, also known as a blastomere, is loosely attached to its neighboring cells. With the morula stage, cell-to-cell contact becomes tighter and leads to a tissue-like appearance of the embryo. Stages between the zygote and the morula stage are called, hence, cleavage stage embryos. After the morula stage, a fluid-filled cavity, the blastocoel, starts appearing in the center, forming the blastocyst with the inner cell mass (ICM) and trophectoderm. The blastocyst eventually implants itself in the wall of the uterus (endometrium) and continues its development. Mammalian embryo development can be divided into two groups based on this implantation event: preimplantation and post-implantation (Jukam et al., 2017). Preimplantation embryo development is also referred to as early embryo development (Figure 2).

Early blastomeres are totipotent, which describes these cells' ability to give rise to any type of tissue, including extraembryonic tissues. Totipotent cells can generate a complete, viable organism. Cells derived from the ICM of blastocysts are considered pluripotent, which describes a cellular state with the ability to form any type of cell except for extraembryonic tissues. Propagation of ICM cells in the cell culture led to the discovery of embryonic stem cells (ESCs), which are self-renewal and pluripotent. Mouse ESCs were the first to be cultured *in vitro* in 1981 (Evans & Kaufman, 1981). Seven years later, this was followed by ESCs derived from a human embryo (Thomson, 1998). Although there have been efforts to establish a cell line derived from the bovine blastocyst, the stable derivation of bovine ESCs, with long-term self-renewing and pluripotency, has only recently been reported (Bogliotti et al., 2018). Apart from ESCs, adult stem cells and progenitor cells also exist in adult tissues. These cells can differentiate into

distinct cell types within their own lineage. Regarding ESCs, depending on the culture conditions, primed (Nichols & Smith, 2009) and naïve (K. Huang et al., 2014; Theunissen et al., 2016) human ESC states have been described. Although transcriptomic analyses show that naïve human ESCs closely resemble epiblast cells, a layer of cells within the ICM, they still represent a distinct cell state that is not identical to cells *in vivo*. Therefore, both naïve and primed ESCs are cell states that exist under *in vitro* culture conditions and serve as valuable models to study events occurring during early development.

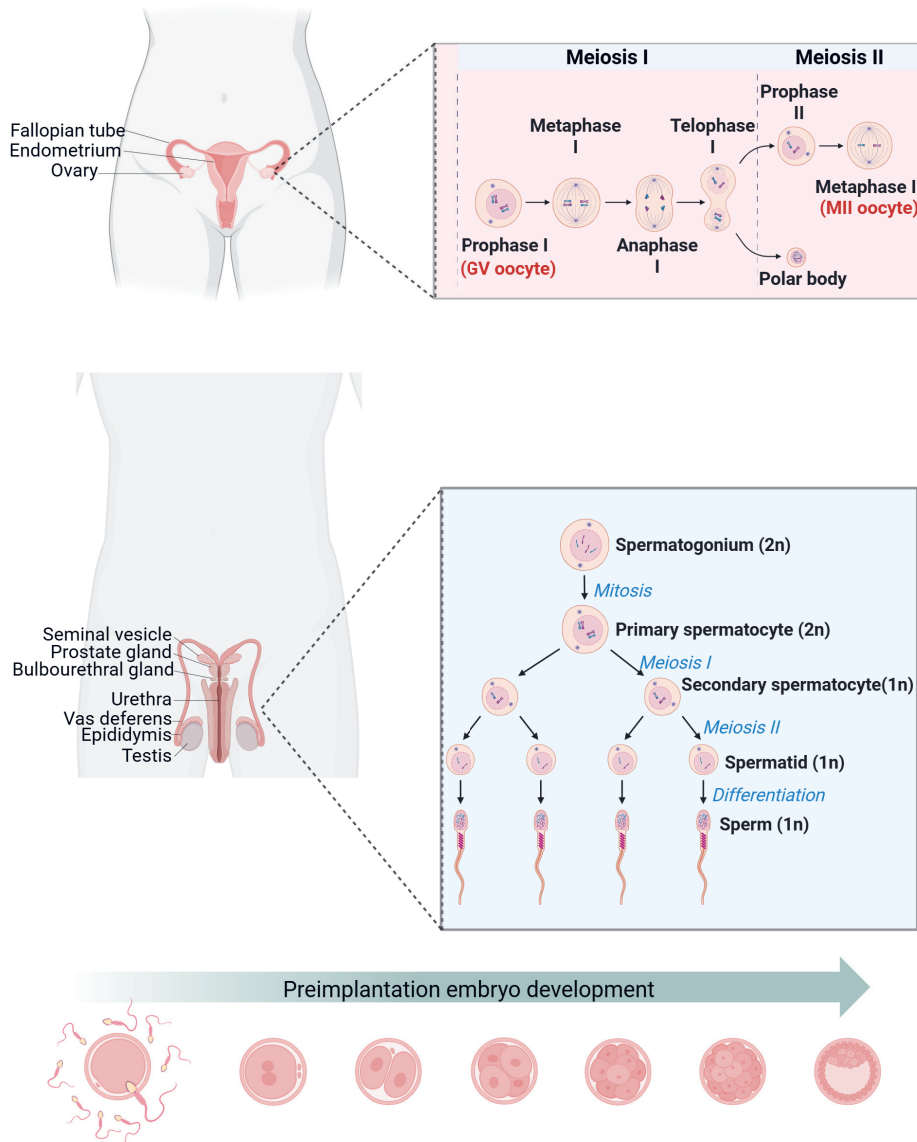


Figure 2. Germ cells and preimplantation embryo development. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/okwm2y8>. Adapted from “Meiosis – Oocyte” and “Male Reproductive System Anatomy”. Retrieved from <https://app.biorender.com/biorender-templates>.

2.3 Assisted reproductive technologies

Attempts to replicate *in vivo* preimplantation embryo development under culture conditions have given rise to human *in vitro* fertilization (IVF) therapy (Figure 3). This foundational work, led by Robert Edwards, was primed with the findings that demonstrate the maturation of human oocytes *in vitro* (Edwards, 1965). The following findings showed that in a laboratory setting, these oocytes could be fertilized by spermatozoa (Edwards et al., 1969), which later enabled culturing human embryos until the blastocyst stage (Steptoe et al., 1971). The breakthrough happened in 1978 when Louise Brown, who is currently a 47-year-old female, became the first human to be born through IVF therapy (Steptoe & Edwards, 1978). For his contributions to the advancement of reproductive technologies, Robert Edwards was awarded the 2010 Nobel Prize in Physiology or Medicine.

A more specialized version of IVF therapy was later introduced, called intracytoplasmic sperm injection (ICSI), in which a denuded oocyte is microinjected with a sperm (Figure 3). Ryuzo Yanagimachi performed the first microinjection of spermatozoa into oocytes in hamsters (Uehara & Yanagimachi, 1976). Improvement of this ICSI procedure later resulted in pregnancies in humans (Palermo et al., 1992). These developments in assisted reproductive technologies (ART) have revolutionized reproductive medicine. The total number of ICSI and IVF children has surpassed 9 million (Adamson et al., 2025).

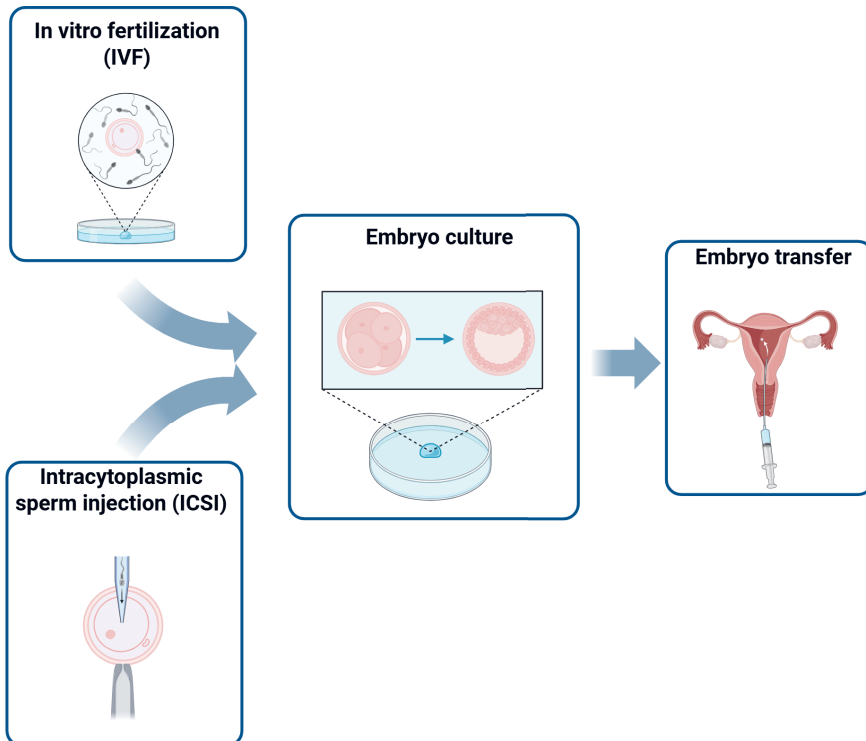


Figure 3. IVF and ICSI procedures. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/xo52vh2>.

Infertility is estimated to affect between 12.6% and 17.5% of couples of reproductive age (Cox et al., 2022). ART offers solutions to individuals who wish to have a child but are unable to do so due to infertility. However, the success rate of IVF has been limited, with a single treatment cycle resulting in pregnancy in about 30% of cases (Calhaz-Jorge et al., 2016). Approximately 10% to 15% of IVF embryos stall their development at the 2- to 4-cell stage (Betts & Madan, 2008). One of the key events occurring during this developmental window is embryonic genome activation (EGA). Therefore, further research is needed to understand the molecular mechanisms underlying EGA. However, research using human embryos is limited by ethical considerations and biological material availability, as only a small number of cells are present during this time of development. The use of closely related model organisms, such as bovine embryos, therefore provides a valuable alternative to study EGA.

2.4 Bovine genome and its evolution

Most of the domesticated cattle breeds we have today were believed to have evolved from currently extinct species, wild ox or aurochs (*Bos primigenius*) (Loftus et al., 1994). Traditionally, in the history of animal husbandry spanning 10,000 years, a single domestication event was thought to have occurred for two major cattle types: zebu (*Bos indicus*) and taurine (*Bos taurus*) cattle. The widely held opinion was that selective breeding of *Bos taurus* cattle gave rise to the emergence of *Bos indicus*, a cattle breed distinguished by the presence of a hump (Machugh et al., 1997). However, mitochondrial DNA (mtDNA) and micro-satellite marker analysis of cattle breeds has pointed to separate aurochs origin populations for Asian (*Bos indicus*) and European-African (*Bos taurus*) domestic cattle, which diverged around 210,000 years ago, well before human-mediated domestication (Andersson, 2001; Loftus et al., 1994; Machugh et al., 1997). Whole genome analysis of dozens of ancient cattle has also demonstrated the division of *Bos taurus* and *Bos indicus* (Verdugo et al., 2019). Despite being named as different species, *Bos indicus* and *Bos taurus* are often considered sub-species, as they can produce fertile offspring (Loftus et al., 1994). To emphasize their sub-species, they may also be referred to as *Bos taurus indicus* and *Bos taurus taurus* (Figure 4).

The taurine cattle genome was sequenced, assembled, and annotated by the Bovine Genome Sequencing and Analysis Consortium as reported in 2009 (Elsik et al., 2009). This marked bovine as the first species in the order *Cetartiodactyla* to have its whole genome sequenced (Seo et al., 2013). The assembly of the cattle genome uncovered high densities of repetitive elements and segmental duplications, which are known drivers of chromosomal rearrangements that shaped the cattle karyotype as we know today (H. A. Lewin, 2009). Synteny-based comparisons of chromosomes among human, bovine, and rodent species demonstrated more similarity between the human and cattle genome organization (Elsik et al., 2009). Amino acid sequence identity between humans and bovines was higher compared to humans and rodents (Elsik et al., 2009).

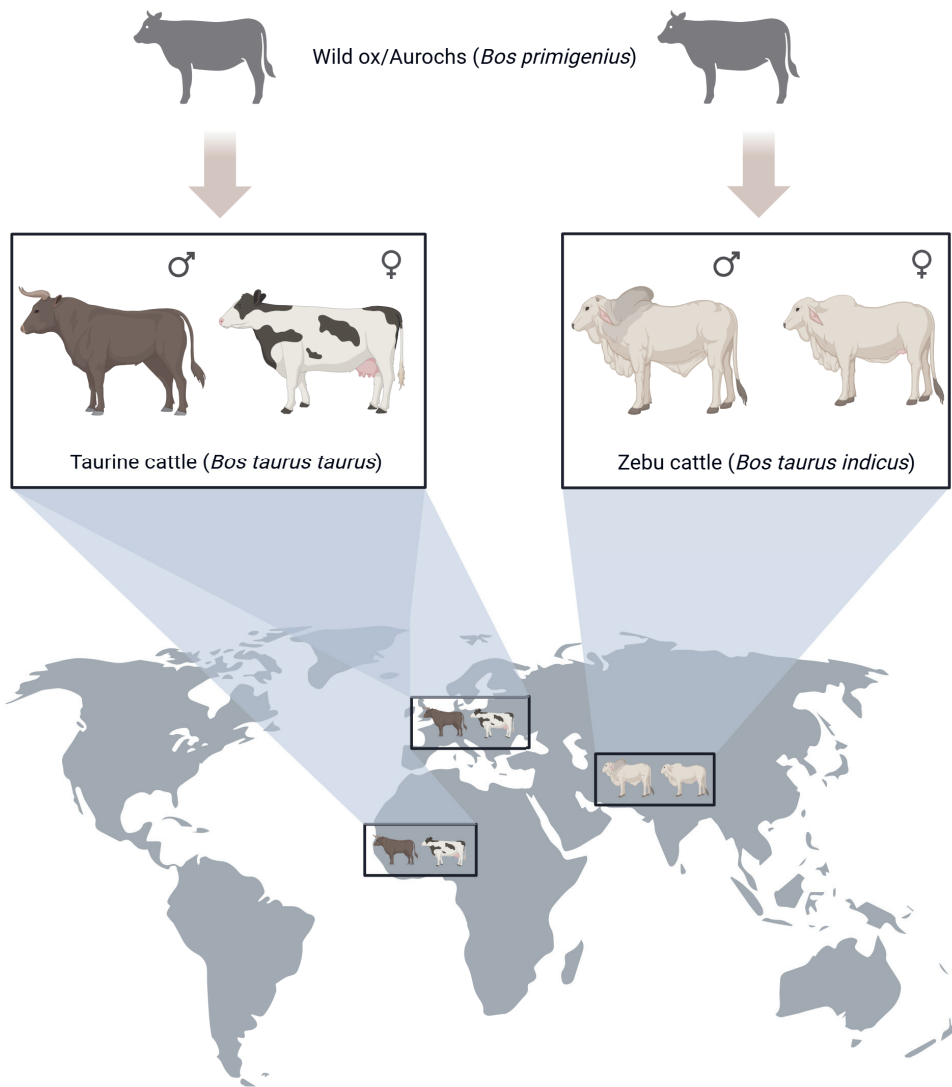


Figure 4. Taurine and indicine cattle. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/4znd0m9>.

The Bovine HapMap Consortium conducted a genome-wide survey based on single-nucleotide polymorphisms (SNPs) with 19 breeds, including humpless taurine (*Bos taurus*) and humped indicine or zebu (*Bos indicus*) cattle (Gibbs et al., 2009). Principal component analysis (PCA) using all SNPs separated taurine (e.g., Holstein) and indicine (e.g., Brahman) breeds with admixed breeds (e.g., Beefmaster) forming an intermediate cluster (Gibbs et al., 2009). While the observed SNP diversity in taurine cattle breeds was similar to that of humans, for indicine cattle breeds, the SNP diversity was higher, suggesting that wild ancestors of cattle, aurochs, faced mild bottlenecks thanks to their large population size (Gibbs et al., 2009). According to the initial results of the 1000 Bull Genomes

Project, which contained whole-genome sequences from 234 bulls, the average number of heterozygous sites for every 1,000 bases was 1.44 (Daetwyler et al., 2014). The rate of heterozygous sites in the cattle population was found to be greater than that observed in human Yoruba (1.03/kb) and European (0.68/kb) populations (Altshuler et al., 2010). This number is substantial, given the selective breeding cattle have experienced, a factor that is known to reduce genomic diversity. These findings point to great historical diversity prior to the domestication of cattle breeds.

Genomic data derived from tens of cattle breeds and hundreds of individuals so far have provided insights into understanding the genomic organization and evolution of the bovine genome, but initiatives such as the Ruminant Telomere-to-Telomere (RT2T) Consortium (Kalbfleisch et al., 2024) will enable an even more comprehensive understanding by uncovering the previously not sequenced parts of the genome, including centromeres and repetitive regions.

2.5 RNA sequencing methods to study early embryo development

RNA sequencing (RNA-seq) is the gold standard method to study transcriptomic states of cells. Conventional bulk RNA-seq methods were primarily designed to work with samples containing a large number of cells, such as cultured cells. Therefore, it has initially been a challenge to perform RNA-seq using samples that contain a low number of cells, such as preimplantation embryos.

Tang et al. introduced the first now well-known method suitable for working with as little material as a single cell (Tang et al., 2009). Following the method developed by Tang et al., several methods capable of profiling the transcriptome of a single cell have been introduced, and these widely used methods are called single-cell RNA-seq (scRNA-seq). The main steps to perform scRNA-seq include reverse transcription, second-strand synthesis, cDNA amplification, and preparation of the sequencing library (Kolodziejczyk et al., 2015).

Second-strand synthesis is either done via poly(A) tailing using terminal deoxynucleotidyl transferase (TdT) or through template switching. The former is used by Quartz-Seq (Sasagawa et al., 2013) and the Tang et al. method, while Smart-Seq (Switching Mechanism At the end of the 5'-end of the RNA Transcript) (Picelli, 2017; Ramsköld et al., 2012), and STRT-Seq (single-cell tagged reverse transcription) (Islam et al., 2011, 2012) utilize the latter. STRT-Seq is focused on sequencing the 5' ends of transcripts, which diminishes the number of reads needed for quantification (Kolodziejczyk et al., 2015).

There are two separate approaches for cDNA amplification: PCR or *in vitro* transcription (IVT). PCR is more popular and used by methods such as Smart-Seq2 (Picelli et al., 2013) and STRT-Seq, while methods that use IVT include CEL-Seq (cell expression by linear amplification and sequencing) (Hashimshony et al., 2012) and MARS-Seq (massively parallel RNA single-cell sequencing) (Jaitin et al., 2014).

One important aspect of scRNA-seq methods is the accuracy of gene expression quantification. Unique molecular identifiers (UMIs) are sequences integrated in the sequencing libraries and used to mark PCR duplicates to improve quantification (Fu et al., 2011; Islam et al., 2014; Kivioja et al., 2012). Methods such as STRT-Seq and Quartz-Seq utilize UMIs. An improved version of Quartz-Seq, which has higher UMI conversion efficiency, has been introduced and is named Quartz-Seq2 (Sasagawa et al., 2018).

Another aspect of scRNA-seq methods is the read coverage across the gene body. For example, STRT-seq provides a higher coverage at the 5' ends of transcripts (Abugessaisa et al., 2022). As a method that profiles 5' ends of mRNAs and enables identification of putative transcription start sites (TSSs) and the regulatory elements around the TSSs, STRT is an efficient method for studying gene regulation. Early development represents a critical window during which extensive transcriptional changes occur, making STRT an ideal method to investigate bovine EGA and the regulatory motifs involved in this process.

Although it is developed as a scRNA-seq method, STRT has been modified to work with solid tissues and cell suspensions (Ezer et al., 2021). In addition, an earlier modified protocol of STRT-Seq (Dong et al., 2018) has recently been used to profile human early embryos at embryonic arrest (T. Wang et al., 2024). More recently, a version of STRT optimized for compatibility with modern sequencing platforms, STRT-N RNA-seq, has been introduced (Boskovic et al., 2023). STRT-N RNA-seq utilizes spike-in RNAs as well as UMIs to provide more accurate quantification of gene expression. Therefore, STRT-N RNA-seq not only provides TSS information but also enables reliable gene expression measurement to better understand bovine EGA.

In the published work of STRT-N RNA-seq, along with the experimental protocol, a detailed protocol for the computational analysis has also been outlined (Boskovic et al., 2023). Two types of analysis are possible with the RNA-seq datasets generated with this method: gene-based and transcript far 5'-end (TFE)-based analysis. The former first defines 5'-UTRs and promoter-proximal upstream regions based on the used gene annotation file and then counts the RNA-seq reads that align to these defined 5' end regions (Figure 5). Reads aligned to genomic regions other than these 5' end regions are not included in the read count and discarded. Since counting the reads relies on the gene models present in the gene annotation file, gene-based analysis cannot detect the expression of unannotated genes.

Gene-based quantification

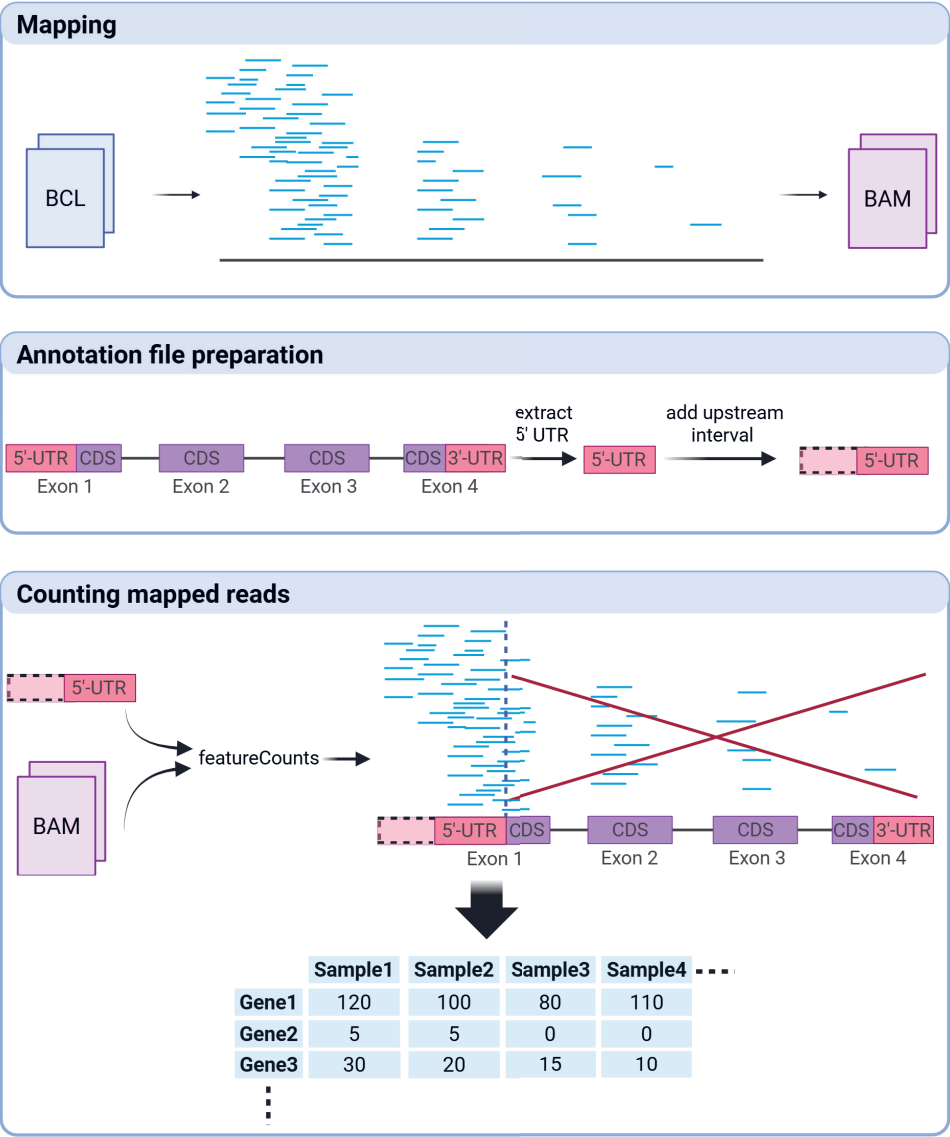


Figure 5. Schematic overview of the gene-based quantification pipeline. Only the main steps are highlighted for simplicity. The original script (STRTN.sh), and its version optimized to process recent Ensembl gene annotations (STRTN_bTau_ens112.sh) can found at <https://github.com/gyazgeldi/STRTN> and https://github.com/baryasar/bTau_Embryo_TFE, respectively. BCL: Binary base call file; BAM: Binary alignment map file; CDS: Coding sequence. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/vqflcii>.

TFE defines a quantification unit for gene expression, introduced earlier (Töhönen et al., 2015). Unlike gene-based analysis, TFE-based analysis quantifies the expression of transcripts assembled *de novo* (“from scratch”). Mapped reads are used to assemble transcripts, from which the first exons are extracted. Overlapping the first exons derived from each sample/replicate and developmental stage results in non-redundant genomic intervals, serving as a global first exon repertoire for the analyzed samples (Figure 6). In this pipeline, the intervals of these first exons are called TFE regions (Boskovic et al., 2023). Reads that align to each TFE region are counted separately.

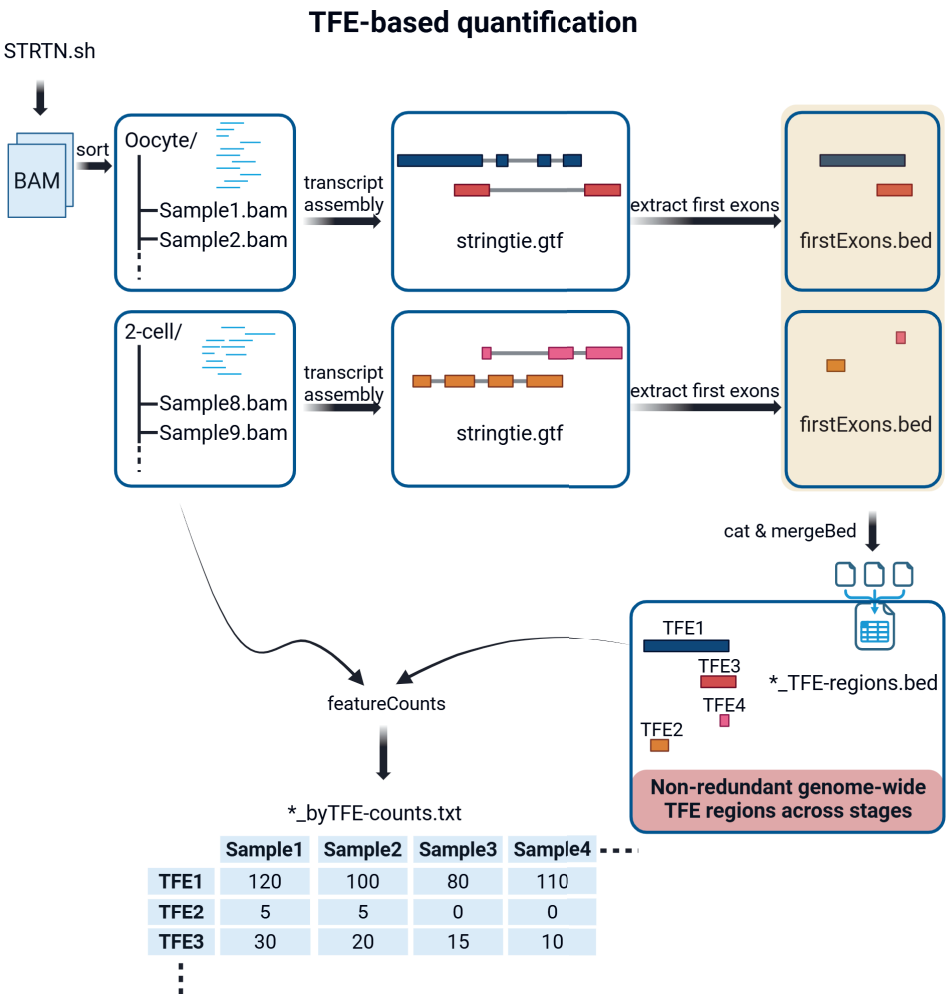


Figure 6. Schematic overview of the TFE-based quantification pipeline. Only the main steps are highlighted. The first exons on the same strand are merged in the pipeline, but strand orientation is not depicted in the figure for simplicity. The original script (STRTN-TFE.sh), which details the input, output, and intermediate files, as well as the tools depicted in the figure or utilized in the original code, can be found at <https://github.com/gyazgeldi/STRTN>. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/4fgkxqw>.

The 5' end of each mapped read is a candidate TSS coordinate. Within each TFE region, these signals are aggregated, and the coordinate with the strongest read support is selected to represent that TFE region as a putative TSS. This selected 1 bp resolution coordinate is referred to as a TFE. TFE coordinates are then compared against gene annotations to assign gene names. Since TFEs represent independent TSS events, multiple TFEs can be associated with a single gene (Figure 7).

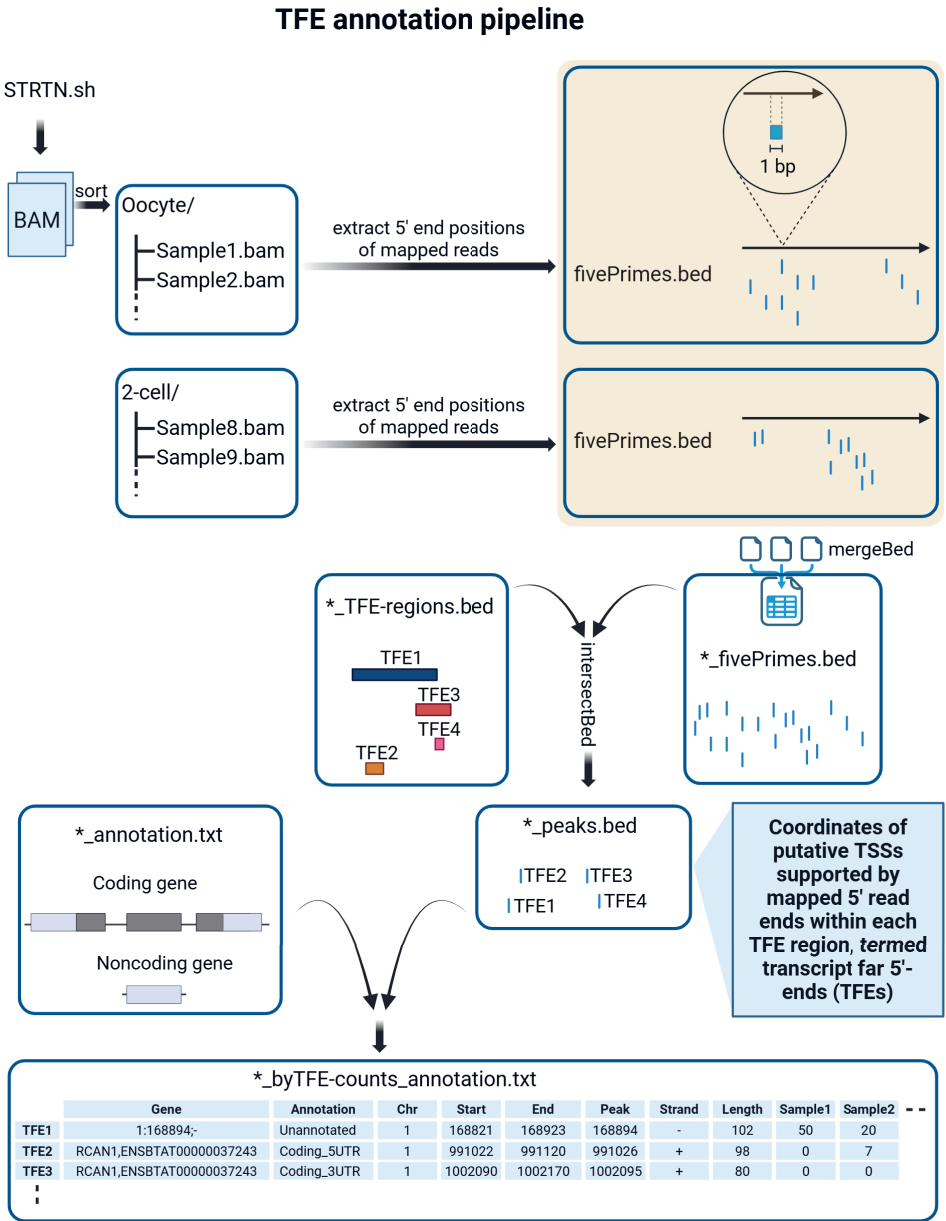


Figure 7. Schematic overview of the pipeline assigning annotations to TFEs. Only the main steps are highlighted for simplicity. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/xrx19sy>.

An additional level of TFE annotation involves assigning genomic feature-based labels to each TFE. Depending on whether TFE corresponds to a coding, a non-coding gene, or does not correspond to any annotated gene, each TFE is assigned one of nine distinct genomic annotations. These genomic annotation features are as follows: unannotated, coding upstream, coding 5'-UTR, coding CDS, intron, coding 3'-UTR, noncoding upstream, noncoding 1st exon, noncoding other exon (Figure 8).

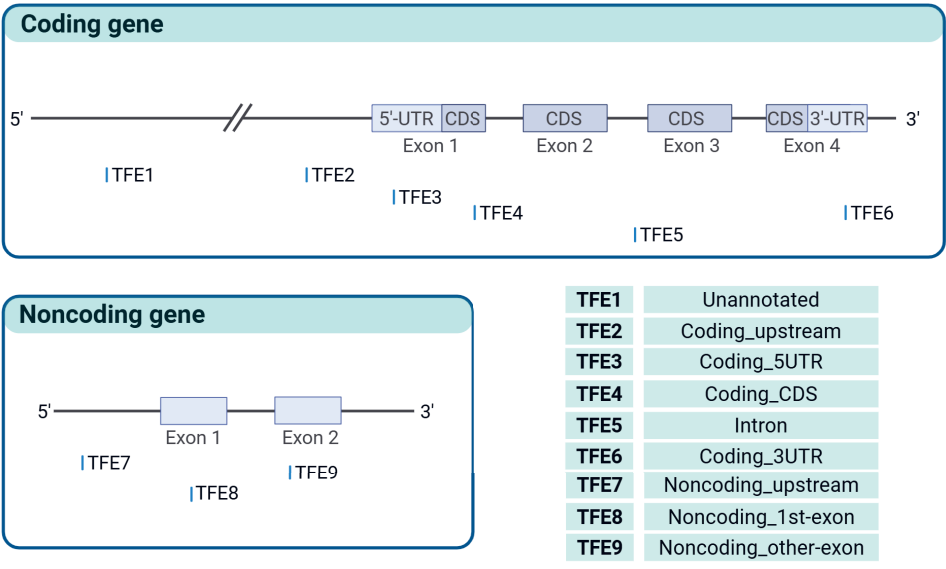


Figure 8. Genomic annotation features of TFEs. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/o174950>.

2.6 Embryonic genome activation and transcriptional studies in mammals

Maternal-to-zygotic transition (MZT) is a time window when maternal mRNA degradation is accompanied by the zygotic transcription from the embryonic genome. The latter process is referred to as zygotic genome activation (ZGA) or embryonic genome activation (EGA). EGA, which describes the transcriptional burst from the zygotic genome following the fertilization of the transcriptionally quiescent oocyte, follows strict temporal and coordinated regulation of cellular processes. In mice, the terms minor and major EGA were coined due to their immediate sequential occurrence, but such generalization to other mammals, including bovines with slightly delayed EGA, might not truly reflect their development (Svoboda, 2018). In the literature, the terms ‘major EGA’ and ‘EGA’ are often used interchangeably, and they define the developmental period when the first large-scale gene upregulation is observed after fertilization and

prior to implantation. Major EGA is species-specific (Kravchenko & Tachibana, 2024; Oomen et al., 2025) and occurs between the mid 1-cell and the late 2-cell stage in mice (Schultz et al., 2018) and during the 4-to-8-cell transition in pigs (Kong et al., 2020), guinea pigs (Guan et al., 2024), opossums (Mahadevaiah et al., 2020), rhesus macaques (Chitwood et al., 2017), and humans (Töhönen et al., 2015), 8-to-16-cell transition in goats (Y. Li et al., 2020) and bovines (Graf et al., 2014) (Figure 9).

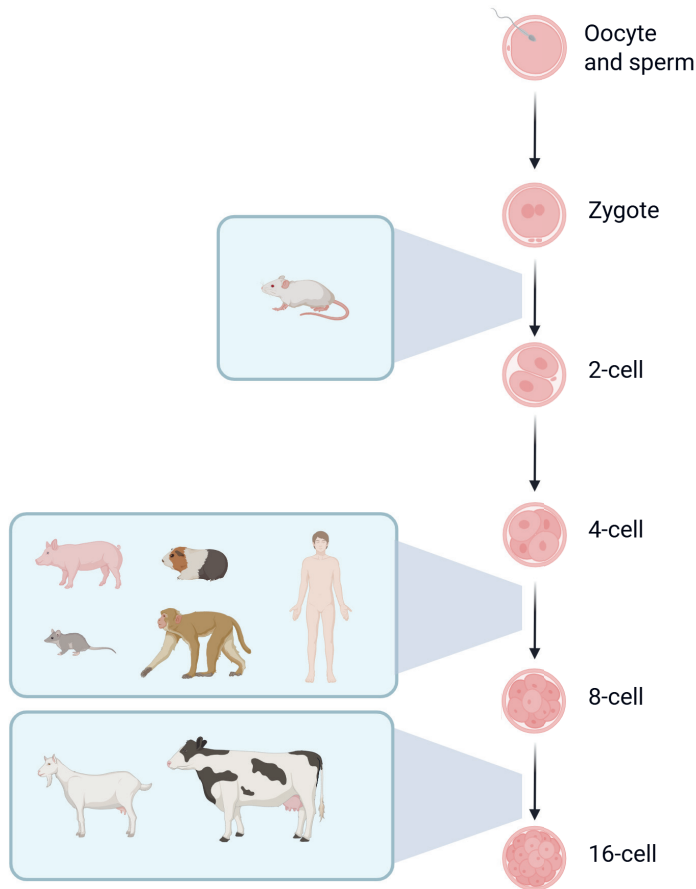


Figure 9. Timing of embryonic genome activation across mammalian species. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/ks8ncmi>.

2.6.1 Mouse

Early comprehensive analysis of mouse oocytes and preimplantation embryos revealed two waves of transcriptional burst from the embryonic genome (Hamatani et al., 2004). Translation of maternal RNA was shown to be required for the initiation of EGA (Hamatani et al., 2004). Allelic expression analysis of mouse early embryos using scRNA-seq demonstrated that maternally expressed

transcripts dominate the transcriptome in the zygote and early 2-cell embryos, with paternally expressed transcripts reaching roughly equal levels at the 4-cell stage (Deng et al., 2014). This observation reflects early processes of EGA and MZT.

In addition to transcriptome analysis, RNA polymerase II (Pol II) profiling was performed to assess the transcriptional activity. Immunostaining of MII oocytes for Pol II did not produce any signal, but after fertilization until the blastocyst stage, Pol II signal was detected (B. Liu et al., 2020). Profiling of Pol II confirmed the presence of Pol II on chromatin not before but after fertilization (B. Liu et al., 2020), in line with transcriptionally quiescent oocytes. Interestingly, global binding of Pol II was not similar to open chromatin profiles throughout mouse early embryo development (B. Liu et al., 2020). Promoters associated with Pol II can be accompanied by a distinct set of factors resulting in pre-initiation, paused, arrested, or terminating complexes (Adelman & Lis, 2012). The term poised was used to describe the state of Pol II located near promoters, together with either of these complexes, without indicating whether transcription is occurring or not (Adelman & Lis, 2012). Therefore, Pol II poising can be observed at genes that are actively transcribed or at those where the transcription machinery remains inactive. Despite the limited RNA transcription at the 1-cell stage mouse embryo, global high enrichment of poised Pol II (at promoter, intergenic, and gene body regions) was observed, suggesting loading of Pol II at these regions prior to its transcriptional activity (B. Liu et al., 2020). After EGA, described Pol II loading was weakly observed at inactive promoters and developmental genes, which was interpreted as a strategy to prevent unwanted activation of such genes during EGA (B. Liu et al., 2020). Specifically, EGA genes showed high Pol II enrichment right before EGA in the early 2-cell stage, indicating a pre-configuration of Pol II at these genes, before they were activated and showed even higher Pol II enrichment (B. Liu et al., 2020).

2.6.2 Human

The main source for studying human preimplantation embryos has been the spare *in vitro* fertilization (IVF) embryos from clinics (Samorinha et al., 2014). These studies initially mainly included morphological examination or identification of factors and conditions to enhance embryo culture for IVF treatments. However, to get a deeper insight into the early developmental processes, utilizing sensitive gene expression profiling methods is fundamental. Here, studies that investigate the human preimplantation embryo development from the perspective of transcriptomics are summarized.

One efficient way of investigation is global transcriptional analysis, which has the potential to provide a better understanding of preimplantation embryo development. However, one initial study (Braude et al., 1988), which demonstrated the human EGA to occur between the 4- and 8-cell embryos, used translation as a proxy for transcription. They showed that α -amanitin, a potent transcriptional inhibitor, stalls the progression of the 4-cell stage human embryos and inhibits

protein synthesis between the 4- and 8-cell stages, due to the absence of transcripts coding for these proteins (Braude et al., 1988). Microarray analysis of early human embryos during the first three days of development revealed gene downregulation to be more pronounced than upregulation, in line with the MZT (Dobson et al., 2004). Another microarray analysis also suggested an initial maternal mRNA clearance phase followed by embryonic gene upregulation (P. Zhang et al., 2009). Two waves of embryonic transcription events were described in human preimplantation embryos, and their treatment with a translation inhibitor suggested that translation of maternal or very early embryonic RNAs is needed for minor EGA to occur (Hamatani et al., 2006). Microarray analysis of individual blastomeres from embryos up to the 8-cell stage revealed that pre-compaction embryos do not show signs of lineage commitment (Galán et al., 2010). Three imaging parameters that predict the development potential of human embryos to the blastocyst stage with 94% sensitivity, associated with four distinct modes of expression patterns of selected genes in individual blastomeres (Wong et al., 2010). Some genes, which are significantly activated in the 2-cell human embryos, were identified and referred to as ‘earlier risers’. Quantitative PCR analysis of these genes in α -amanitin-treated 2-cell embryos showed the absence of early riser transcripts, indicating that embryonic transcription also occurs at the 2-cell human embryos (Vassena et al., 2011). A later study suggested the initiation of the transcription to occur at the 1-cell stage human embryos (Asami et al., 2022).

Single-cell RNA-seq analysis of human embryos up to the 8-cell stage showed upregulation of 79 genes already at the pronuclear or zygote stage (Xue et al., 2013). Nearly half of these genes maintain their expression levels in the 2- and 4-cell stage, with a subsequent downregulation at the 8-cell stage (Xue et al., 2013). Other studies with single-cell RNA-seq profiling of human embryos confirmed major EGA to occur during the 4-to-8-cell transition (Töhönen et al., 2015; Yan et al., 2013). Blastomeres of the same stage showed mostly similar transcriptional profiles, but the expression of downregulated genes varied more than the upregulated ones in 8-cell blastomeres, indicating the maternal mRNA degradation to occur more heterogeneously (Yan et al., 2013). Based on the reanalysis of human single-cell RNA-seq datasets, active repression of transcription was suggested to be present prior to EGA (Boroviak et al., 2018). Another scRNA-seq study investigating the dynamics in human preimplantation embryos revealed elevated levels of gene expression from the Y chromosome between embryonic day 3 (E3) and 4 (E4), indicative of the MZT (Petropoulos et al., 2016). Furthermore, with the help of SNPs detected in reads from scRNA-seq, they observed biallelic transcripts derived from X chromosome genes in male E3 cells, suggesting remaining maternal transcripts. The data showed that full maternal mRNA clearance is completed in E4 embryos. Comparison of biparental embryos to uniparental embryos using scRNA-seq showed that maternal mRNA clearance, whose completion is suggested to be mediated by EGA (Bogliotti et al., 2020; Xue et al., 2013), is compromised in uniparental embryos, suggesting allele-specific expression in early embryos (Leng et al., 2019). Another study

performing transcriptome profiling of the same types of human embryos showed that EGA occurs in androgenetic but not in parthenogenetic embryos, indicating that the paternal genome is needed for human EGA (Yuan et al., 2023). A study focusing on MZT in human embryos revealed two modes of maternal mRNA clearance, EGA-dependent and -independent (Sha et al., 2020).

Poly(A) tail profiling of human early developmental stages revealed that 3' UTRs of some transcripts are shortened, with their proportion reaching to 60% in the 1-, 2-, and 4-cell stages and decreasing to around 40% at the 8-cell stage and getting stabilized around 15% in blastocysts, which was the percentage also observed in somatic cells (Y. Liu et al., 2023). The transcripts with shortened 3' UTRs were found to get polyadenylated again, during which non-A residues, mostly consecutive U residues, got incorporated (Y. Liu et al., 2023). This event occurring as a means of MZT was referred to as maternal mRNA remodeling (Y. Liu et al., 2023). Inhibition of polyadenylation following fertilization using an adenosine analog caused an arrest at the 1-cell stage (Y. Liu et al., 2023), which could not be attributed to inhibition of transcription, as such embryos were previously shown to proceed to the 8-cell stage (X. Chen et al., 2019; Xia et al., 2019). Therefore, maternal mRNA remodeling was suggested to be required for the cleavage of the 1-cell human embryo.

2.6.3 Bovine

Initial efforts to define the timing of EGA in cattle included [³H]uridine incorporation experiments, which showed EGA to occur during the 8- to 16-cell stage in bovine embryos (Camous et al., 1986; Frei et al., 1989). Inhibition of transcription in IVF bovine embryos by α -amanitin did not allow embryos to reach past the 8-cell stage, demonstrating the need for zygotic transcription at this stage (Barnes & First, 1991). Later experiments involved longer incubations of embryos with [³H]uridine, enabling the detection of RNA synthesis from the zygotic genome as early as the 2- to 4-cell stage (Hyttel et al., 1996; Memili et al., 1998; Plante et al., 1994; Viuff et al., 1996). One study even showed [³H]uridine incorporation in zygote and 2-cell embryos (Memili & First, 1999). The 4-cell stage was reported to be the stage when ribosomal RNA transcription occurs based on fluorescence *in situ* hybridization (FISH) and silver staining (Viuff et al., 1998). Measuring dynamic levels of total, messenger, and poly(A)-tailed RNA revealed an overall increase in total and poly(A)-tailed RNA content after the 8-cell stage, coinciding with the onset of bovine EGA (Gilbert et al., 2009).

Reverse transcription polymerase chain reaction (RT-PCR) and microarray-based approaches were used to investigate the players involved in bovine EGA before RNA sequencing became widely available and affordable. Using an Affymetrix DNA microarray specifically designed for bovine, one of the early bovine embryo studies listed genes expressed from the embryonic genome by comparing 8-cell embryos treated with or without α -amanitin (Misirlioglu et al., 2006). A subsequent study using a similar strategy with α -amanitin showed that transcription activity affects the presence of proteins primarily found in the

nucleolus, such as RNA polymerase I and topoisomerase (Svarcova et al., 2007). Another microarray analysis identified a high number of differentially expressed genes for the 4- and 8-cell stage between *in vivo* and *in vitro*-produced bovine embryos (Kepkova et al., 2011). Even though commercial microarrays included a large number of genes, their limitation in terms of covering oocyte- and early embryo-specific genes led some researchers to use custom-designed arrays with over a thousand transcripts (Vallée et al., 2009). In this study, almost all of the genes on the array were found to be expressed in at least one of the stages during *in vivo* development, with the 2- and 8-cell embryos not showing a critical difference (Vallée et al., 2009). Another bovine-specific microarray was used to analyze ICM and trophectoderm (TE) of *in vivo* blastocysts, revealing that difference between the two cell types were more pronounced in *in vivo* embryos than in *in vitro* embryos (Hosseini et al., 2015). Despite their contribution to the transcriptomic understanding of embryo development, microarrays were replaced with RNA sequencing, which is not limited to only annotated transcripts and can reveal novel transcripts.

The very first RNA sequencing study applied to cattle early embryos included *in vitro* cultured embryos that reached either the blastocyst stage on day 8 or those that failed to transition from morula to blastocyst on day 8, referred to as normal and degenerative embryos, respectively (W. Huang & Khatib, 2010). Overall gene expression between the two types of embryos appeared similar, with only 47 differentially expressed genes detected. Alternative splicing was noted as a common event in the transcriptome of cattle blastocysts, and specific alternative transcripts were suggested to indicate abnormal embryo development. However, the results from this study should be interpreted cautiously due to the absence of replicates. The study by Huang and Khatib utilized pools of embryos, prompting other researchers to create RNA-seq libraries of single blastocysts to investigate the variability in gene expression among different embryos (Chitwood et al., 2013).

The following study on bovine embryo development was composed of a more comprehensive setup that included oocytes, cleavage stage embryos, and blastocysts. To reveal the genes involved in bovine EGA, the researchers employed three different strategies: (1) distinguishing transcripts exclusively produced from the embryonic genome by eliminating those from oocytes, (2) identifying paternal transcripts enabled by fertilizing distant subspecies of *Bos taurus*, and (3) analyzing reads that are potentially non-spliced nascent transcripts judged by their ability to map to introns (Graf et al., 2014). The 8-cell stage showed the highest number of genes identified by all three approaches: 129 first-detected transcripts, 395 paternally expressed transcripts, and 3,965 incompletely spliced transcripts. This unique approach is distinct from differential gene expression analysis, which revealed the greatest gene upregulation during the 8-to-16-cell transition, when major EGA takes place in *Bos taurus* (Graf et al., 2014).

RNA-seq of *in vivo* bovine preimplantation developmental stages identified the largest number of gene upregulation events during the 4-to-8-cell transition (Jiang et al., 2014), suggesting this stage as the major EGA in *in vivo* bovine embryos, in contrast to the widely accepted timing of EGA during the 8-to-16-cell

transition in IVF bovine embryos. This study also suggested that the transcriptomes of human and bovine early embryos are more similar to each other than to that of mouse (Jiang et al., 2014). Investigation of oocyte maturation by performing RNA-seq on individual GV and MII oocytes demonstrated that 85.7% of the upregulated mRNAs with complete 3' UTR contained cytoplasmic polyadenylation signal (Reyes et al., 2015). The authors suggested that the observed increase in transcripts during oocyte maturation may stem from cytoplasmic polyadenylation (Reyes et al., 2015). Small RNA-seq of oocytes, 8-cell, and blastocyst stage embryos demonstrated that the 8-cell bovine embryos contained a much higher abundance of miRNAs but markedly lower, nearly undetectable levels of Piwi-interacting-like RNAs compared with the other stages (Cuthbert et al., 2019). Combining small RNA-seq with α -amanitin showed that miRNA gets transcribed from the embryonic genome at the 2-cell stage for the first time, increasing its expression at the morula and blastocyst stages (Paulson et al., 2022). High levels of miRNAs may be implicated in the clearance of maternal RNAs. One transcriptomic study, which analyzed 2–4 cell bovine embryos produced by sperm samples derived from high and low fertility bovines, identified *ATXN2L* as a differentially expressed gene (Gross et al., 2019). *ATXN2L* depletion at the zygote stage resulted in 22.9% rise in blastocyst formation rate, indicating a potential role for this gene in embryo development (Gross et al., 2019).

The initial research for studying early bovine embryo development at the single-cell level included the analysis of the abundance of a subset of annotated bovine genes (Negrón-Pérez et al., 2017; Wei et al., 2017). The only single-cell transcriptomic study on the major bovine EGA analyzed a total of 14 IVF bovine embryos at the 8-cell and 16-cell stages (Lavagi et al., 2018). Pseudo-time line and clustering analysis showed that some blastomeres do not develop in a synchronous fashion. Interestingly, scRNA-seq was not able to detect 109 out of the 129 genes that were found to be expressed at the 8-cell stage in the study by Graf et al. (2014).

RNA-seq of embryos treated with or without Pol II inhibitor α -amanitin allowed researchers to identify transcripts carried over from the oocyte to the embryo and helped discover the requirement of EGA for maternal mRNA clearance (Bogliotti et al., 2020). Embryos were sampled 70–72 h postfertilization, capturing embryos at the 8-to-16-cell stage, which contained 2459 transcripts of embryonic origin and 719 transcripts of maternal origin. Of these, 905 transcripts required transcription for their clearance during this stage (Bogliotti et al., 2020). Twenty-nine TF-binding motifs were found to be located at the 5' ends of 2459 genes that are α -amanitin sensitive. These 29 motifs were predicted binding sites for ten TFs, namely KLF4, KLF5, KLF9, OTX2, YY1, GSC, EHF, ETV2, CREBZF and SMAD3 (Bogliotti et al., 2020). All of these genes, except for *SMAD3*, showed increased expression while transitioning from the MII oocyte to the 8-to-16-cell stage embryos. Nine of these TFs were predicted to potentially target 1548 out of 2459 embryonically transcribed genes during EGA. Among these TF genes, *KLF5* and *GSC* are the only genes not expressed in MII oocytes, which points to

their possible integral roles in gene regulation during bovine EGA in preimplantation development (Bogliotti et al., 2020).

Ribosome profiling of bovine oocytes and early embryos showed that transcription and translation were well correlated in the oocyte, 2-cell, morula, and blastocyst stages, but at the 8-cell stage, there was a shift, with translation not completely reflecting transcription (L. Zhu et al., 2022). Single-cell RNA-seq of bovine blastocyst before its expansion revealed that X chromosome dosage compensation occurs between the morula and late blastocyst stage (B. Hu et al., 2024). Thousands of novel genes, many of which showed dynamic expression during early embryo development, were identified through a combination of short- and long-read RNA-seq on bovine oocytes and embryos (Schettini et al., 2024). Recently, KLF and DUX factor binding motifs were found to be enriched around the transcription start sites of EGA genes (Oomen et al., 2025). Collectively, the abovementioned studies summarize key findings on the transcriptomics of early bovine embryo development.

2.6.4 Expression of repetitive DNA sequences in preimplantation embryos

Repetitive DNA sequences or repeats make up approximately half of the human genome (Liao et al., 2023). Repeats can be positioned in tandem, such as satellite DNA, or as interspersed sequences, such as retrotransposons (Liao et al., 2023). Retrotransposons with long terminal repeats (LTRs) are referred to as LTR retrotransposons or endogenous retroviruses (ERVs) (P. J. Thompson et al., 2016). LTR-lacking retrotransposons also exist, and the main ones include long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and SINE variable-number tandem-repeat Alu (SVA) (P. J. Thompson et al., 2016).

Analysis of highly expressed repeats in early human embryos revealed two clusters in expression patterns, separated by EGA (Grow et al., 2015). The pre-EGA repeat cluster was influenced by the deposited maternal transcripts, but with the onset of EGA, the repeat expression profile changed, for example, with a marked increase in human ERV K (HERVK) expression (Grow et al., 2015). HERVK elements can be categorized based on the LTRs they contain, for example, LTR5a, LTR5b, and LTR5H subtypes. In human embryos, LTR5H, which contained an OCT4 binding motif, was expressed, whereas the LTR5a and LTR5b subtypes lacking such a motif were not (Grow et al., 2015). It points to the selective expression of HERVK elements regulated by human embryonic proteins. Since many of HERVK retained intact open reading frame (ORF), retroviral HERVK proteins have the potential to be produced in human cells (Subramanian et al., 2011). Interestingly, the presence of HERVK-derived retroviral proteins was reported in human blastocysts after observation of viral-like particles through transmission electron microscopy (Grow et al., 2015). The specific immune response elicited by the transport of such viral proteins to the cytoplasm was suggested to protect early embryos against viral infections (Grow et al., 2015).

A member of the ERV family, MLT2A1 (P. J. Thompson et al., 2016), exhibited increased RNA expression as well as chromatin accessibility at the highest levels in 4- and 8-cell stage human embryos (L. Liu et al., 2019). Depletion of MLT2A1 in human embryos led to embryonic arrest and the loss of the EGA transcriptome signature (Xiang et al., 2026). Accessible MLT2A1 elements were enriched with binding motifs of DUX4 and OTX2 (L. Liu et al., 2019), both of which were shown to activate MLT2A1 (Hendrickson et al., 2017; Q. Wang et al., 2025). Similarly, at the onset of bovine EGA, another ERV family member MLT1A0 exhibited high expression and chromatin accessibility (Halstead et al., 2020). Expression of murine ERV L (MERVL) was shown to mediate the transition from totipotency to pluripotency, as its elimination blocked the progression of early mouse embryo development (Sakashita et al., 2023). Expression of ERVs in early mammalian embryos could be a conserved and necessary event.

Non-LTR retrotransposon SVAs were also highly expressed at the 8-cell and morula stage human embryos, but not in differentiated epithelial or endothelial cells (L. Gao et al., 2018). Among repeat elements, chromatin accessibility was most enriched at SVAs in human embryos from the 2-cell to blastocyst stage (L. Li et al., 2018). Stage-specific expression of SINEs and LINEs was observed in human (Yandlm & Karakulah, 2019) and bovine (Halstead et al., 2020) embryos, with species-specific dynamic changes (Oomen et al., 2025). Interestingly, expression of DNA transposons was reported in early embryos of several mammalian species, including bovine (Oomen et al., 2025). Expression of satellite DNA, such as HSATII and ACRO1, was also demonstrated at the cleavage stage embryos (Hendrickson et al., 2017) with the marked expression of most satellites at the 4-cell stage human embryos (Yandlm & Karakulah, 2019).

The studies described above present the relationship between mammalian preimplantation embryos and repetitive DNA sequences. Although some of the studies demonstrate the potential roles of transcripts derived from repeats, our understanding of these transcripts in the context of early embryo development is still limited.

2.7 Epigenetic regulation in preimplantation embryos

During preimplantation embryo development, drastic changes occur not only in the transcriptome but also in the epigenome, which is the state of the genome, characterized by chemical modifications to the DNA and histone proteins, along with the chromatin structure and high-order genome organization. In an analogy, the epigenome was regarded as an electrical licensor, whose correct setup is needed for the electrical switch to function, just as a properly configured epigenome is required for transcription or EGA to occur (Zou et al., 2024).

2.7.1 Mouse

Histone H3 K4 tri-methylation (H3K4me3), a histone modification mostly indicative of transcriptional activity (Ruthenburg et al., 2007), is typically found as sharp peaks in the promoters, enriched at their CpG-rich regions. In mice, uncommon broader H3K4me3 peaks, referred to as non-canonical H3K4me3 (ncH3K4me3), were observed not only in MII oocytes but also in maternal alleles of zygotes and early 2-cell stage embryos, which were replaced by canonical H3K4me3 following major EGA at the late 2-cell stage, but not in the presence of α -amanitin (B. Zhang et al., 2016). Allelic reprogramming was also observed for the paternal genome, which depleted its H3K4me3 peaks genome-wide following fertilization, keeping this epigenetic state until major EGA (B. Zhang et al., 2016). The paternal allele of the genome regained the H3K4me3 peaks at post-EGA stages in a similar profile observed in somatic cells (B. Zhang et al., 2016), demonstrating the relationship between epigenetic reprogramming and EGA.

Histone H3 K27 tri-methylation (H3K27me3) is another histone modification deposited by Polycomb repressive complex 2 (PRC2), a member of Polycomb group (PcG) proteins (Lanzuolo & Orlando, 2012). H3K27me3 is typically found at promoters with strong enrichment in somatic cells, in late blastocysts and post-implantation embryos in mice (Zheng et al., 2016). These canonical H3K27me3 profiles were also shown to exist in sperms, but they were globally lost in the promoters and distal regions of the paternal genome after fertilization (Zheng et al., 2016). Similar H3K27me3 depletion was also observed at the promoters but not at the distal regions of the maternal genome (Zheng et al., 2016). These distal regions with H3K27me3 mark were referred to as non-canonical (ncH3K27me3) (Q. Xu & Xie, 2018). Unlike H3K4me3 reprogramming (B. Zhang et al., 2016), the reset of H3K27me3 profiles during early development was found not to be dependent on EGA (Zheng et al., 2016).

Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) analysis in mice embryos following minor but before major EGA showed globally low-intensity permissive chromatin with the exception of loci with murine endogenous retrovirus L (MERVL), where the strongest ATAC-seq enrichment was observed (J. Wu et al., 2016). These accessible regions in the proximity of MERVLs, which produce transcripts with non-definitive 3' ends, disappeared upon blocking minor EGA, resulting in a state similar to post-EGA stages (J. Wu et al., 2016). The function of this promiscuous transcription of MERVLs confined to pre-minor EGA stages in mice is yet to be understood. Following major EGA, global repressive chromatin was observed with open chromatin getting restricted to enhancers and promoters of developmentally regulated genes, correlating with their expression (J. Wu et al., 2016). Allelic analysis of open chromatin revealed similar accessible regions in both parental genomes after EGA, with a minor portion of the genome following allele-specific chromatin accessibility (J. Wu et al., 2016).

2.7.2 *Drosophila* and Zebrafish

In *Drosophila melanogaster*, Zelda, a maternal factor with a major role in EGA, reduces nucleosome occupancy at regions where it binds, allowing enhancer-containing DNA regions wrapped around nucleosomes to become accessible and aid transcription (Sun et al., 2015). Opening of Zelda-bound regions is thought to induce the expression of the very first set of zygotic genes (Schulz et al., 2015).

In zebrafish (*Danio rerio*), both sperm and oocytes showed canonical H3K4me3 and H3K27me3 profiles, but they differed in terms of DNA methylation status of their putative enhancers (B. Zhang et al., 2018), with the former exhibiting a fully methylated pattern and the latter showing typical partial methylation of distal regulatory regions (Stadler et al., 2011). Following fertilization, enhancers also became hypermethylated (B. Zhang et al., 2018). Regarding acetyl histone modifications, histone H3 K27 acetylation (H3K27ac) peaks found at promoters in gametes were retained in pre-EGA embryos, whereas those found at enhancers in gametes were lost soon after fertilization, indicating a loss of epigenetic memory (B. Zhang et al., 2018). Promoters with H3K27ac peaks kept during pre-EGA embryos were referred to as primed because of accompanying DNA hypomethylation and histone variant H2AFV (Murphy et al., 2018), signs of a permissive chromatin state. Based on these findings, a proposed model suggests that primed promoters found in pre-EGA embryos resolve into either active or repressed promoters upon EGA (B. Zhang et al., 2018).

2.7.3 Human

Open chromatin analysis using human three pronuclei (3PN) zygotes, which are abnormal embryos with three sets of chromosomes, revealed high and low accessibility regions at CpG-rich and CpG-poor promoters, respectively (J. Wu et al., 2018). A correlation between promoter openness and both gene expression and CpG density was observed (J. Wu et al., 2018). This is in line with the hypothesis that suggested mammalian CpG-rich promoters were key elements in allowing Pol II to perform transcription (Fenouil et al., 2012). ATAC-seq peaks found at the distal regions in a stage-specific manner were enriched in TF motifs, allowing the identification of DUX4 as a minor EGA factor and GSC, PITX1, TFAP2C, and GATA proteins as major EGA factors (J. Wu et al., 2018). A notable finding of this analysis was the widespread open chromatin observed at the 2-cell stage, with the majority of the distal peaks overlapping DNA hypomethylated domains at human oocytes (J. Wu et al., 2018), a modification strongly correlated with open chromatin. It suggests that these distal ATAC-seq peaks are inherited from the oocyte to the 2-cell stage. The finding that 86% of promoter ATAC-seq peaks at the 2-cell stage were also found at the 8-cell stage, although these loci were only expressed during the latter suggests that pre-EGA embryos establish a primed chromatin state, where TFs and transcriptional machinery remain ready for activation with access to DNA and initiate transcription during EGA (J. Wu et al., 2018). Pre-EGA distal ATAC-seq peaks were lost following

EGA, whereas CpG-rich promoter ATAC-seq peaks remained, with the overall changes in open chromatin during EGA being dependent on transcription (J. Wu et al., 2018).

H3K4me3 modifications were observed as canonical strong peaks at promoters, as opposed to ncH3K4me3 seen in mice (B. Zhang et al., 2016), both in GV and MII human oocytes, which spread to distal CpG-rich and hypomethylated regions after fertilization in pre-EGA human embryos (Xia et al., 2019). Half of these promoters with strong H3K4me3 peaks were activated at the 8-cell stage, while the other half were inactivated following EGA (Xia et al., 2019). Increased H3K4me3 and open chromatin at the 4-cell stage were described as priming the genome for EGA, which is then resolved at the 8-cell stage (Xia et al., 2019). Similar priming was also described at distal regulatory regions, including enhancers, where sites with H3K4me3 and open chromatin at the 4-cell stage were either marked by H3K27ac and remained active after EGA or became marked by H3K27me3 and remained poised for potential future activation (Xia et al., 2019). This epigenetic transition, referred to as priming-resolving, was dependent on transcription (Xia et al., 2019), demonstrating the interconnectedness of epigenetic and transcriptomic changes during human early embryo development.

H3K27me3, which is found at the promoters of developmental genes (Schwartz & Pirrotta, 2007), was seen at human oocytes but globally lost at the 8-cell stage embryos (Xia et al., 2019). Global depletion of H3K27me3 was delayed on the maternal allele, which still retained H3K27me3 marks present in oocytes at the 2-cell stage, whereas those from sperm were already lost at the same stage (Xia et al., 2019).

Profiles of cytosine methylation at the CpG dinucleotides go through dynamic changes during early embryo development. An early study that was based on reduced representation bisulfite sequencing (RRBS) showed global erasure of methyl marks from the paternal genome (Smith et al., 2014). However, RRBS can detect 5–10% of the CpG dinucleotides in the genome, requiring a method with better coverage, such as whole genome bisulfite sequencing (WGBS). WGBS demonstrated that high methylation levels of the human gametes were lost at the blastocyst stage, primarily in the paternal and, to a lesser extent, in the maternal genome (Okada et al., 2014). The exceptions to this demethylation event were observed at the imprinted differentially methylated regions (DMRs) and gene bodies, as well as at repeat sequences (Okada et al., 2014). A more advanced technique, single-cell DNA methylome sequencing, revealed *de novo* DNA methylation at numerous genomic loci along with previously observed genome-wide methylation (P. Zhu et al., 2018), which is a phenomenon also observed in rhesus monkey preimplantation embryo development (F. Gao et al., 2017). *De novo* methylation occurring at thousands of genomic sites during human early embryo development was transient, lost at the blastocyst stage, and happened in two waves, one of which coincided with human EGA (P. Zhu et al., 2018). *De novo* methylation at the time of EGA might be a strategy of the embryo to prevent the firing of non-EGA genes, as inactive promoters are known to be hypermethylated. To better distinguish the differences in maternal and paternal genome

methylation patterns, single-cell DNA methylome sequencing was performed on uniparental embryos, namely, gynogenetic and androgenetic human blastocysts (J. Xu et al., 2021). Androgenetic blastocysts, which carry genetic material only from the father, showed a greater loss of global methylation at CpG dinucleotides as compared to the gynogenetic blastocysts, embryos with genetic contribution coming only from the mother (J. Xu et al., 2021). This finding provides another level of evidence for the greater demethylation the paternal methylome undergoes. Overall, these findings suggest a dynamic balance with gains and losses of the methyl group throughout preimplantation development.

2.7.4 Bovine

Early immunostaining experiments for 5-methyl cytosine showed unequal signals between the parental pronuclei, with the paternal pronuclei lacking a DNA methylation signal (Lepikhov et al., 2008). WGBS analysis of *in vivo* and *in vitro* bovine developmental stages identified two main groups. One group contained sperm with overall high DNA methylation profiles, whereas the other contained oocytes and embryos, which had low levels of methylation (Duan et al., 2019). The contrast in DNA methylation status between the sperm and embryos is due to the global demethylation, which occurs after fertilization and continues until the 8-cell stage. Following major EGA, a two-fold increase of global DNA methylation signal was detected. The DNA methylation around the TSS across all stages was low, in line with the general unmethylated states of promoters. In the human genome, early estimates pointed to around 60% of promoters being associated with CpG islands, most of which do not exhibit dynamic changes and are always unmethylated except for those that show tissue-specific methylation (Bernstein et al., 2007). Therefore, as authors also pointed out, the global DNA methylation dynamics presented in this study across bovine developmental stages reflect mostly the noncoding regions (Duan et al., 2019). Notably, immunostaining for H3K9me2, which is thought to protect against DNA demethylation (Rose & Klose, 2014), showed signal in the maternal but not in the paternal pronucleus (Lepikhov et al., 2008).

Open chromatin landscape of bovine preimplantation stages showed a notable change from the 4-cell to 8-cell stage in terms of ATAC-seq peaks, indicating chromatin relaxation possibly to enable major EGA (Halstead et al., 2020). Overall chromatin opening progressed between the 2-cell and morula stages, with many peaks confined to a single stage and some cumulatively observed across consecutive stages. Halstead et al. (2020) found that chromatin remodeling requires transcription, as they did not observe chromatin remodeling in the absence of transcription. This suggests that a reciprocal relationship may exist between transcription and chromatin remodeling, indicating a potentially interdependent mechanism.

Immunostaining for H3K4me3 demonstrated a speckled pattern in the paternal pronucleus (Lepikhov et al., 2008), whereas the signal in the maternal pronucleus was more spread. In contrast, an independent immunostaining of H3K4me3

showed signal only for the maternal pronucleus (L. Yang et al., 2025). Higher resolution of H3K4me3 profiling across several bovine developmental stages was also performed (Lu et al., 2021; C. Zhou et al., 2023). Similar to mouse oocytes (B. Zhang et al., 2016), ncH3K4me3 signals were also observed in bovine oocytes (Lu et al., 2021) as well as in bovine pre-EGA stages (C. Zhou et al., 2023). Following major bovine EGA, broad ncH3K4me3 peaks were replaced by canonical narrow H3K4me3 peaks (C. Zhou et al., 2023).

Interestingly, profiling of H3K27ac also revealed broad peaks for bovine pre-EGA stages, but the signal intensity was much less compared to H3K4me3 (C. Zhou et al., 2023). Histone H3 K9 tri-methylation (H3K9me3) signals were mostly observed at gene bodies prior to bovine major EGA, after which intergenic signals were detected (C. Zhou et al., 2023). Notably, immunostaining for H3K9me3 and H3K27me3 showed signal only in the maternal pronucleus of a bovine IVF zygote (L. Yang et al., 2025). H3K27me3 profiling showed an intergenic distribution in oocytes, followed by an intragenic distribution in pre-EGA stages, and again an intergenic distribution in post-EGA stages (C. Zhou et al., 2023). The post-EGA stages, which showed increased intergenic H3K27me3 genome coverage (C. Zhou et al., 2023), exhibited higher global signal intensity than the pre-EGA stages in H3K27me3 immunostaining (C. Zhou et al., 2019). Furthermore, the reported loss of H3K27me3 signal in the 4-, 8-, and 16-cell stages (Lu et al., 2021) was suggested to result from the low number of cells used in the experiment rather than biological effects (C. Zhou et al., 2023). These studies show that the epigenome during early development is highly dynamic and its relationship with early cellular processes is yet to be explored.

2.8 Roles of PRD-like homeobox genes in preimplantation embryo development

2.8.1 Homeobox genes

Homeobox genes belong to a gene superfamily (Holland, 2013), which typically contains a 180-bp sequence, coding for a homeodomain with a length of 60 amino acid. This length can range from a minimum of 54 to a maximum of 97 amino acids in length (Bürglin, 1994). While the homeobox gene was initially discovered in *Drosophila melanogaster* (McGinnis et al., 1984), it was shown later to exist in the genomes of all metazoans, including vertebrates, plants, and fungi (Gehring et al., 1994). Even protists were shown to contain genes encoding homeodomain (Bürglin, 2011; Derelle et al., 2007). Homeobox gene superfamily is ancient and arose earlier than the evolutionary divergence of animals, plants, and fungi (Ryan et al., 2006). Some homeobox genes can be organized in the genome in clusters such as the HOX cluster, a group of genes regulating body axis formation, and can be present as a single cluster or multiple clusters depending on the clade (Bürglin, 2011). Although many extensively studied homeobox genes were identified as homeotic genes, which are involved in body

patterning, homeobox and homeotic are distinct terms (Bürglin, 2011). The former describes a genomic sequence feature, whereas the latter describes a function during development. Functions beyond the homeotic roles for homeobox superfamily genes have been defined (Bürglin, 2011). Indeed, homeobox genes are mostly known for their roles in embryo development and differentiation through modulating gene expression as they encode transcription factors (TFs) with DNA-binding abilities. This ability is facilitated by the homeodomain that is composed of three alpha helices. Nuclear magnetic resonance and X-ray crystallography experiments show that helix 3 is responsible for the sequence-specific DNA contacts and is hence also known as the recognition helix (Bürglin, 2011). Position 50 of the homeodomain or position 9 of helix 3 has been shown to be critical for sequence specificity, as its change to another amino acid altered the homeodomain-recognized DNA sequence (Hanes & Brent, 1989). Positions 48 and 51 are tryptophan (W) and asparagine (N), and conserved in almost all homeodomains (Bürglin, 2011).

2.8.2 Classification of homeobox genes

Thousands of homeobox genes were identified in various organisms among metazoans (Bürglin, 2011) since the very first homeobox gene was discovered (McGinnis et al., 1984). The more homeobox genes were identified, the more variants of the homeodomain were discovered, which necessitated regular updates on the classification of these genes (Bürglin, 2011; Bürglin & Affolter, 2016; Gehring et al., 1994; Holland, 2013; Holland et al., 2007). Classification of the homeobox genes primarily relies on the phylogenetic trees generated by comparing homeodomains (Bürglin & Affolter, 2016). A secondary factor for their classification is the types of domains and motifs present in addition to homeodomain (Bürglin & Affolter, 2016).

For gene classification, three main terms are generally used with the following hierarchical order: superclass, class, and family (Holland, 2013). These terms help understand the evolutionary history of homeobox genes, but for the usage of these terms, there is no consensus, and it is possible to see slightly different classifications for the same gene. For example, all homeobox genes as a whole can be grouped under a superclass by some authors (Holland, 2013; Holland et al., 2007), whereas others prefer grouping several homeobox gene classes under the TALE superclass and do not categorize homeobox as a superclass (Bürglin & Affolter, 2016). Nevertheless, similar homeobox gene classes were defined. One of the initial comprehensive phylogenetic analyses described ten distinct homeodomain classes in animals (Ryan et al., 2006). Another study focusing on human homeobox genes classified eleven homeobox gene classes (Holland et al., 2007). A more extensive survey of homeobox genes described 16 homeobox gene classes: *ANTP*, *PRD*, *PRD-like*, *POU*, *HNF*, *CUT*, *LIM*, *ZF*, *CERS*, *PROS*, *SIX/SO*, *IRO*, *MKX*, *TGIF*, *PBC*, and *MEIS* (Bürglin & Affolter, 2016). Some of these gene classes contain only a few gene families, highlighting their conserved roles, whereas some gene classes, such as *ANTP* and *PRD*, contain tens of genes,

implying an expansion of gene number and possible diversification in their roles. Eight gene families have been reported to be lost in humans but present in fruit fly, indicating evolution of genes resulting in species-specific traits (Ryan et al., 2006).

2.8.3 Classification of PRD and PRD-like class

One of the largest classes of homeobox genes is the Paired or PRD class. In the proteins of this class, in addition to that of homeodomain, another helix-turn-helix motif, indicative of a DNA-binding domain, was identified (Burri et al., 1989). Later, its crystal structure in complex with DNA revealed one more helix-turn-helix motif, resulting in the identification of PAI and RED subdomains located in the N- and C-terminus of the PRD class proteins, respectively (W. Xu et al., 1995). *PAX* (Paired-box) gene family is a known evolutionarily conserved member of PRD class genes (B. Thompson et al., 2020).

A group of proteins with close similarity to the PRD class proteins based on homeodomain primary sequence but without a PRD domain also exists. While some classify these proteins under the PRD class (Holland, 2013; Holland et al., 2007), others consider them as a separate class named PRD-like or PRDL (Bürglin, 2011; Bürglin & Affolter, 2016). As opposed to PRD class with three helix-turn-helix motifs, PRDL class proteins contain one or two helix-turn-helix motifs, either of which is formed by the homeodomain but not a separate domain. Those who consider PRD as a broader class describe three families based on the residue at position 50 of the homeodomain: S50- or Pax-, K50-, and Q50-type (Galliot et al., 1999). On the other hand, those who consider PRDL as a distinct class classify it based on the absence of a serine residue at position 50 (Bürglin, 2011). In this case, the K50-, and Q50-type proteins fall under the PRDL class (Katayama et al., 2018). The presence of an R-type (Yaşar et al., 2024) PRDL protein was also shown. It is thought that the fusion of an ancestral Q50-type PRDL gene with a gene containing a PRD domain-encoding sequence gave rise to the *PAX* gene family following a Q50 to S50 transition (Galliot et al., 1999).

2.8.4 PRD-like class transcription factors from an evolutionary perspective

One of the initial reports on the evolution of PRDL class genes revealed four genes belonging to this family and named them: *ARGFX*, *DPRX*, *TPRX1*, and *DUXA* (Booth & Holland, 2007). Although initial evidence suggested the presence of a pseudogene called *TPRX2P* in the human genome (Booth & Holland, 2007), later experimental evidence (Madissoon et al., 2016) supported its designation as *TPRX2*. Expression of these four genes in human germ cells suggested their involvement in processes related to reproduction (Booth & Holland, 2007). Two other members of PRDL homeobox genes, *DUXB* and *LEUTX*, were discovered and named in a separate study (Holland et al., 2007). No orthologues

of these six genes were found in the mouse genome (Booth & Holland, 2007; Holland et al., 2007; Royall et al., 2018; Zhong & Holland, 2011). In bovines, *DPRX* (Maeso et al., 2016) and *DUXB* (Leidenroth & Hewitt, 2010) were identified as pseudogenes. In addition, an extra *TPRX* gene, named *TPRX3*, was identified in the bovine genome (T. D. Lewin et al., 2021). An example of a PRDL homeobox gene that is found in the genomes of humans, bovines, and mice is *NOBOX*, which was first identified in growing oocytes and was attributed a role in folliculogenesis (Suzumori et al., 2002). Because of their suggested role in early embryo development, a subset of PRDL class homeobox genes, specifically, *ARGFX*, *DPRX*, *LEUTX*, and *TPRX*, were categorized under a single term and are collectively referred to as the Eutherian Totipotent Cell Homeobox (ETCHbox) genes (Maeso et al., 2016).

Although homeobox genes are highly conserved, an exception was noted for highly divergent ETCHbox genes (T. D. Lewin et al., 2021). Gene duplication of *CRX*, another homeobox gene, followed by sequence evolution, is suggested to have given rise to the ETCHbox genes (Maeso et al., 2016). Notably, an Alu element overlapping with an exon of *ARGFX* was identified in humans (Booth & Holland, 2007). This exon is thought to emerge via the exonization process (G. Li & Holland, 2010), which occurs through the accumulation of mutations at Alu elements, resulting in functional splice sites (Gal-mark et al., 2008). Bovine *ARGFX* has one less exon than human *ARGFX*, which is likely because of the exonization mediated by primate-specific Alu elements (Kriegs et al., 2007).

2.8.5 *DUX* family genes

DUX family genes are thought to be derived from a single homeobox *DUX* (*sDUX*) gene, found in organisms such as opossum, wallaby, and platypus (Leidenroth & Hewitt, 2010). Duplication of *sDUX* in placental mammals gave rise to *DUXA*, *DUXB*, *DUXC*, and *Duxbl*, all of which are considered paralogous as they arose through gene duplication (Clapp et al., 2007). *Duxbl* is specific to rodents, but remnants of homologous sequences can be found in other placental mammals (Leidenroth & Hewitt, 2010). These paralogous genes contain introns, but *DUX* genes without any introns also exist (Figure 10).

The emergence of intronless *DUX4* and *Dux* is hypothesized to be the result of two separate retrotransposition events (Leidenroth & Hewitt, 2010). *DUXA* was initially proposed as the origin of *DUX4* (Booth & Holland, 2007). In another study, either *DUXC* or *Duxbl* was suggested to be the ancestor of *DUX4* (Snider et al., 2010). Subsequent evidence presented stronger support for *DUXC* as the source of these retrotransposition events. Retrotransposition or retroposition is the mechanism by which a gene is duplicated through the reverse transcription of an mRNA derived from the gene to be duplicated (Kaessmann et al., 2009). Retrotransposed genes are generally considered nonfunctional unless the genomic integration site allows nearby promoter acquisition, rendering them retrogenes (Holland et al., 2007). Since the presence of mRNA synthesized from the gene that is retroposed is required at the time of retrotransposition, and because this

event must occur in germ cells for the altered genetic information to be passed down, *DUXC* is expected to be expressed in oocytes and/or sperm cells. In connection with this, *DUXC* expression has been demonstrated in canine testis (Snider et al., 2010). Whether the full-length *DUXC* mRNA is synthesized in germ cells is yet to be confirmed.

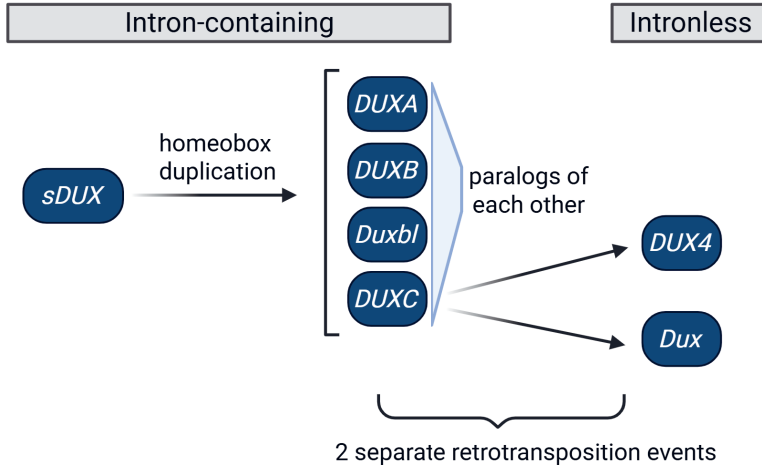


Figure 10. Evolution of *DUX* genes. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/9k0s08q>.

An initial model suggested that *DUXC* and *DUX4* were present in a common ancestor (Clapp et al., 2007). However, with the availability of improved genome assemblies, this model was updated, proposing only *DUXC* as the gene found in the common ancestor of *DUX4*, *DUXC*, and *Dux* (Leidenroth et al., 2012). *DUXC* and *DUX4* are mutually exclusive across placental mammal genomes, suggesting that they act as counterparts in different organisms. Similarly, *Dux* is thought to be the rodent equivalent and does not co-exist in the genome with either *DUXC* or *DUX4* (Leidenroth et al., 2012).

There are two main lines of evolutionary evidence supporting the functional similarity of *DUXC* and *DUX4* (Leidenroth & Hewitt, 2010). First, the C-terminal domains of both proteins are conserved across placental mammals based on alignments of their primary protein sequences. Second, a maximum likelihood tree generated using homeodomains forms a distinct branch for *DUX4* and *DUXC*, away from *DUXA*, *DUXB*, *DUXBL*, and *DUX*.

Both *DUXC* and *DUX4* are found in multiple tandem copies, with the genomic region containing *DUX4* specifically referred to as the D4Z4 array. Even though homologous sequences of the D4Z4 array have been identified in primates, no homologues have been found in more distantly related species, including bovine (Clapp et al., 2007). The reason for this is likely the introns that are present in *DUXC* but not in *DUX4*. Furthermore, since retrotransposition results in the birth of retrogenes at positions unrelated to their original locus, no syntenic relationship is retained between such genes, including *DUX4* and *DUXC*.

2.8.5.1 History of *DUX4*

DUX4 ORF was, for the first time, predicted in the D4Z4 array of a facioscapulohumeral muscular dystrophy (FSHD) patient in 1999 (Gabriëls et al., 1999). Several years before this finding, investigations on the D4Z4 locus in humans had already begun (Deutekom et al., 1993; Hewitt et al., 1994; Wijmenga et al., 1992), which enabled its characterization and ultimately made the identification of *DUX4* possible (Gabriëls et al., 1999), see Figure 11. Human D4Z4 locus was initially shown to be flanked by EcoRI restriction sites, whose excision released a DNA fragment containing the entire D4Z4 array (Wijmenga et al., 1992). FSHD was first characterized by the length of these cut out EcoRI fragments that showed differential migration between healthy individuals and individuals with FSHD, detected by Southern blot using a probe called p13E-11, annealing to one end of this locus (Wijmenga et al., 1992). The length of the EcoRI fragments was found to be shorter in FSHD patients (Wijmenga et al., 1992). The following year, EcoRI fragments derived from random FSHD patients revealed a detailed map of these fragments, revealing tandemly repeated internal 3.2 kb units, and the sequence on the other end of the array was detected by Southern blot using a probe called pLAM1 probe (Deutekom et al., 1993). p13E-11 and pLAM1 nomenclatures commonly used in the literature for annotating the D4Z4 locus originate from these early hybridization experiments. These sequences are likely specific to humans and are inappropriate terms for the characterization of D4Z4 homologues in non-primate species.

Sequencing of one D4Z4 tandem repeat led to the prediction of 12–100 internal repeat copies in this locus of healthy individuals (Hewitt et al., 1994). The presence of two homeobox sequences in the D4Z4 array was also reported, but these were initially thought not to encode a protein (Hewitt et al., 1994). *DUX4* was first identified and named after the sequencing of a DNA fragment containing only two D4Z4 tandem repeats of FSHD patient origin, revealing a genomic sequence of 13.5 kb in length (Gabriëls et al., 1999). This study identified a putative *DUX4* promoter with TACAA and GC boxes. Mutating these motifs reduced luciferase reporter expression, indicating their importance for *DUX4* expression (Gabriëls et al., 1999). *In vitro* transcription and translation assays produced the *DUX4* protein product that formed a dimer with a molecular weight of 75 kDa (Gabriëls et al., 1999). However, this was an artificial way of producing *DUX4* mRNAs and proteins, and until 2007, the expression of *DUX4* in a human tissue had not been reported (Dixit et al., 2007), see Figure 11.

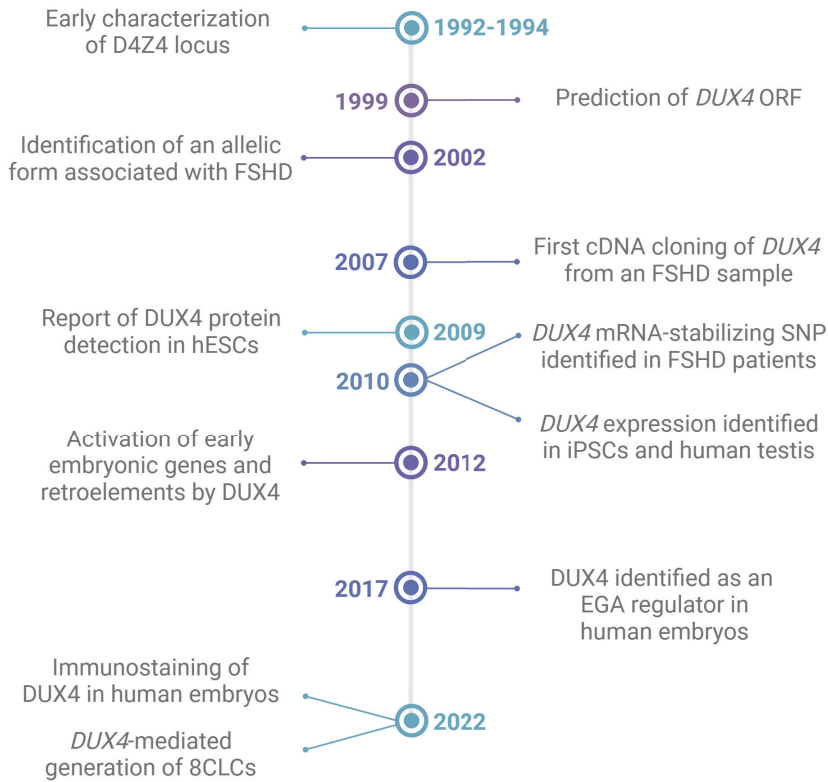


Figure 11. Key findings on *DUX4* in chronological order. This figure was created in BioRender. Yaşar, B. (2026) <https://BioRender.com/kael40s>.

The search for genes responsible for developing FSHD phenotype paved the way for the identification of a specific allelic form of the D4Z4 locus associated with this disease (Lemmers et al., 2002; Van Geel et al., 2002). The first cDNA cloning of *DUX4* was performed using primary myoblasts of FSHD patients and revealed two isoforms of *DUX4* with one or two introns, both of which were located downstream of the *DUX4* coding sequence (CDS) in exon 1 (Dixit et al., 2007). The cloned *DUX4* cDNA was derived from the most distal D4Z4 repeat unit, extending into the downstream polyadenylation signal, which was thought to mediate the manifestation of FSHD by stabilizing these *DUX4* transcript isoforms absent in control samples (Dixit et al., 2007; Snider et al., 2010). Later, an SNP downstream of the last D4Z4 repeat was shown to contribute to the polyadenylation-mediated stabilization of *DUX4* transcripts in FSHD patients, resulting in a toxic gain of function (Lemmers et al., 2010). Even though *DUX4* expression was also observed in control human myoblasts, it was a short isoform, unlike the full-length *DUX4* detected in myoblasts derived from individuals with FSHD (Snider et al., 2010). Interestingly, one in every thousand cultured FSHD muscle cells was estimated to express full-length *DUX4*, whose ectopic expression in healthy human muscle cells caused apoptosis, unlike the short isoform of *DUXC* (Snider et al., 2010).

The finding that linked *DUX4* with early development was made in 2009, where the full-length isoform of *DUX4* expression was shown through RT-PCR in human embryonic stem cells (hESCs) (Snider et al., 2009). Using the same experimental approach, human testis and induced pluripotent stem cells (iPSCs) were later also shown to express the full-length isoform of *DUX4* (Snider et al., 2010). Earlier studies had focused more on the toxicity effect of *DUX4*, considering its connection with the FSHD manifestation (Bosnakovski et al., 2009; Bosnakovski, Xu, et al., 2008) until results supporting its early developmental role were reported. Specifically, *DUX4* was found to activate retroelements and early embryonic genes such as *TPRX1* and *ZSCAN4* (Geng et al., 2012). *DUX4* binding motif identified by chromatin immunoprecipitation followed by sequencing (ChIP-seq) was shown to be bound directly by *DUX4* using electrophoretic mobility shift assay (EMSA) (Geng et al., 2012). Detecting the expression of its target genes *in vivo* in human testis demonstrated the biological relevance of *DUX4* activity, identifying it as a transcriptional activator in the context of early development and as a gene misexpressed in the context of FSHD pathophysiology (Geng et al., 2012). Following these reports that suggest a link between *DUX4* and early development, analyses of RNA-seq datasets derived from preimplantation human embryos demonstrated the first direct evidence for the role of *DUX4* in EGA, as presented in 2017 by two distinct published studies (De Iaco et al., 2017; Hendrickson et al., 2017) and a work submitted as a preprint (Töhönen et al., 2017). Identification of *DUX4* binding motifs in the promoters of cleavage stage genes (De Iaco et al., 2017; Hendrickson et al., 2017), the activation of *DUX4* target genes (J. M. Whiddon et al., 2013) in early embryos, and the activation of EGA genes in *DUX4*-overexpressed hESCs (De Iaco et al., 2017) suggested that *DUX4* drives early embryo development by functioning as a transcriptional activator. Direct activation of a cleavage stage gene *ZSCAN4* by *DUX4* was also demonstrated using luciferase reporter (Hendrickson et al., 2017). Furthermore, evidence for the *DUX4*-mediated activation of human endogenous retroviruses (HERVs) (Hendrickson et al., 2017; J. L. Whiddon et al., 2017), a definitive feature of cleavage stage human embryos (Göke et al., 2015), was provided. Immunostaining analysis of early human embryos supported prior findings with protein-level data and revealed nuclear enrichment of endogenous *DUX4* protein in the 2- and 4- but not 8-cell stage, suggesting its main activity prior to or during EGA (Vuoristo et al., 2022). Furthermore, pulsed induction of *DUX4* expression reprogrammed hESCs to an 8-cell-like stage (Yoshihara et al., 2022). These functional experiments support the role *DUX4* plays in driving human EGA.

The findings summarized here present a sequence of publications in a chronological order to demonstrate how *DUX4*, which was traditionally researched in the context of the human disease FSHD, later turned out to be a critical player in human preimplantation embryo development and became the focus of developmental biology research.

2.8.5.2 *DUX4* mRNAs and proteins in early human embryos

DUX4 is expressed only within the first 48 hours of embryo development, but it initiates various transcriptional programs that its effect continues after its transcription halts (Vuoristo et al., 2022). A similar effect of *DUX4* was also observed in multinucleated FSHD myotubes, where *DUX4* is initially found in only a few nuclei, diffuses into neighboring nuclei, and activates its target genes that remain detectable even after *DUX4* becomes undetectable (Rickard et al., 2015).

During human early embryo development, the peak expression of *DUX4* was observed at the zygote stage (Vuoristo et al., 2022) but an earlier report showed the highest *DUX4* expression at the 4-cell stage (Hendrickson et al., 2017). The former relied on an RNA-seq dataset prepared with a 5'-end-based method (Töhönen et al., 2015), while the latter utilized a 3'-end-based method (Yan et al., 2013). In addition, quantification of the latter did not benefit from spike-in normalization unlike the former. The transcriptome in early developmental period is highly dynamic due to maternal mRNA degradation as well as cell divisions, which halve the volume of blastomeres and their cellular content (Töhönen et al., 2015). Therefore, strategies such as the equal addition of exogenous RNA to each sample for normalization of measured transcripts help prevent biases arising from this dynamic nature of early embryonic development and provide more reliable quantification. Furthermore, methods that mitigate additional biases (Boskovic et al., 2023), such as PCR duplication through the use of UMIs (Kivioja et al., 2012), may further increase the quantification accuracy and enable better transcriptome characterization.

DUX4 proteins detected by antibody staining showed elevated signals at the 2- and 4-cell stage with a clearance at the 8-cell stage (Vuoristo et al., 2022). The knockdown of *DUX4* in 3PN human zygotes, which was confirmed by a sharp decrease in *DUX4* antibody staining, neither caused major changes in the transcriptome of early embryos nor embryonic arrest (Vuoristo et al., 2022).

2.8.5.3 *DUX4* and repeat elements

Induction of *DUX4* expression in human primary myoblasts followed by ChIP-seq revealed *DUX4* binding sites in the mammalian apparent LTR-Retrotransposons (MaLRs) (Geng et al., 2012). MaLRs or ERVL-MaLRs are traces of ancient retroviral infections and fall under the non-autonomous LTR retrotransposons (Liao et al., 2023; Smit, 1993). Direct binding of *DUX4* to ERVL-MaLR elements was confirmed by EMSA (Geng et al., 2012). In line with this, enhancer-like regions activated by *DUX4* were shown to contain ERVL-MaLR elements (Vuoristo et al., 2022). Enhancer activation by *DUX4* was validated using CRISPR activation that targets ERVL-MaLR elements-containing enhancer region of *ZSCAN4* (Vuoristo et al., 2022), a known target gene of *DUX4* (Geng et al., 2012). Reanalysis of an earlier ChIP-seq dataset (Geng et al., 2012) with a newer genome assembly revealed that approximately two-thirds of all *DUX4* binding sites were in a repetitive element, with one-third overlapping with

ERVL-MaLR elements (J. M. Whiddon et al., 2013). DUX4-bound repetitive elements were more conserved than the flanking sequences among placental mammals and primates, indicating the functional importance of these nucleotide positions (J. M. Whiddon et al., 2013). Satellite repeat family, such as HSATII subtype, was enriched among DUX4 binding sites, and DUX4-mediated transcription of HSATII was demonstrated (J. M. Whiddon et al., 2013). Similarly, expression of ACRO1 and HSATII satellite repeats was observed in cleavage stage embryos as well as in *DUX4*-overexpressed cells (Hendrickson et al., 2017). Bidirectional expression of HSATII was confirmed by RNA fluorescence *in situ* hybridization (RNA-FISH) in *DUX4*-expressing myoblasts (Shadle et al., 2019). An overlap of HSATII RNA-FISH signal and immunofluorescence from double-stranded RNA (dsRNA) antibody in the nucleus was reported, implying the formation of a nuclear foci composed of dsRNAs as a result of HSATII expression (Shadle et al., 2019). Sequencing of DUX4-induced nuclear dsRNAs after their pull-down proved the presence of dsRNAs (Shadle et al., 2019). Demonstrating the presence of dsRNAs was enabled by dsRNA editing mediated by ADAR enzymes in the cell, resulting in the detection of A-to-G mismatches between the RNA-seq read and reference genome at positions where ADAR enzymes edited adenine to inosine (Bahn et al., 2012). ADAR-edited dsRNAs were also detected in early embryos with endogenous *DUX4* expression (Shadle et al., 2019), suggesting that it is a molecular event occurring during normal embryo development.

2.8.5.4 DUX4 and nonsense-mediated mRNA decay

One of the pathways through which RNA gets degraded is the nonsense-mediated mRNA decay (NMD) pathway (Lykke-Andersen & Jensen, 2015). RNA isoforms with a premature stop codon positioned upstream of splice junctions are good substrates for the NMD pathway (Lykke-Andersen & Jensen, 2015). DUX4 was shown to inhibit NMD, causing levels of such RNA isoforms to rise, including its own RNA, which was identified as an NMD substrate (Feng et al., 2015). One component of NMD pathway is UPF1 (L. Huang et al., 2011). A time course experiment showed a reciprocal decrease of UPF1 in response to an increase of DUX4 protein levels after its induced expression (Feng et al., 2015). DUX4-mediated UPF1 degradation was proteasome-dependent and proposed to be the route by which DUX4 inhibited NMD (Feng et al., 2015). Inhibition of NMD has also been linked to nuclear dsRNA foci that form in response to DUX4. One NMD factor, the exon junction complex component EIF4A3, was suggested to be sequestered by these dsRNA foci, contributing to NMD inhibition (Shadle et al., 2017).

2.8.5.5 Toxicity of DUX4

The toxicity of *DUX4* was initially reported through its ectopic expression in multiple cell lines (Kowaljow et al., 2007). The same effect was also shown in mouse myoblast cells, in which *DUX4* repressed genes that regulates glutathione redox pathways (Bosnakovski, Xu, et al., 2008). Byproducts created by cellular respiration are normally reduced by antioxidants, components of glutathione redox pathways, preventing oxidative stress. *DUX4* expression was proposed to make cells more vulnerable to oxidative stress, contributing to cell toxicity, as stress-inducing agents showed more cell death in *DUX4*-expressing cells than controls, which was reversed by the addition of antioxidants (Bosnakovski, Xu, et al., 2008). Notably, dual induction of *DUX4* with either mouse *Pax3* or *Pax7* rescued the toxicity caused by *DUX4* (Bosnakovski, Xu, et al., 2008). *In vivo* toxic effect of *DUX4* was demonstrated by introducing its mRNA into the developing zebrafish embryos (Mitsuhashi et al., 2013). Interestingly, injection of *DUX4* mRNA encoding a C-terminal domain (CTD)-truncated form of *DUX4* did not show the same effect, indicating that the *DUX4* CTD is responsible for the toxicity (Mitsuhashi et al., 2013). Similarly, induction of *DUX4c*, which is an ORF located 42 kb upstream of the *D4Z4* locus with an inverted orientation and encodes the same protein as *DUX4*, but with a truncated C-terminus, did not result in cytotoxicity (Bosnakovski, Lamb, et al., 2008). A correlation was observed between the CTD-mediated transcription activation capacity and the toxicity (Choi et al., 2016). High-throughput screen of 44,000 small molecules to prevent *DUX4*-mediated cell death showed that most of the candidate compounds protected against cytotoxicity by enhancing the ability of cells to combat oxidative stress (Bosnakovski et al., 2014). *DUX4* toxicity was shown to be dependent on p53 (Wallace et al., 2011), which was later confirmed in a different experimental setting (Feng et al., 2015). Conversely, in other cell types, the toxicity effect was not dependent on p53, indicating cell-type specific regulation (Shadle et al., 2017).

MYC and dsRNA defense pathways were identified after a small interfering RNA (siRNA) screen targeting about 7,000 genes in the rhabdomyosarcoma cell line with induced *DUX4* expression aimed to find partners, together with *DUX4*, that contribute to cytotoxicity (Shadle et al., 2017). MYC is a known inducer of apoptosis or programmed cell death pathway (McMahon, 2014). *DUX4* stabilized *MYC* mRNAs, likely as a result of NMD inhibition, which in turn upregulated pro-apoptotic genes such as *BCL2L1*, resulting in cell death, a phenotypic outcome of cytotoxicity (Shadle et al., 2017). Antibody against dsRNAs showed signal in some but not all nuclei of *DUX4*-expressing cells, probably due to asynchronous expression of ectopic *DUX4* (Shadle et al., 2017). Therefore, another possible route for cell death was thought to be mediated by the antiviral defense pathway activated through nuclear dsRNA, which is a typical signal for viral infection (Shadle et al., 2017). Since apoptosis can be a part of a developmental process, as observed in organs such as testis (Snider et al., 2010; J. M. Whiddon et al., 2013) and thymus (Das & Chadwick, 2016), where

DUX4 expression was observed, it was suggested that one of the normal functions of *DUX4* could be mediating apoptosis (Shadle et al., 2017). Another link between *DUX4* and toxicity was formed through HSATII, whose knockdown resulted in reduced *DUX4*-mediated dsRNA and toxicity (Shadle et al., 2019). Moreover, *DUX4* caused impaired DNA repair through expression of HSATII, which scavenged the complex composed of KDM2A/B and PRC1, causing global loss of H2AK119Ub signal and the prevention of 53BP1 recruitment, a protein responsible for DNA repair (Arends et al., 2024).

A link between the *DUX4*-mediated cytotoxicity and NMD was suggested (Campbell et al., 2023). Translation of isoforms with premature termination codons that would normally be degraded by NMD resulted in the accumulation of truncated proteins, many of which were RNA-binding proteins (RBPs) with a potential toxicity effect (Campbell et al., 2023). This effect was experimentally validated for one of the truncated RBPs, specifically SRSF3 (Campbell et al., 2023), proposing another route through which *DUX4* exerts its cytotoxicity.

Forced continuous expression of *DUX4* results in toxicity, but in early embryos, it is transiently expressed (De Iaco et al., 2017; Hendrickson et al., 2017; Vuoristo et al., 2022). To observe the toxicity of *DUX4*, whose expression was intended to mimic the early embryonic period, a 4h to 6h pulse of *DUX4* expression was induced in a myoblast cell line, causing a moderate level of cell death (Resnick et al., 2019). Similarly, the cytotoxic effect of *DUX4* was mostly overcome when a pulse of 15 min *DUX4* expression was induced in hESCs (Yoshihara et al., 2022). However, constant or 1h induction of *DUX4* in hESCs caused a widespread cell death, with the 30 min induction improving cell survival but not as much as the 15 min induction (Yoshihara et al., 2022). These results suggest that the true biological relevance of *DUX4* can be captured when its expression is induced transiently, similar to its natural expression, which lasts only a short period of time during early development. Furthermore, these findings highlight that ectopic expression of *DUX4* in distinct cell lines, such as undifferentiated hESCs and lineage-committed myoblasts, can lead to cytotoxic effects to different extents, likely due to the distinct chromatin architecture of these cell types.

2.8.5.6 The effect of *DUX4* on translation

Comparison of *DUX4*-expressing myoblasts and controls uncovered distinct changes at the protein and mRNA levels, suggesting the inefficient translation of produced transcripts (Jagannathan et al., 2019). Methionine incorporation assay confirmed that *DUX4* caused a slowdown in protein translation (Jagannathan et al., 2019). In another study, polysome profiling demonstrated a higher 80S fraction and fewer polysomes in *DUX4*-expressing cells compared to controls (Campbell et al., 2023), validating the earlier report. *DUX4*-mediated fragmentation of the Golgi apparatus was also observed (Jagannathan et al., 2019). *DUX4* was shown to suppress not only translation initiation but also elongation (Hamm et al., 2023). The suppression of the former happens through dephosphorylation

of 4E-binding protein 1 (4EBP1), which in its hypophosphorylated form sequesters eukaryotic initiation factor 4E (eIF4E), while that of the latter occurs by phosphorylation-mediated inhibition of eukaryotic elongation factor 2 (eEF2) (Hamm et al., 2023). Polysome profiling and ribosome footprinting showed mostly reduced polysome association and loss of 5' ribosome occupancy in *DUX4*-induced myoblast cells, respectively (Hamm et al., 2023). Those mRNAs with increased polysome association, with multiple ribosomes sitting on these RNAs, decreased in protein levels, indicating suboptimal translation machinery (Hamm et al., 2023). While *DUX4* was shown to inhibit translation, translation of its target genes was not affected, which is thought to be because of the less structured 5'-UTRs these genes have (Hamm et al., 2023). Such a translation inhibition effect of *DUX4* might be a modulating the translational machinery in early embryos to prevent translation of maternal mRNAs until they are degraded.

2.8.5.7 C-terminal domain of DUX4

Immunoprecipitation of *DUX4* followed by mass spectrometry revealed two main gene activators as binding partners of *DUX4*: p300 and CBP (Choi et al., 2016). *DUX4* CTD, which is conserved with *DUXC* (Clapp et al., 2007), was identified as an essential region for transcription activation and p300 interaction (Choi et al., 2016). Identifying shared downregulated genes mediated by *DUX4* and its CTD-truncated form suggested that CTD is only responsible for activation but not repression activity of *DUX4* (Choi et al., 2016). Additional evidence for the connection of *DUX4* and transcriptional machinery came from a study in which *DUX4* binding partners were identified (Vuoristo et al., 2022). Several subunits of the multisubunit Mediator complex, which relays transcription activation signals from TFs to Pol II (X. Chen et al., 2021), were identified as interaction partners of *DUX4* (Vuoristo et al., 2022). Specifically, MED15 was experimentally shown to be recruited through the KIX binding motif found in *DUX4* CTD (Vuoristo et al., 2022).

DUX4 binding sites were enriched with H3K27ac signals, which had a drop in the center of its profile, indicative of nucleosome displacement (Choi et al., 2016). Although many *DUX4* binding sites showed nucleosome displacement after *DUX4* binding as a result of its pioneer activity, some of these sites were already accessible prior to *DUX4* binding (Choi et al., 2016), indicating that *DUX4* binds to both accessible and inaccessible regions in the genome. Notably, pioneer activity for *DUX4* was also suggested in another study (De Iaco et al., 2017).

Two of the Ewing-like sarcoma patients contained translocation-mediated *DUX4* fusions, bringing the *DUX4* C-terminus together with the high mobility group (HMG) box protein CIC (Kawamura-Saito et al., 2006). The transcription repressor activity of CIC was reversed, causing the activation of CIC target genes, which indicates the capacity of the C-terminal region of *DUX4* in transcription activation (Kawamura-Saito et al., 2006). Moreover, co-immunoprecipitation with *DUX4* CTD revealed the top binding partner as STAT1 in *DUX4*-expressing

myoblasts (Spens et al., 2023). The (L)LxxL(L) motif in *DUX4* CTD was shown to be responsible for binding to STAT1 and suppression of interferon-gamma-stimulated genes, which was facilitated by preventing Pol II recruitment (Spens et al., 2023).

2.8.5.8 Further characterization of *DUX4*

Upon induced expression of *DUX4* in human myoblasts, expression of histone variants H3.X and H3.Y was observed, and subsequent incorporation of these histones at the *DUX4* target genes increased the mRNA levels and sustained the expression of *DUX4* target genes (Resnick et al., 2019). Histone variants H3.X and H3.Y were found to be more associated with euchromatin (Wiedemann et al., 2010). It is possible that *DUX4* not only mediates transcription activation directly through its CTD but also indirectly by promoting relaxed chromatin with the help of H3.X and H3.Y. Interestingly, immunofluorescence signals from H3.X/Y in the 8- but not in 4-cell stage human blastomeres were also reported (Taubenschmid-Stowers et al., 2022). The relevance of this potential mechanism in early embryos needs to be demonstrated experimentally.

In healthy muscle cells and iPSCs, the *DUX4* locus was shown to be kept in a repressed state by nucleosome remodeling deacetylase (NuRD) and chromatin assembly factor 1 (CAF-1) (Campbell et al., 2018). Integration of MBD3L family proteins into the NuRD complex was demonstrated to hinder its DNA binding ability and reverse *DUX4* repression, resulting in the expression of *DUX4* in FSHD muscle cells (Campbell et al., 2018). However, whether these findings are also true in early embryo setting is yet to be investigated.

An association between *DUX4* expression and cancer was also made. Analysis of close to 10,000 samples of cancer showed expression of *DUX4* in many solid cancer types, including ovary, bladder, thymus, esophagus, kidney, breast, soft tissue, lung, stomach, cervix, endometrium, and testicular germ cell cancer (Chew et al., 2019). Of note, detected *DUX4* levels in cancer samples ranged from two to ten transcripts per million (Chew et al., 2019). Inability to previously identify *DUX4* as a cancer gene was thought to stem from its repetitive multi-copy nature, making its detection challenging (Chew et al., 2019). *DUX4* lowered major histocompatibility complex class I RNA expression (Chew et al., 2019) and translation (Hamm et al., 2023), which was suggested to be a way of evading the immune system in some cancers.

2.8.5.9 Human *DUX4* homologue mouse *Dux*

An equivalent of human *DUX4* in mouse is *Dux*, both of which were suggested as the major driver of EGA (De Iaco et al., 2017). Similar to *DUX4* (Vuoristo et al., 2022), *Dux* is also expressed at the onset of major EGA and activates the transcription of EGA-related genes (De Iaco et al., 2017; Hendrickson et al., 2017). Its ectopic expression converted mouse embryonic stem cells (mESCs) into a 2-cell-like cell (2CLC) state (Macfarlan et al., 2012), which is also observed

spontaneously in some cells of cultured mESCs with a dependency on *Dux* (De Iaco et al., 2017; Hendrickson et al., 2017). Which TFs activate the expression of DUX family EGA-inducing proteins has been an intriguing question.

Dux-mediated EGA induction in mouse was suggested to be initiated by inherited maternal factors (Hendrickson et al., 2017). A screen through cloning of genes involved in epigenetic and chromatin regulation in mouse oocytes and/or zygotes and through their subsequent ectopic expression in mESCs was made (Eckersley-Maslin et al., 2019). This screen identified DPPA2 and DPPA4 as factors resulting in 2CLCs through activation of *Dux* expression, which was supported by ChIP-seq results revealing binding of these factors to *Dux* promoters (Eckersley-Maslin et al., 2019). DPPA2 and DPPA4 were also found to be necessary for the formation of *Dux*-expressing 2CLCs (De Iaco et al., 2019). Although 2CLCs are a practical model to study EGA in cell culture because of their continuous growth capacity, replacing the need for culturing IVF embryos, maternal and maternal-zygotic knockout (KO) experiments showed that DPPA2 and DPPA4 were not essential in mouse embryos for EGA to occur (Z. Chen et al., 2021).

A comparison of homeobox gene expression profiles between mESCs and totipotent blastomere-like cells (TBLCs) (H. Shen et al., 2021) revealed *Obox4* as a gene with ten times higher expression in the latter (Y. Guo et al., 2024). Because of its capacity to activate *Dux* as well as most of its target genes, and the compromised embryonic development observed in *Dux-Obox4* double-KO embryos but not in *Obox4* KO embryos, OBOX was suggested to act as a redundant factor of DUX (Y. Guo et al., 2024). Similarly, OBOX family proteins were shown to regulate mouse EGA by contributing to chromatin opening that facilitates Pol II loading at regulatory elements (Ji et al., 2023). Even though *Obox* gene family is absent from the human genome, it is syntenic to human *TPRX2* (Maeso et al., 2016).

Another maternal factor aiding 2CLC formation through expression of *Dux* was found to be NELFA (Z. Hu et al., 2020). However, the binding of DPPA2/4 and NELFA to *Dux* promoters was proposed to be indirect, as no specific binding motif was suggested (Grow et al., 2021). Therefore, researchers continued their search for alternative TFs that can activate *Dux*.

In light of the finding demonstrating the DNA-damaging agent-induced *Zscan4* expression (Storm et al., 2014), which is also directly induced by DUX (Hendrickson et al., 2017), mESCs were treated with DNA-damaging agents and exhibited signs of 2CLCs, including the expression of *Dux* (Grow et al., 2021). It raised the possibility of involvement of a factor both in *Dux* expression and the DNA-damage response (DDR) pathway. Through depletion experiments, p53, a member of DDR, was found to be necessary for *Dux* expression in 2CLCs that are formed by DNA-damage stimulation of mESCs (Grow et al., 2021). Activation of *Dux* expression by p53 was also demonstrated in the absence of DPPA2 and DPPA4 (Grow et al., 2021), which are other inducers of *Dux* (Eckersley-Maslin et al., 2019). In contrast, elimination of p53 from early mouse embryos *in vivo* through a maternal-zygotic KO experiment still showed DUX-positive blastomeres at the early 2-cell stage (Grow et al., 2021). It implies that several

factors may contribute to *Dux* expression, and the players involved in *Dux* activation might differ between *in vivo* and *in vitro* conditions. In other words, 2CLCs might not fully reflect the 2-cell stage mouse embryos, and results obtained from 2CLCs must be interpreted with caution.

In search of finding a relevance for p53 in a human setting, an LTR10C element bearing a p53 binding site was found to be necessary for the expression of *DUX4* using a human cell model (Grow et al., 2021). Knocking out *TP53*, the gene encoding P53, prevented significant activation of *DUX4* in hESCs (X. Zhang et al., 2023), supporting the involvement of P53 as a factor that acts upstream of *DUX4*. However, it is uncertain yet whether *DUX4* expression through binding of p53 to LTR10C element is a mechanism favored in early human embryos *in vivo*.

Both in human and mouse embryos, shortening and elongation of telomeres were observed at minor and major EGA, respectively (X. Zhang et al., 2023). In mESCs, *Zscan4* was previously found to be responsible for telomere elongation needed for retaining genomic stability and normal karyotype, a typical feature of mESCs unlike many other cell lines (Zalzman et al., 2010). However, silencing of *Zscan4* in mouse embryos did not prevent the elongation of telomeres observed during major EGA (X. Zhang et al., 2023). To identify the causal effect of telomere elongation, the researchers used low doses of α -amanitin that allowed EGA to occur but impaired telomere elongation (X. Zhang et al., 2023). In embryos with impaired telomere elongation but with partial EGA, *Dux* expression was observed, indicating that telomere elongation is necessary for *Dux* repression (X. Zhang et al., 2023).

2.8.5.10 The role of *Dux* and its timely repression in mouse embryos

Even though *Dux* was initially reported to be necessary for early embryo development (De Iaco et al., 2017), KO experiments targeting the *Dux* locus in mice suggested that the role of *DUX* in mouse EGA is not essential (Z. Chen & Zhang, 2019; De Iaco et al., 2020).

Similar to human *DUX4*, mouse *Dux* is found in a locus with macrosatellite repeats (Clapp et al., 2007). Downstream of the *Dux* cluster, a truncated version of *Dux*, named *Gm4981*, is located and transcribed during oogenesis (Z. Chen & Zhang, 2019). Depletion of this entire locus by zygotic CRISPR-Cas9 injection was performed, resulting in mono- and bi-allelic deletions of the targeted genomic region, neither of which impaired the preimplantation embryo development (Z. Chen & Zhang, 2019). To assess its effect on full development, the same approach was used, initially producing *Dux* heterozygous pups with no developmental defects (Z. Chen & Zhang, 2019). Crossing the resulting *Dux* heterozygous offspring yielded zygotic KO pups without abnormalities (Z. Chen & Zhang, 2019). However, since the oocytes that formed this offspring were heterozygous and might have contained expressed *Gm4981* transcripts, which are expressed in oocytes and might have been carried out to the *Dux* KO embryos, maternal and zygotic KO pups were produced after crossing of zygotic *Dux* KO

embryos (Z. Chen & Zhang, 2019). Zygotic or maternal and zygotic *Dux* KO embryos did not show embryonic arrest, and both gave rise to pups that can reach adulthood with lower developmental potential and litter size (Z. Chen & Zhang, 2019).

Similar observation was also made by another group demonstrating that *Dux* depletion does not inhibit the full development of pups, but it significantly reduces litter size (M. Guo et al., 2019). It was suggested that the primary role of DUX is not to initiate EGA but to promote it by activating the expression of its target genes (M. Guo et al., 2019).

Another finding was that the timely clearance of *Dux* transcripts was crucial for early embryo development, as prolonged expression of *Dux* led to developmental arrest (M. Guo et al., 2019). Consistently, silencing of *Dux*, which was mediated through LINE1 RNA, was shown to be needed to exit the 2-cell stage (Percharde et al., 2018). This is in line with another study demonstrating the need for *Dux* repression for the embryo to continue past the EGA (X. Zhang et al., 2023). Another protein, CCCTC-binding factor (CTCF), which is a zinc-finger TF that mediates binding within and between chromosomes to define certain genomic boundaries (Ong & Corces, 2014), was suggested to counteract the 2-cell program in mESC (Olbrich et al., 2021). Compelling evidence for the involvement of the mouse- and rat-specific factor DUXBL (Leidenroth & Hewitt, 2010; S. L. Wu et al., 2010) in facilitating exit from the totipotent 2-cell stage was recently provided. Based on the proposed mechanism (Vega-Sendino et al., 2024), *Duxbl* is expressed in response to DUX, whose DNA-bound regions become accessible to the DUXBL protein, which then initiates a repressive program by interacting with cofactors such as TRIM33/TRIM24, known to interact with TRIM28 (Herquel et al., 2011). Notably, TRIM28, also known as Kap1, was earlier shown to mediate 2-cell exit via repression of *Dux* (Percharde et al., 2018).

A connection among nucleoli, *Dux* loci, and 2-cell exit was also made. Exit from the 2-cell stage was associated with *Dux* repression through the formation of mature nucleoli, as the knockdown of nucleolar components such as RNA Pol I, Fibrillarin, and Nucleolin in mESCs promoted *Dux* expression and 2-cell program (Xie et al., 2022). Since nucleoli are not matured and only appear as nucleolar precursor bodies until the late 2-cell stage embryos, maturation of nucleoli following major EGA was suggested to be an event that pulls *Dux* loci physically inward for repression (Xie et al., 2022). Although nucleoli do not have membranes, the association of *Dux* loci with this compartment was ensured through liquid-liquid phase separation (Xie et al., 2022). The mechanism proposed for *Dux* repression and exit from the 2-cell stage is insightful (Xie et al., 2022), however, in the absence of *Dux*, embryos may simply bypass the need for *Dux* repression, or other factors compensate for the loss of *Dux* function, given that they can still progress through the 2-cell stage, as evidenced by viable *Dux* KO pups (Z. Chen & Zhang, 2019; M. Guo et al., 2019).

These reports are insightful to understand the role of DUX in EGA and embryo development, however, it is noteworthy to mention that EGA is a species-

specific event (Lu et al., 2021) and many players, which are transiently expressed during human early development (Töhönen et al., 2015), are absent in the mouse genome (Zhong & Holland, 2011). This suggests that distinct regulatory networks might exist in different species during early development. Therefore, the insights gained from *Dux* KO experiments cannot be directly applied to human embryo development for *DUX4*. Although *DUX4* knockdown in human embryos caused only minimal changes (Vuoristo et al., 2022), whether the complete depletion of *DUX4* causes defects in preimplantation embryo development is not yet known.

2.9 Roles of PRDL TFs in early embryo development

2.9.1 Identification of PRDL TFs in early embryo development

An early study that performed RNA sequencing of more than 300 blastomeres and oocytes identified 32 and 129 genes significantly upregulated during minor and major EGA, respectively, and highlighted the involvement of specific TF families during this early developmental period in humans (Töhönen et al., 2015). Three distinct TF families, which can bind to the promoters of genes upregulated during major EGA, were identified: PRD-like (PRDL), T-box, and basic leucine zipper (bZIP) (Töhönen et al., 2015). Focusing on PRDL homeobox genes, several of which were transiently upregulated in cleavage stage embryos, a time point in development earlier than the appearance of Yamanaka factors (Takahashi & Yamanaka, 2006), seven previously poorly annotated novel TF genes, namely *ARGFX*, *CPHX1*, *CPHX2*, *DPRX*, *DUXA*, *DUXB*, and *LEUTX*, were identified (Töhönen et al., 2015). Based on the stage at which their transcription was most pronounced, PRDL homeobox TF genes were referred to as maternal (*GSC*, *HESX1*, *ISX*, *NOBOX*, *OTX2*, *PITX2*, *RAX2*, and *TPRXL*) or embryonically activated (*ARGFX*, *DPRX*, *DUXA*, and *LEUTX*) (Töhönen et al., 2015). Genes with transcription activity at both maternal and embryonic stages were categorized as mixed (*CPHX1*, *CPHX2*, *DUXB*) (Töhönen et al., 2015).

Alu elements overlapping the promoters of many genes upregulated during EGA were identified, one of which was *ZSCAN4*, an activated gene during the oocyte-to-4-cell transition (Töhönen et al., 2015). Luciferase assay in HEK293 cells demonstrated the activation capacity of *ZSCAN4* promoter containing Alu elements, with the cotransfection of *DUXA* increasing the signal slightly (Töhönen et al., 2015). However, a later study using human iPSCs showed that cotransfection of *DUX4* increased the luciferase signal, but deleting proximal Alu elements upstream of *ZSCAN4*, did not affect its activation (Hendrickson et al., 2017). Furthermore, cotransfection of *DUXA* did not yield any activation (Hendrickson et al., 2017), as opposed to the earlier report (Töhönen et al., 2015). As different cell types were used for the luciferase assay in each study, the observed differences may be due to cell type-specific effects. Of note, targeting Alu element-containing regulatory regions using CRISPR-Cas9-based gene activation (CRISPRa), together with targeting the promoters of core reprogramming factors (*OCT4*, *SOX2*, *KLF4*, *MYC*, *LIN28A*) in human foreskin fibroblasts,

improved the efficiency of iPSC derivation (Weltner et al., 2018). A similar approach was also successfully applied to a more differentiated cell line, human lymphoblastoid cell lines, resulting in efficient iPSCs generation and highlighting the contribution of Alu elements to early developmental gene regulatory programs (Sokka et al., 2022). Similarly, Alu elements containing PLAG1 binding motifs were enriched within a *de novo* DNA motif present in the promoters of 74 out of the 129 genes (Töhönen et al., 2015) upregulated during human major EGA (Madissoon et al., 2019). Functional experiment in mice, in which *Plagl1*-KO embryos were generated, showed that PLAG1 regulated embryonic development through binding to the regulatory elements (Madissoon et al., 2019), specifically, B1 elements, the rodent counterparts of human Alu elements (Labuda et al., 1991).

2.9.2 Molecular cloning of PRDL TF homeobox genes

In light of the TSS information obtained from the earlier dataset (Töhönen et al., 2015), several PRDL class transcripts were isolated (Tang et al., 2010) from cDNA libraries derived from human ovary or 8-cell blastomeres (Madissoon et al., 2016). Molecular cloning of these cDNA fragments, which involved inserting them into cloning plasmids and propagating them in bacteria to determine their full-length sequences by Sanger sequencing, confirmed the presence of and revealed the structures of *ARGFX*, *CPHX1*, *CPHX2*, *DPRX*, *DUXA*, *DUXB*, *NOBOX*, *TPRX1*, and *TPRX2* (Madissoon et al., 2016).

Experimental validation for *ARGFX* mRNA derived from human testis was performed earlier using 3' RACE and RT-PCR (G. Li & Holland, 2010). Several other splice variants of *ARGFX* were identified later (Madissoon et al., 2016). Similarly, molecular cloning of *NOBOX* was performed using human oocytes, and its detection in early embryos required higher PCR cycles, suggesting the presence of only maternal residual transcripts at these stages (Huntriss et al., 2006). Molecular cloning revealed a previously unexplored *NOBOX* isoform (Madissoon et al., 2016). The same study also reported novel splice variants for *TPRX1* and *TPRX2* (Madissoon et al., 2016), which were predicted earlier (Booth & Holland, 2007).

To characterize the roles of these cloned PRDL homeobox genes, they were overexpressed in hESCs, and the genes up- or downregulated by each PRDL factor were identified. Based on the number of genes each factor regulated, they were identified as activators or repressors. *DPRX*, which downregulated the highest number of genes, was identified as a repressor, while *CPHX1* and *CPHX2* acted as activators (Madissoon et al., 2016).

Molecular cloning of *LEUTX* reported in a separate study enabled identification of a novel first exon (Jouhilahti et al., 2016). Elevated expression levels of *LEUTX* were reported at the onset of human EGA using publicly available datasets (Töhönen et al., 2015; Xue et al., 2013; Yan et al., 2013), with its peak expression observed at the 8-cell stage across multiple human early embryo RNA-seq datasets (Jouhilahti et al., 2016; Petropoulos et al., 2016; Töhönen et al., 2015; Xue et al., 2013; Yan et al., 2013).

2.9.3 Characterization of PRDL TF homeobox genes

2.9.3.1 *LEUTX*

LEUTX (*leucine twenty homeobox*) gene was named after the leucine residue at the highly conserved position 20 of *LEUTX* homeodomain (Holland et al., 2007). Before *LEUTX* was discovered, TAATCC was identified as a core consensus motif bound by the PRDL homeodomain protein Otx in sea urchin (Mao et al., 1994). Similarly, human OTX2, a K50-type PRDL TF, bound the same core motif (Jolma et al., 2013). As another K50-type PRDL TF, *LEUTX* was shown to activate transcription through the TAATCC motif. This was demonstrated by the loss of *LEUTX* transcriptional activity when the interaction between the motif and protein was perturbed by the K57A mutation, which was confirmed in structural modeling (Katayama et al., 2018). Screening *LEUTX* in human genotype databases revealed that no homozygous mutations affecting its functional role were found, suggesting a critical role in embryo development (Katayama et al., 2018). However, knockdown of *LEUTX* in human 3PN embryos had only minor effects on EGA (Zou et al., 2022), implying that compensatory factors might exist.

Overexpression of the novel isoform of *LEUTX* in hESCs activated the majority of EGA genes, whereas *DPRX* overexpression caused downregulation of those genes (Jouhilahti et al., 2016). A model was suggested, in which *LEUTX* expression from the embryonic genome aids in the activation of EGA genes, with one of these genes potentially being *DPRX*, which initiates a repression program to balance the transcriptome of the embryo and ensure its developmental progression (Jouhilahti et al., 2016).

Ectopic expression of *LEUTX* was also performed in more differentiated cells. Cell integrity of human fibroblasts was reported to be disrupted upon *LEUTX* overexpression (Maeso et al., 2016). The overlap between genes downregulated by both *LEUTX* and *TPRX1* and genes upregulated by *ARGFX* included 124 genes, suggesting a regulatory network in which *LEUTX* and *TPRX1* take on repressor roles, counteracted by the activator role of *ARGFX* (Maeso et al., 2016). Although this is in contrast with the earlier model (Jouhilahti et al., 2016), the roles of genes can differ depending on the distinct regulatory pathways in which they participate. In addition, the differences might be a result of the cell type used overexpression experiments. In a more recent study, overexpression of human and marmoset *LEUTX* in fibroblasts showed differences in downstream gene expression, which were attributed to the highly divergent nature of *LEUTX*, as shown by the presence of six residues under positive selection (T. D. Lewin et al., 2023). Earlier, residue 7 in the N-terminus of *LEUTX* homeodomain was shown to be under positive selection, indicating that slightly different amino acids at this position evolved in different lineages to maintain DNA binding in a context-dependent manner (T. D. Lewin et al., 2021). Structural modelling of *LEUTX* in complex with DNA suggested that the N7 residue in the N-terminal arm may contact the minor groove (T. D. Lewin et al., 2021).

Ectopic expression of bovine *LEUTX* and *ARGFX* in fibroblasts demonstrated that *LEUTX* changed the expression of a similar number of genes in both directions, whereas the *ARGFX* downregulated more genes than it upregulated, suggesting a repressive role for *ARGFX* (T. D. Lewin et al., 2022). A significant overlap was seen between the genes regulated directly or indirectly by *ARGFX* and *LEUTX* (T. D. Lewin et al., 2022), indicating their common functions, as opposed to their human counterparts (Maeso et al., 2016). Comparing the effects of human and bovine overexpression of PRDL class genes showed the highest overlap between genes downregulated by *LEUTX* overexpression in both species (Maeso et al., 2016). A high number of overlap was also observed for genes upregulated by human and bovine *LEUTX* (Maeso et al., 2016), pointing to a partial shared function between the two species.

Mass spectrometry-based analyses revealed five stable and 124 dynamic interaction partners of *LEUTX* (Gawriyski et al., 2023). Proteins forming stable interactions with *LEUTX* included EP300 (also known as p300) and CBP (Gawriyski et al., 2023), which contain a KIX domain that can interact with 9aaTADs (Piskacek et al., 2007; Radhakrishnan et al., 1997), a domain found near the C-terminus of *LEUTX*. EP300 and CBP are coactivators of transcription (Ogryzko et al., 1996), and their interaction with *LEUTX* supports its activator role suggested earlier (Jouhilahti et al., 2016). This interaction was lost when the 9aaTAD was deleted but retained when the K57A mutation was introduced (Gawriyski et al., 2023), which drastically reduced *LEUTX*-mediated transcriptional activation in luciferase assays (Jouhilahti et al., 2016). Similar to *DUX4* (Vuoristo et al., 2022), *LEUTX* also upregulated enhancers, many of which were shared with those activated by *DUX4* (Gawriyski et al., 2023). Comparison of *LEUTX*-bound sites identified by ChIP-seq with the open chromatin regions in early human embryos (J. Wu et al., 2018) showed that *LEUTX* binds accessible regions of the chromosome at the 8-cell stage (Gawriyski et al., 2023).

LEUTX was suggested to act the activator of porcine EGA, by opening chromatin at EGA genes and mediating histone acetylation (K. Zhou et al., 2024). In porcine expanded pluripotent stem cells, which are cultured cells containing heterogenous subpopulations with features of totipotency, naïve pluripotency or primed pluripotency, *LEUTX* was identified as a factor promoting totipotency (L. Wu et al., 2025). Overexpression of *LEUTX* in these cells induced 4-cell-like state, exhibiting totipotency signatures (L. Wu et al., 2025).

2.9.3.2 *TPRX* genes

There are two main *TPRX* genes in humans, *TPRX1* and *TPRX2*, whereas bovines have three: *TPRX1*, *TPRX2*, and *TPRX3* (T. D. Lewin et al., 2021). Although another *TPRX*-related gene, *TPRXL*, exists in humans, due to its unusual evolutionary relationship with *TPRX* genes, it was not designated as *TPRX3* and is categorized separately (Booth & Holland, 2007). Rodents do not have *Tprx* genes, which were initially thought to be due to gene loss (Wilming et al., 2015; Zhong & Holland, 2011), but a later report showed that duplication of *Tprx2* followed

by extensive sequence divergence gave rise to *Obox*, found in many copies and syntenic with human *TPRX2* (Maeso et al., 2016). Notably, OBOX proteins were identified as key regulators of mouse EGA (Ji et al., 2023).

Human *TPRX1* and *TPRX2* contain 9aaTAD involved in transcription activation (Gawriyski et al., 2024). In line with this, MED15, a protein relaying signals from TFs to Mediator complex (X. Chen et al., 2021), was found to be a binding partner of *TPRX2* (Gawriyski et al., 2024). Similarly, *TPRX1* was shown to interact with transcription coactivator EP300 (Gawriyski et al., 2024). Stable interaction between members of the SWI/SNF complex and *TPRX2* was also reported (Gawriyski et al., 2024). ChIP-seq for *TPRX2* showed that *TPRX2* binds to the genomic region close to *TPRX1* and *TPRX2* genes (Gawriyski et al., 2024), but whether *TPRX2* regulates *TPRX1* or its own expression is not yet known.

TPRX1 expression was initially shown to be induced by ectopic *DUX4* expression (Geng et al., 2012). Later, *TPRX1* was used as a marker gene to identify 8-cell-like cells (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022), in which *DUX4* was also expressed. A recent study, which suggests maternal PRDL TF OTX2 as an important regulator of human EGA, demonstrated OTX2-driven *TPRX1* and *TPRX2* activation (Q. Wang et al., 2025).

Knockdown of *TPRX1* and *TPRX2* was performed in human 3PN embryos using siRNAs. *TPRXL*, a *TPRX1*-derived gene (Booth & Holland, 2007), which was shown to be translated in oocytes (Zou et al., 2022), was also included in the knockdown experiments using morpholino. Individual knockdown of *TPRX1* and *TPRX2* had minor effects in the transcriptome, whereas moderate effects were observed for *TPRXL* knockdown (Zou et al., 2022). Severe developmental delay was observed when all three genes were simultaneously depleted at the 1-cell stage (Zou et al., 2022). Yet, the number of embryos that reached the morula stage was not zero. The transcriptomes of these morula stage embryos resembled those of the 8-cell stage, indicating a developmental compromise in the absence of *TPRX* factors, despite the developmental stage reached (Zou et al., 2022).

2.9.3.3 *DUXA* and *DUXB*

DUXA was proposed to be a competitive inhibitor of *DUX4*, with the ability to bind the same genomic regions as *DUX4* (Bosnakovski et al., 2023). Fusion of the activation domain VP64 (Beerli et al., 1998) to the C-terminus of *DUXA* converted this protein to a transcription activator and replicated the cytotoxicity effects exerted by *DUX4* (Bosnakovski et al., 2023). The activation capacity of *DUXA*-VP64 was demonstrated by a luciferase assay, which showed no activation for *DUXB*-VP64 (Bosnakovski et al., 2023). *DUX4*- and *DUXA*-VP64-mediated gene upregulation involved commonly regulated genes, including many EGA genes (Bosnakovski et al., 2023).

ChIP-seq experiment revealed *DUXB* binding in the *DUX4* genomic region (Gawriyski et al., 2024). A binding motif for *DUXB* was not previously available, and the *de novo* motif identified in the same study probably provided the first known binding motif for *DUXB* (Gawriyski et al., 2024). *DPRX*, *EGR2*, *KLF3*,

and CRX were suggested to regulate *DUXA* expression as they showed binding at *DUXA* promoter region (Gawriyski et al., 2024).

DUXA KO cloned bovine embryos show lower blastocyst formation rate, whereas its overexpression increased it (Shi et al., 2025), suggesting its role in modulating early development. In another study, knockdown of *DUXA* in bovine IVF embryos showed downregulation of major EGA genes, suggesting its role in EGA (J. Zhang et al., 2025). However, *DUXA* and *DUXC* show sequence similarity (Leidenroth & Hewitt, 2010), and whether siRNAs targeting *DUXA* caused degradation of *DUXC* mRNA in this study (J. Zhang et al., 2025) remains an open question.

2.9.3.4 ARGFX

ARGFX (*arginine-fifty homeobox*) gene is named after the arginine residue at position 50, which is very uncommon among homeobox genes (Booth & Holland, 2007). Because of having an unusual arginine residue at position 50 of the homeo-domain, *ARGFX* was predicted to have a distinct DNA binding motif (Booth & Holland, 2007).

A mutant version of *ARGFX* containing a 13 bp deletion that causes a frame-shift in the CDS was identified in bovines. This allele was homozygous in some Jersey bovine individuals, suggesting that *ARGFX* is either redundant or that its role is retained in a truncated form (T. D. Lewin et al., 2022). Ectopic expression of the mutant *ARGFX* was still able to cause downstream effects, but to a lesser extent form (T. D. Lewin et al., 2022). Similar to *LEUTX* (Gawriyski et al., 2023), *KIX* domain-containing transcription coactivators interacted with *ARGFX* but not *DPRX* (Gawriyski et al., 2024), which aligns with its previously suggested repressor role (Jouhilahti et al., 2016).

A maternal allelic ratio (MAR) of mRNAs in human embryos was suggested as a metric to judge the degree of MZT (Q. Guo et al., 2025). A high MAR indicates a higher fraction of maternal transcripts and is therefore negatively correlated with both EGA and MZT. The top two genes showing the strongest negative correlation between this ratio and gene expression were *DPRX* and *ARGFX*, highlighting them as potential markers of MZT progression. A cutoff value for MAR was defined, which classified 8-cell embryos with high-MAR as transcriptionally arrested status (TAS) embryos (Q. Guo et al., 2025). Both *DPRX* and *ARGFX* were downregulated in TAS embryos, suggesting that failure to activate these genes prevents the oocyte-to-embryo transition program from continuing and arrests embryonic development. Furthermore, separate knockdown of *DPRX* and *ARGFX* did not cause severe effects, implying redundancy between these two genes, as also supported by transcriptome profiling. Double knockdown of both *DPRX* and *ARGFX* compromised embryo development on day 4, suggesting their cooperative function in ensuring MZT.

2.10 Cell culture models for studying early embryo development

Two main states of hESCs were observed, depending on the cell culture medium in which the cells derived from the ICM of the blastocyst stage human embryos are maintained: the naïve and primed states. While the *in vivo* counterpart of the latter is post-implantation stage human embryos, the former is considered an equivalent of the epiblast cells in the blastocyst. However, none of these cell models are appropriate for studying EGA, which occurs at a much earlier time during preimplantation development. In order to provide a model system, by which EGA can be studied in a petri dish, several research groups have recently generated cells that resemble the transcriptome profiles of 8-cell-like cells (8CLCs) (Mazid et al., 2022; Moya-Jódar et al., 2023; Taubenschmid-Stowers et al., 2022; Yoshihara et al., 2022; X. Yu et al., 2022).

The first report of 8CLCs (Taubenschmid-Stowers et al., 2022) came from an observation of a rare cell population among naïve hESCs cultured under PXGL conditions (G. Guo et al., 2017). These cells showed elevated expression of human EGA marker genes, such as *LEUTX*, *ZSCAN4*, *DUXA*, *MDB3L3*, and *TRIM49* in two independent datasets (G. Guo et al., 2017; Macfarlan et al., 2012), making up approximately 1.5% of the entire hESC population in each dataset (Taubenschmid-Stowers et al., 2022). Transcript profiles of 8-cell embryos and 8CLCs were found to be similar, with the latter not observed in datasets derived from primed state ESCs or differentiated cells (Taubenschmid-Stowers et al., 2022).

The generation of 8CLCs was next demonstrated through a stepwise cell culturing process. For this purpose, cell culture medium with 4 chemicals and LIF (4CL) was defined (Mazid et al., 2022). Initially, primed hESCs were maintained in 4CL medium, containing H3K27 methyltransferase EZH2 inhibitor and histone deacetylase (HDAC) inhibitor, resulting in the formation of naïve hESCs (Mazid et al., 2022). Culturing these naïve cells in higher doses of both inhibitors for 5 days, referred to as enhanced 4CL (e4CL) medium, resulted in the generation of 8CLCs with the expression of totipotency genes such as *TPRX1* and *DUXA/B* (Mazid et al., 2022). In addition to the stepwise formation of e4CL-day5 cells with 8CLC features, culturing primed hESCs directly in e4CL medium for 7 days also gave rise to 8CLCs, named direct e4CL-day7 cells (Mazid et al., 2022). The former, which make up the 11.9% of cells in culture, showed higher similarity to blastomeres of the 8-cell stage embryos (Mazid et al., 2022). It was noted that the extended passaging of cells maintained in e4CL medium was not possible, probably due to the nature of totipotency that does not exhibit self-renewal and is distinct from it (Surani et al., 2007).

Chemical-induced 8CLCs (ci8CLCs) were the third report for the generation of 8CLCs. Firstly, human pluripotent stem cells (hPSCs) were screened with chemicals, resulting in the generation of epiblast-like hPSCs (prEpiSCs), which exhibited greater resemblance to epiblast cells of the blastocyst compared to other reported naïve hESCs (X. Yu et al., 2022). In order to convert prEpiSCs to

8CLCs, several chemicals, including H3K27 methyltransferase EZH2 inhibitor, a p53 activator, and apoptosis inhibitors were used, which led to the formation of ci8CLCs (X. Yu et al., 2022).

Transgene-mediated generation of 8CLCs was also reported. Induced expression of *DUX4* for 15 min in primed hESCs produced, 24 h post-induction, cells with a transcriptome similar to that of 8-cell embryos (Yoshihara et al., 2022). These 8CLCs, also referred to as induced blastomere-like (iBM) cells, made up 6.6% of the cultured cells, which increased to 17% following their enrichment through fluorescent-activated cell sorting using iBM cell-specific surface protein NaPi2b encoded by *SLC34A2* (Yoshihara et al., 2022). The iBM cells exhibited decreased expression of pluripotency marker genes (*SOX2* and *NANOG*) and increased expression of cleavage-stage genes (*LEUTX* and *DUX4*) (Yoshihara et al., 2022). With the disappearance of DUX4 protein at 48 h post-induction, iBM cells reverted to a state similar to the primed pluripotent state (Yoshihara et al., 2022).

Another group demonstrated 8CLCs as a subpopulation in human naïve induced pluripotent stem cells (iPSCs) cultured in 5iLAF (Moya-Jódar et al., 2023; Theunissen et al., 2014). This subpopulation formed 1.7% of the cultured cells analyzed (Moya-Jódar et al., 2023) and showed similarity to other 8CLCs generated by other groups (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022).

Following the report of 8CLC models from independent groups with distinct approaches, an integrative analysis containing datasets generated from these 8CLC models, as well as early human embryos, was performed (Yoshihara & Kere, 2023). After integrating these datasets, unsupervised clustering of single-cell transcriptomes identified clusters corresponding to human embryonic developmental stages (Yoshihara & Kere, 2023). Among all 8CLC models, the iBM cells were composed of cells, 99% of which were classified into the cluster of 8-cell stage embryos, whereas for other 8CLC models, this percentage ranged between 3%-25%, characterizing the iBM cells as the 8CLC model with the greatest similarity to 8-cell stage embryos (Yoshihara & Kere, 2023). A more recent integrative analysis (Paloviita et al., 2026) showed that among all 8CLCs, those generated through *DUX4* expression (Yoshihara et al., 2022) and through culture in e4CL medium (Mazid et al., 2022) most closely resembled 8-cell embryos. While the availability of such cell models is of great value for studying human EGA without the need for embryos, it is important to note that 8CLCs represent a distinct cell state that is very similar, but not identical, to 8-cell blastomeres.

3 AIMS

The objective of this thesis is to investigate embryonic genome activation (EGA) using the bovine as a model organism to gain insight into the molecular players involved in this critical process, through which the embryonic genome takes over the control of development. The studies presented here were aimed at contributing to the understanding of human and bovine preimplantation embryos development, which are relevant for improving *in vitro* fertilization (IVF) embryo practices used in humans for clinical purposes and in cattle for agricultural purposes. The specific aims of each study presented in the thesis are given below.

1. To identify and clone bovine orthologues of PRD-like homeobox genes whose expressions are transient and almost confined only to the time of human EGA.
2. To examine the 5'-end-focused transcriptional profiling of bovine oocytes and preimplantation embryos. More specifically, by focusing on the time of bovine EGA, we aimed to identify regulatory motifs of EGA genes and key regulators of bovine EGA.
3. To identify and isolate *DUXC* mRNA from the early bovine embryos for accurate gene annotation and characterization of its locus across several cattle breeds, and to assess the dependency of *DUXC* mRNAs on embryonic genome transcription by reanalyzing publicly available datasets derived from bovine early embryos treated with transcriptional inhibitors.
4. To assess the effect of oxygen levels during *in vitro* culture of bovine embryos on their transcriptome.

4 MATERIALS AND METHODS

4.1 Bovine IVF procedures and embryo culture (Study I, II, III, and IV)

Oocyte collection from bovine ovaries with subsequent *in vitro* maturation (IVM), *in vitro* fertilization (IVF), and embryo culture were performed in all studies as previously described (Org et al., 2019). Briefly, ovaries that were placed in a 0.9% sterile NaCl solution at 37 °C were transported from a slaughterhouse in Estonia to the IVF laboratory at the Estonian University of Life Sciences. The ovaries were washed twice with the same solution that they were kept in. Cumulus oocyte complexes (COCs) that are found in ovarian follicles of 2–8 mm size were aspirated with the help of a vacuum pump (Minitüb GmbH). Collected COCs were washed and matured in IVM medium, containing 0.05 µg/ml EGF, 0.8% bovine serum albumin (BSA), 1 mM L-glutamine, 0.5 mM pyruvate. For IVM of every 50 COCs, one well of four-well Nunc plates (Thermo Fisher Scientific) was used and incubated at 38.5 °C and 5% CO₂ for 22–24 h.

After thawing frozen Holstein bull semen, oocytes and sperms were co-incubated in IVF medium (Boskovic et al., 2025) at 38.5 °C and 5% CO₂ for 22–24 h. IVF medium contained 1 µM epinephrine, 10 µM hypotaurine, 30 µg/ml heparin, 0.6% BSA, 20 µM penicillamine, and 0.25 mM pyruvate. Following IVF and vortex-mediated removal of cumulus cells, zygote stage embryos were incubated in modified synthetic oviduct fluid (SOF), containing 0.2 mM L-glutamine, MEM amino acid solutions (Sigma-Aldrich), BME (Sigma-Aldrich), 0.8% BSA, and 0.25 mM pyruvate. The collected developmental stages included GV and MII oocyte as well as the zygote, 2-, 4-, 8-, and 16-cell embryo, and blastocyst. For study II and IV, single oocytes or embryos were collected, whereas for study I and III, pools of oocytes or embryos were used. Embryo culture was performed under hypoxia conditions for study I, II and III. Experimental design for study IV required culturing at normoxia and ultrahypoxia conditions in addition to hypoxia.

4.2 Ethical considerations (Study I, II, III, and IV)

Studies included in this doctoral thesis did not involve the use of live animals. The ovaries used for oocyte collection for IVF embryo culture were obtained post-mortem from a slaughterhouse in Estonia, and no ethical approval was required.

4.3 Gene prediction (Study I and III)

For the gene prediction in study I, reanalysis of a publicly available RNA-seq dataset (Graf et al., 2014) derived from bovine oocytes and early embryos, as well as evolutionary analysis were performed. On the Galaxy platform (Afgan et al., 2018), HISAT2 (Kim et al., 2019), StringTie (Pertea et al., 2015), and samtools

merge (H. Li et al., 2009) tools were used for mapping, transcript assembly, and merging replicates based on the developmental stage, respectively. Built-in genome assembly *bosTau9* and *ncbiRefSeq* gene annotation retrieved through UCSC Table Browser (Karolchik et al., 2004) using Apr. 2018 (*ARS-UCD1.2/bosTau9*) assembly were used. StringTie merge function was used to merge all files into one annotation file representative of bovine early embryo transcriptome. This resulting file was used to examine specific loci. These loci were determined through *tblastn* (Johnson et al., 2008) search against the bovine genome using putative primary human *PRDL* protein sequences (*ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, *TPRX1*, and *TPRX2*) identified in another study (Madissoon et al., 2016). Assembled transcripts at each identified locus were examined manually and confirmed to serve as predicted genes except for *TPRX2*, which required manual homeobox and splice site search to predict its structure. The *TPRX* locus was examined whether *TPRX3*, a previously suggested *TPRX* duplicate in bovine (T. D. Lewin et al., 2021), was predicted through a gene prediction approach. Upon confirmation, *TPRX3* was also included in our gene set.

Study III focused on a single gene, namely *DUXC*, which already contained several Ensembl gene predictions from release 104 (*ENSBTAT00000081668*, *ENSBTAT00000070667*, and *ENSBTAT00000080304*) and 112 (*ENSBTAT00000111847*). To find *DUXC* gene predictions, similar to study I, a homology-based *tblastn* search was used to find the bovine orthologue of porcine *DUXC* (Nip et al., 2023). Since these predictions lacked 3'- and 5'-UTRs, upstream and downstream of the genomic sequences were included in the predicted gene sequences to allow capturing the entire coding sequence. Furthermore, since the dataset generated in study II suggested a potential novel 5' end specifically for *ENSBTAT00000111847*, this gene prediction was modified accordingly.

4.4 Designing PCR primers (Study I and III)

After gene predictions for the genes of interest, their nucleotide sequences were extracted in the spliced forms. The template sequence for each gene was used as input for NCBI Primer-BLAST (Ye et al., 2012) to design PCR primers. In the primer design tool, *Bos taurus* was selected for primer pair specificity, while the repeat regions were excluded. Primers that have the lowest self-complementarity score were selected.

4.5 Preparation of cDNA library for molecular cloning (Study I and III)

For the preparation of cDNA libraries, pooled bovine oocytes or early embryos were used, with the number of oocytes or embryos ranging from three to five. Collected oocytes and embryos were derived from two separate batches. The developmental stages used for study I included the MII oocyte, 8-, and 16-cell

stage, whereas only the 8-cell stage embryos were used for study III. At the stage of sample collection, embryos or oocytes were placed in STRT lysis buffer, which was prepared and frozen beforehand. The lysis buffer composed of 0.525 μ M STRT-T30 V-LNA (TTAAGCAGTGGTATCAACGCAGAGTCGACT₂₉ + V), 1 U/ μ l RiboLock RNase Inhibitor (Thermo Fisher Scientific), 2.5 mM dNTP mix (Thermo Fisher Scientific), 0.1% Triton™ X-100 (Sigma-Aldrich), 10 mM Tris pH 8.0, and 5 mM Pierce™ DTT (Thermo Fisher Scientific).

The protocol for preparation of cDNA libraries followed the STRT-N (Boskovic et al., 2023) method until the fragmentation stage with some modifications. Briefly, samples that were placed in STRT buffer and frozen were immediately subjected to incubation at 80 °C for 2 min without allowing time for thawing at room temperature. This procedure was applied to ensure denaturation of RNAs found in the samples. Next, thawed samples were put on ice before starting the reaction for first-strand synthesis. The following components were used to mediate reverse transcription: 0.95 μ M template switching oligo (TSO; Biotin-ACGACGCTCTTCCGATCTN₈rGrGrG), 4.76 U/ μ l SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific), 1x SuperScript™ IV RT buffer (Thermo Fisher Scientific), 0.95 U/ μ l RiboLock RNase Inhibitor (Thermo Fisher Scientific), 4.8 mM DTT, and 952 mM Betaine solution. The program that was used for reverse transcription reaction included an initial incubation at 42 °C for 60 min. This was followed by ten cycles of tandem incubation at 50 °C for 2 min and 42 °C for 2 mins to allow reverse transcriptase to function at regions with secondary structure. The reverse transcription reaction was finalized by incubation at 85 °C for 5 min, resulting in the inactivation of the enzyme.

Reaction cleanup was performed using Dynabeads™ MyOne™ Carboxylic Acid (ThermoFisher Scientific) along with capture buffer (25% PEG-600, 2.25 M KCl). Next, PCR was performed using 1 μ M primer 1 (Biotin-ACACGACGCTCTTCCGATCT), 1 μ M universal primer (AAGCAGTGGTA-TCAACGCAGAGT), 0.197 mM dNTP mix (Thermo Fisher Scientific), 3% DMSO, 1x Phusion GC buffer (Thermo Fisher Scientific), and 0.04 U/ μ l Phusion DNA Polymerase (Thermo Fisher Scientific). The program that was used to amplify cDNA libraries started with denaturation at 98 °C for 10 s and was followed by two cycles of 55 °C for 30 s and 72 °C for 5 min. Next, another denaturation step at 98 °C for 10 s, which was followed by 23 cycles of 62 °C for 30 s and 72 °C for 5 min, was performed. Finally, the generated PCR products were incubated with exonuclease I (Sigma-Aldrich) and Shrimp Alkaline Phosphate (New England Biolabs), which enabled the removal of the primers and dNTPs that remained in the reaction tube.

4.6 PCR for molecular cloning (Study I and III)

As input for gene-specific PCR, 1 ul of cDNA library was used. For the amplification of *ARGFX*, *DUXA*, and *TPRX2*, 16-cell-stage-derived cDNA libraries were used; for *DUXC*, *LEUTX*, and *TPRX3*, 8-cell-stage-derived cDNA libraries were used; for *NOBOX*, MII oocyte-derived cDNA libraries were used. For the amplification of *TPRX1*, both 8- and 16-cell-derived cDNA libraries were used. Primer sequences used in both studies are shown in Table 1. PCR reactions for study I and III slightly differed.

Table 1. Primers sequences used to amplify PRD-like homeobox genes

Gene	Strand	Forward/Reverse Primer	Stage
<i>ARGFX</i>	+	AGGGTATTTACCATAAGGACACA/GTGCACTGTAATGGAGGGGT	16-cell
<i>DUXA</i>	-	CAAGACGGTCCTTCAGACAAGA/TGGATTGGTCACTAGGGCTG	16-cell
		TTATTGGTCTCTCTCTCCCCAG/TCCTAGGGCTGCATGGTTC	
<i>DUXC</i>	-	GCATTGGAGAGAGCATCGGG/TGAACGTGGCTAGGGGGAGAA	8-cell
<i>LEUTX</i>	+	GCAACACCTGGTCGAATCACT/GGGTTTTGCGACCATACGTG	8-cell
		TCAGGCAACACCTGGTCGAA/AGCCTGAGAATACGGAGTGG	
<i>NOBOX</i>	-	GTGCGGGAGTCCAGGC/CTGTGGAGCAAGGAGAGCAA	MI oocyte
<i>TPRX1</i>	-	CAGCAGGAGGATGCAAGAACC/TCAGCGTATCATCAGTGTGCCA	8-/16-cell
<i>TPRX2</i>	+	AGCATCAAACCTCCGACAGCA/ACAGGTAAGGATAGAGACAAGCA	16-cell
<i>TPRX3</i>	-	GCAGCAGTGAGGAACCAAGA/AGAGGCACTGAAATAGACAGGAA	8-cell

For study I, Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific) was used to perform the PCR reaction in line with the protocol provided by the manufacturer. The program used for PCR started with a denaturation step at 98 °C for 30 s. This was followed by 30 cycles of the following incubations: 98 °C for 10 s; 64 °C for 30 s; and 72 °C for 80 s. Final extension at 72 °C for 10 min was performed before the samples were cooled down to 4 °C.

For study III, Q5® High-Fidelity DNA Polymerase (New England Biolabs) was used to perform the PCR reaction in line with the protocol provided by the manufacturer. PCR reactions were either performed in the absence or presence of betaine with a final concentration of 1M. The program used for PCR started with a denaturation step at 98 °C for 3 min. This was followed by 35 cycles of the following incubations: 98 °C for 10 s; 66 °C or 68 °C, with or without betaine, respectively, for 20 s; and 72 °C for 105 s (Yaşar et al., 2026). Final extension at 72 °C for 5 min was performed before the samples were cooled down to 4 °C.

Amplified PCR fragments were excised from 2% agarose gel under UV light based on their expected lengths and extracted in accordance with the protocol of GeneJet Gel Extraction Kit (Thermo Fisher Scientific). Recovered blunt-end fragments were used as substrates for pCR4Blunt-TOPO vector using the Zero

Blunt TOPO PCR Cloning kit (Invitrogen). After the generation of plasmids containing full-length genes of interest, they were transformed into *E. coli* and incubated overnight at 37 °C. The next day, individual colonies were picked and used for inoculating LB medium containing ampicillin and incubated for 12–16 h at 37 °C. Plasmids were extracted using NucleoSpin Plasmid DNA Purification kit (Macherey Nagel). To evaluate the generation of correct plasmids by the size of the insert, plasmids were excised by EcoRI, which helped select the clones to be Sanger sequenced using T7 and T3 primers (Core Facility of Genomics, Institute of Genomics, University of Tartu). Molecular cloning experiments were performed in triplicate for *ARGFX*, *DUXA*, *DUXC*, *LEUTX*, *TPRX1*, *TPRX2*, and *TPRX3*, and in duplicates for *NOBOX*. UCSC Genome Browser Blat (Kent, 2002) was utilized to visualize the structures of cloned full-length genes.

4.7 Generation of the *TPRX3* overexpression plasmid (Study I)

Homeodomain-encoding *TPRX3* clone that was placed into pCR4Blunt-TOPO plasmid was used as input DNA material for amplifying this gene. The PCR amplification reaction contained primers bearing AgeI and NotI restriction sites as follows, respectively: 5'-AATATaccggtACCATGGGACCTGCAGGGAGGC AGC-3' and 5'-ATAAAgcggccgcCTATAGATTCAGACACTGCTGGGGG-3'. Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific) and the corresponding protocol provided by the manufacturer were used to perform PCR. The program used for PCR started with a denaturation step at 98 °C for 30 s. This was followed by 30 cycles of the following incubations: 98 °C for 10 s; 67 °C for 30 s; and 72 °C for 80 s. Final extension at 72 °C for 10 min was performed before the samples were cooled down to 4 °C. Amplified *TPRX3* fragment that contained AgeI and NotI restriction sites was cut and inserted into TetON piggyBac plasmid backbone that was also cut by the same restriction enzymes. To confirm the correct insert of the resulting pB-TetOn-bgi-*TPRX3*-IRES-GFP-PGK-Puro plasmid, Sanger sequencing was performed (Core Facility of Genomics, Institute of Genomics, University of Tartu).

4.8 *TPRX3* overexpression in bovine fibroblasts for RNA-seq (Study I)

To perform overexpression of *TPRX3* in bovine fetal fibroblast (BFF) cells, they were grown in cell culture until 70–80% confluency, after which cells were split using Trypsin-EDTA (0.25%) (Thermo Fisher Scientific). The Neon® Electroporation System (Thermo Fisher Scientific) was used to transfect the plasmids using 100 µl tips in accordance with the manufacturer's protocol that was optimized for adherent cells. The following plasmids with given amounts were used for each electroporation event with 1600 V, 20 ms, 1 pulse settings: 0.5 µg

of pPB-CAG-rtTA-M2-IN, 0.5 µg of pPB-CAG-hyPBase-IRES-Venus, and 1 µg of pB-TetOn-bgi-TPRX3-IRES-GFP-PGK-Puro. After 24 h, the expression of Venus was verified under the microscope using the GFP channel with emission at 486 nm. To select the cells with the plasmids, antibiotic selection was performed. Puromycin (0.2–0.4 µg/ml) and G418 (200–400 µg/ml) were used for approximately 3 days and 7 days after electroporation. Cells were cultured in two groups, one serving as the control group and the other representing the group in which *TPRX3* was overexpressed. For overexpression, 1 µg/ml doxycycline was added and incubated for 24 h. In the control group, no treatment was applied. Cells were collected 24 h post-treatment.

4.9 Reanalysis of whole-genome sequencing datasets (Study I)

NCBI SRA was used to download the following datasets: PRJNA183919, PRJNA184837, and ERX1815535. Orphan reads were excluded in each dataset, and mapping of pairs against the bosTau9 genome assembly was performed using Bowtie (Langmead et al., 2009) version 1.2.3 with the following parameters: “--chunkmbs 1024 -a -m 1 --best --strata -S -f --ff -X 2500”.

4.10 Preparation of cDNA library for RNA-seq (Study I, II and IV)

For Study I, BFF cells with and without doxycycline treatment were collected and used for RNA extraction using the Direct-zol™ RNA Microprep kit (Zymo Research). SMARTer® Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian (Takara) was used to prepare the cDNA libraries according to the manufacturer’s protocol. The concentrations of samples were assessed by Qubit Fluorometer (Thermo Fisher Scientific), and 47.9 ng of cDNA library from each sample was sent for sequencing. RNA-seq was performed on the Illumina NovaSeq 6000 platform.

For Study II and IV, the STRT-N protocol was performed to prepare cDNA libraries (Boskovic et al., 2023). Briefly, single bovine oocyte or preimplantation embryo was placed in a single well of a cell lysate plate containing lysis buffer. Immediately after the collection of single oocytes and embryos, the cell lysate plate was placed in a freezer and kept at –80 °C. Next, the plate was first transferred to 80 °C for 2 min incubation, and then to ice. To perform the reverse transcription reaction, an assembled reaction mix containing the ERCC Spike-ins mix was added. The following steps were used for the reverse transcription reaction: an initial incubation at 42 °C for 60 min; 10 cycles of tandem incubation at 50 °C for 2 min and 42 °C for 2 mins. Each sample was cleaned up, which was followed by the addition of barcoded primers. Next, cDNA amplification with 23 PCR

cycles was performed, after which the samples were pooled, cleaned up, and fragmented. The resulting fragments were extracted, end-repaired, and ligated with the sequencing adapters prior to amplification of the final library. Prepared libraries were sequenced at the Biomedicum Functional Genomics Unit (FuGU), University of Helsinki (Helsinki, Finland).

4.11 RNA-seq analysis (Study I, II, and IV)

For study I, as recommended in the protocol of SMARTer[®] Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian (Takara), using cutadapt (v.3.1), trimming of the first three nucleotides of Read 2 was performed (Martin, 2011). Using the same tool, adapter sequences found in both of the paired reads were also trimmed. For mapping the resulting paired-end RNA-seq reads against the bosTau9 genome assembly, HISAT2 (Kim et al., 2019) was used with the option --rna-strandness RF. StringTie (Pertea et al., 2015) was used to generate a count matrix using the option -e, and with the reverse strand (RF) specified. To produce a file compatible with DESeq2 analysis. The used gene annotation file was retrieved on 06.10.2023 using the genePredToGtf utility (Perez et al., 2025), which included pulling the gene annotation file directly from the UCSC public database and converting it to a GTF file. Differential expression analysis was performed in R using DESeq2 (Love et al., 2014). The design was formulated considering the doxycycline treatment status and the replicate. The control group was used as the reference. The default Benjamini and Hochberg method was used to compute the adjusted p-values. Median of ratios normalization method was used for quantifying the expression of *TPRX3*. Gene set enrichment analysis was performed for biological processes using the genome-wide bovine annotation (Carlson, 2019; T. Wu et al., 2021). Publicly available datasets generated by Lewin et al. (T. D. Lewin et al., 2022), Graf et al. (Graf et al., 2014), and Paulson et al. (Paulson et al., 2022) were utilized for comparison. To assess the statistical significance of upregulated genes shared among the compared datasets, calculation of p-values was performed using Fisher's Exact Test.

For study II, base call (BCL) files provided by the sequencing facility were used as input files along with the bovine genome assembly and gene annotation file to run STRTN-bTau_ens112.sh (https://github.com/baryasar/bTau_Embryo_TFE) using the options -a ens and -s 8M3S75T6B as described earlier (Boskovic et al., 2023). Genome assembly and gene annotation files were ARS-UCD1.3 and ARS-UCD1.3.112 (Ensembl version 112), respectively. The gene annotation file was slightly modified to contain the genomic information derived from the cDNA sequences of *ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, *TPRX1*, *TPRX2*, and *TPRX3*. The computations were performed at the High Performance Computing Center at the University of Tartu (University of Tartu, 2018). PCR duplicates are automatically removed by the pipeline using Picard Mark-Duplicates (Broad Institute, 2019) and the BARCODE_TAG = RX option. This is done by collapsing the reads that contain identical mapping positions

and UMIs. Genomic features and sequences with the NKLS prefixes were removed. TFE analysis was performed using the STRTN-TFE.sh from <https://github.com/gyazgeldi/STRTN>.

RNA spike-in feature of DESeq2 (v.1.44.0) was used for the normalization of TFE counts in R. The following samples were removed after quality control: 1, 11, 13, 23, 26, 37, 41. TFEs that have zero counts across samples were also removed. To identify TFEs with high variability, coefficients of variation were calculated, and a gamma generalized linear model (GLM) was used for fitting spike-in controls in R using the package MASS (v.7.3.61) (Venables & Ripley, 2002). Fluctuations of the control spike-in RNAs were used to model the technical variation in gene expression, and the significance of variation in gene expression was tested (Krjutškov et al., 2016). TFEs with the highest 20,000 CV² (coefficient of variation squared) and significant adjusted p-values (< 0.05) were selected to compute pairwise Pearson correlations. Variance stabilizing transformation (VST) values of the selected TFEs were used for computation and heatmap visualization using the pheatmap package (v.1.0.12). Percentages of TFEs corresponding to each genomic annotation were calculated for each stage. Similarly, relative poly(A)-tailed RNA content was calculated to demonstrate its global temporal change.

DESeq2 (v.1.44.0) was used for the differential gene expression, considering TFE counts as gene counts. The default Wald test of DESeq2 was used to compute differential expression across stages using filtered TFE counts. Each developmental stage was included in the design formula. For extracting the results, developmentally consecutive stages in pairs were set using the results() function. The results were filtered to include adjusted p-values lower than 0.05 and log2 fold changes greater than and less than zero for up- and downregulated TFEs, respectively. The R package clusterProfiler (v.4.12.6) was used for gene ontology enrichment analysis (S. Xu et al., 2024). The function enrichGO() was utilized with the following parameters: ENSEMBL as key type; org.Bt.eg.db as organism database; Biological Process (BP) as subontology group.

Differentially expressed TFEs detected in each consecutive developmental stage comparison were used for the analysis. Heatmap and Uniform Manifold Approximation and Projection (UMAP) visualizations were generated by the Seurat (v.5.1.0) package. For the latter eight principal components were used. Spike-in RNAs were used for normalization. Stage-specific TFEs were identified from normalized expression data. Scaled data from Seurat were used as input to the DoHeatmap() function to generate heatmaps with marker genes. Transcription factors annotated in AnimalTFDB 4.0 (W. K. Shen et al., 2023) were highlighted using the ComplexHeatmap package (v.2.20.0). Scaled data from Seurat were used as input to the kmeans() function to compute k-means clustering using the parameters centers = 6, nstart = 1000 and iter.max = 20. R packages ggpubfigs (v.0.0.1) (Steenwyk & Rokas, 2021) and viridis (v.0.6.5) (Garnier et al., 2024) with colorblind-friendly color palettes were utilized. Genomic intervals were visualized using the R package Gviz (v.1.48.0) (Hahne & Ivanek, 2016). TFEs identified in the stage-specific expression analysis were utilized to perform *de*

novo motif analysis. Specifically, the 16-cell-specific TFEs were selected. DeepTSS (Grigoriadis et al., 2022) was used to filter out the false-positive TFEs before the *de novo* motif analysis. Promoter sequences (–400 to + 100 bp) of the TFEs were extracted using bedtools getfasta (v.2.30.0) (Quinlan & Hall, 2010). *De novo* motif generation was performed by XSTREME (Grant & Bailey, 2021) from the MEME suite (Bailey et al., 2015).

For study IV, preprocessing of RNA-seq data was performed with the STRT-N pipeline detailed in <https://github.com/gyazgeldi/STRTN>. Input files included the bosTau9 genome assembly and RefSeq ARS-UCD.1.2 gene annotation file. Quality control and a fluctuation test were performed to retain samples and genes suitable for analysis. The R package Seurat (v.4.4.0) was used for generating PCA and UMAP plots. DESeq2 (v.1.40.2) was used for differential gene expression analysis with the normalization utilizing the spike-in RNAs. The R package ggplot2 (v.3.5.0) was used for visualizations. Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2025), and gene ontology analysis was performed.

4.12 Analysis of *DUXC* genomic organization across cattle breeds (Study III)

For study III, genome organization analysis was initially performed for Hereford (ARS-UCD2.0) and Holstein (NCBA_BosT1.0) (T. T. Li et al., 2023) genome assemblies. UCSC faToTwobit utility (Perez et al., 2025) was used to convert the FASTA file of each assembly to 2bit format. The resulting 2bit files were used to perform standalone *in silico* PCR tool (Perez et al., 2025) using the options -out=fa -maxSize=7000 along with the forward and reverse primers CCCGTGCCCCAGACCCC and CCGGGCCCGTGGAAGCCCACTGGCT, respectively. *In silico*-generated fragments were aligned for each breed using MAFFT (v.7.505) with the options --localpair --maxiterate 1000. IQ-TREE 2 (v.2.1.3) (Minh et al., 2020) was used to generate phylogenetic trees. Tree files were visualized using the R packages ape (v.5.8.1) (Paradis & Schliep, 2019) and ggtree (v.3.14.0) (G. Yu, 2020).

More comprehensive comparative analysis included the following cattle breeds: Jersey (Jang et al., 2023), Hanwoo (Jang et al., 2023), Holstein (T. T. Li et al., 2023), Hereford, Tibetan, Mongolian, Tuli, Wagyu (Pineda et al., 2025), and Yunling. The Holstein genome assembly was downloaded from the China National Center for Bioinformation, and genome assemblies for the rest were downloaded from the NCBI. All assemblies utilized long-read sequencing. Only chromosome 7 was used for the analysis, which was extracted using samtools (v.1.19) (H. Li et al., 2009). To extract the *DUXC* loci across breeds, a custom-defined internal *DUXC* repeat unit was mapped to chromosome 7 from each breed. For indexing and alignment minimap2 (H. Li, 2018) was used with the following settings: -x asm20 -c --eqx. The boundaries of the array were extended 6 kb upstream and downstream. The resulting *DUXC* locus from each breed was

aligned to each other using minimap2, generating 36 alignments. R package SVbyEye (v.0.99.0) (Porubsky et al., 2025) was used to visualize alignments across breeds. To demonstrate the composition of the *DUXC* array in each breed, alignment of the internal *DUXC* repeat unit against the *DUXC* locus from each breed was performed using minimap2. In addition, to incorporate *DUXC* gene annotations based on the full-length *DUXC* sequence, the minimap2 mode that considers splicing was utilized. The output SAM files were in order converted to BAM, GTF, and GFF3 using samtools (v.1.19) (H. Li et al., 2009), StringTie (v.2.2.1) (Pertea et al., 2015), and AGAT (Dainat, 2024), respectively. R package SVbyEye (v.0.99.0) (Porubsky et al., 2025) was used to visualize the resulting files.

4.13 Reanalysis of publicly available RNA-seq datasets (Study III)

To decide whether the internal or distal *DUXC* copy is expressed, the RNA-seq dataset from a published study (L. Zhu et al., 2022) was utilized. Genomic regions where internal and distal *DUXC* are located were extracted from chromosome 7 of the NCBA_BosT1.0 genome assembly. Alignment of RNA-seq reads to each extracted genomic region was performed using HISAT2 (v.2.2.1) (Kim et al., 2019). SAM-to-BAM conversion, sorting, and indexing were performed using samtools (v.1.6) (H. Li et al., 2009) with the flag -F 4 used for keeping only mapped reads. Resulting files of the same stage were merged using samtools merge, which is followed by running StringTie (v.2.1.7) (Pertea et al., 2015). Results were visualized by both Integrative Genomics Viewer (v.2.9.2) (Robinson et al., 2011) and Gviz (v.1.48.0) (Hahne & Ivanek, 2016).

Another RNA-seq dataset (J. Zhang et al., 2025) including all stages of preimplantation development was also utilized. Input genome assembly and gene annotation files were ARS-UCD1.3 and ARS-UCD1.3.112 (Ensembl version 112), with the latter slightly modified to contain the cDNA evidence-based structures of eight PRDL homeobox genes. Chromosome 7 was masked to contain only one copy of *DUXC* unmasked, using bedtools maskfasta (v.2.30.0) (Quinlan & Hall, 2010). HISAT2 (v.2.2.1) (Kim et al., 2019) was used for mapping. For quantifying gene expression, featureCounts (v.2.0.1) was used. Differential gene expression was performed in R using DeSeq2 (v.1.44.0) with each stage considered a separate condition in the design formula. Variance-stabilizing transformation (VST) was applied, and z-scores were calculated to generate a heatmap using pheatmap (v.1.0.13), RColorBrewer (v.1.1.3) and viridis (v.0.6.5) (Garnier et al., 2024). After masking all *DUXC* copies except for one, reanalysis of the dataset generated in study II was also performed using the same approach as described for study II. Computations for reanalysis were performed at the High Performance Computing Center at the University of Tartu (University of Tartu, 2018).

5 RESULTS AND DISCUSSION

5.1 Study I: Molecular cloning of bovine IVF embryo-derived PRD-like homeobox genes

5.1.1 Experimental validation of the expression of seven PRD-like gene during early bovine development

We carried out cDNA cloning of several PRDL homeobox genes to reveal their full-length sequence and confirm their expression in bovine oocytes and/or early embryos. This experimental approach was primed by the reanalysis of an RNA-seq dataset (Graf et al., 2014), enabling gene prediction and primer design (Figure 12).

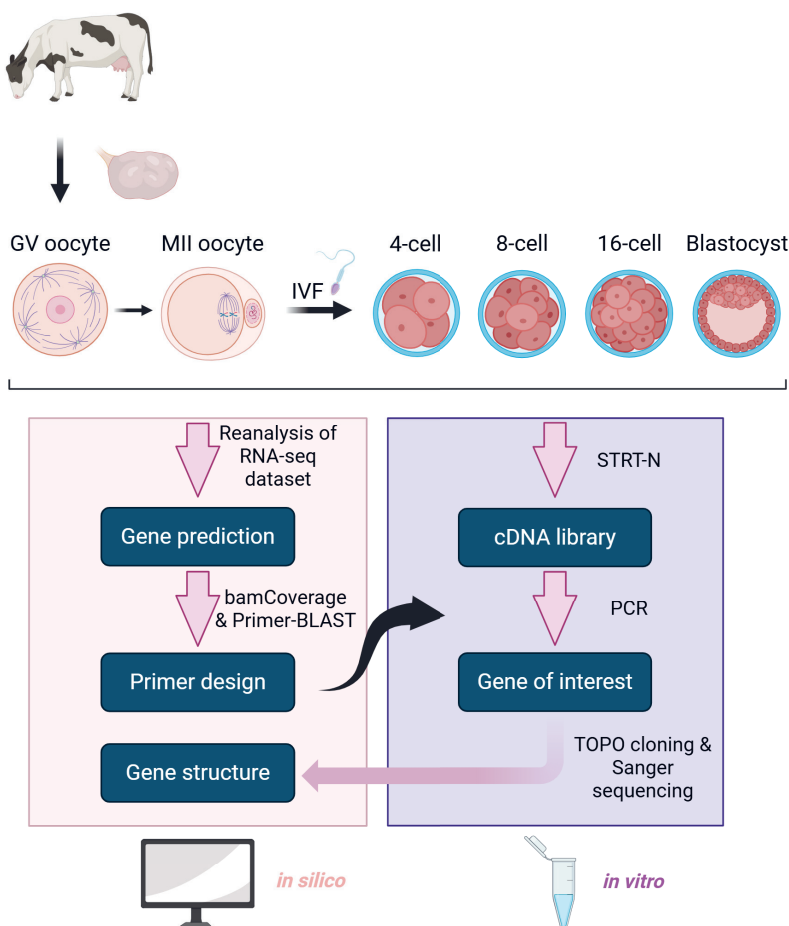


Figure 12. Overview of the molecular cloning experiments. This figure was originally published in Yasar et al., BMC Genomics (2024) under CC BY 4.0 and created in <https://BioRender.com>.

Computational analysis was combined with the experimental work, in which preparation of cDNA libraries using bovine early developmental stages, gene-specific cDNA amplification, and eventually TOPO cloning and Sanger sequencing led to the identification of full-length gene sequences (Figure 12). Among the genes investigated, *TPRX2* was not annotated, whereas *ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, and *TPRX1* had inaccurate predictions in the NCBI RefSeq genes as well as the Ensembl Gene Predictions. Notably, our gene prediction approach did not result in any prediction for *DPRX*, which was suggested to be a pseudogene in bovine in another study (Maeso et al., 2016).

Our analysis allowed us to determine the expected lengths of the homeo-domain-encoding cDNA fragments, revealing 999 nucleotides (nt), 774 nt, 1809 nt, 1949 nt, 1264 nt, and 1207nt for *ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, *TPRX1*, and *TPRX2*, respectively. Guided by the gene expression levels of the given PRDL homeobox genes, we selected the following combinations of genes and developmental stages: *NOBOX* from MII oocyte; *TPRX2* and *LEUTX* from 8-cell; *ARGFX* and *DUXA* from 16-cell embryos. For *TPRX1*, both 8- and 16-cell stages were used as the source. The PCR result demonstrated the first line of experimental evidence for the full-length expression of these genes in bovine early developmental stages, as we observed prominent bands at the expected lengths in the agarose gel. In addition to the genes mentioned above, we also tested the expression of *TPRX3* in 8-cell stage embryos, which was pointed out as a *TPRX* duplicate (T. D. Lewin et al., 2021) through evolutionary analysis but not confirmed experimentally. Our cloning experiments showed that homeodomain-encoding *TPRX3* mRNA was present in 8-cell bovine IVF embryos.

Sanger sequencing and alignment of these cDNA clones revealed their gene structures and identified several transcript isoforms. Our cDNA clones for *DUXA*, *LEUTX*, *TPRX1*, *TPRX2*, and *TPRX3* mapped to chromosome 18, whereas those for *ARGFX* and *NOBOX* mapped to chromosome 1 and 4, respectively. A total of 20 transcript variants of the PRDL genes were isolated. Each gene contained at least one variant with a homeobox, but isoforms without a homeobox were also identified. The highest number of transcript variants was observed for *TPRX1* and *TPRX2*, the former having six distinct variants.

With respect to RefSeq annotations, our results showed experimental confirmations of the expression of novel exons, previously reported only as predictions (T. D. Lewin et al., 2021). These include two exons in the 5' end of *ARGFX* and one exon in the 5' end of *DUXA* and *LEUTX*. Although experimental evidence was presented for *NOBOX* (HQ589330) earlier (Tripurani et al., 2011), our results revealed another exon upstream of the previously reported exon 1. NCBI gene annotations included a prediction for *TPRX1* but not for *TPRX2* or *TPRX3*. Although these three bovine genes had been predicted independently in a separate study (T. D. Lewin et al., 2021), our results showed a more comprehensive picture, revealing several distinct isoforms in addition to the previously predicted homeodomain-only variants based on evolutionary analysis. Furthermore, we also demonstrated the syntenic relationship between bovine and human for *DUXA* as anchor genes flanking *DUXA* such as *USP29*, *ZIM3*, *AURKC*, and *ZNF543*

were observed. In summary, our results provide embryo-derived, reliable gene structures, building upon previous structures based solely on computational predictions.

5.1.2 Identification of SNVs in PRD-like homeobox transcription factor genes

We identified single-nucleotide variants (SNVs) in the seven PRDL homeobox genes investigated as a result of aligning Holstein cattle-derived full-length cDNA sequences to the reference genome with Hereford origin. These SNVs were found across the gene body. To confirm that observed SNVs were indeed due to breed-related differences, we reanalyzed whole-genome sequencing (WGS) datasets and observed that these SNVs were also present in other WGS datasets of Holstein origin (PRJNA183919 and PRJNA184837). Out of the 58 unique SNVs or indels identified in the cDNA clones of seven genes, 37 were found in the reanalyzed datasets, whereas 21 SNVs were specific to our cDNA clones. These SNVs were not only in the 5'-UTR or 3'-UTR but also within the CDS, raising the question of whether they alter residues in the homeodomain, a functional domain important for DNA binding by PRDL TFs. Our investigation showed that almost none of the SNVs were found in regions coding for the homeodomain. One exception was one *DUXA* cDNA clone. Since we cannot speculate on the developmental potential of the embryo or embryos from which this *DUXA* cDNA clone was derived, we can conclude that the observed SNVs did not affect functional domains in the proteins, highlighting the potentially important roles of PRDL proteins during preimplantation bovine embryo development.

5.1.3 Functional domains found in PRD-like proteins

Based on the cDNA sequences of PRDL genes generated in this study, for each gene, we selected an open reading frame that codes for a protein with a homeodomain or homeodomains, leading to the derivation of putative protein sequences of *ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, *TPRX1*, *TPRX2*, and *TPRX3*. We performed multiple sequence alignment using these sequences and confirmed the presence of certain residues conserved in the homeodomain: tryptophan (W), asparagine (N), and arginine (R) at position 48, 51, and 53, respectively. Position 50 is generally used to categorize PRDL proteins, and we showed that *DUXA* and *NOBOX* were Q-type, and *LEUTX*, *TPRX1*, *TPRX2*, and *TPRX3* were K-type PRDL TF. Although such a categorization was not made for PRDL proteins with an arginine at position 50, we referred to *ARGFX* as R-type. Arginine at this position in the homeodomain is uncommon, with only a few exceptions reported in ZFH class proteins, namely *ATBF1* (Morinaga et al., 1991) and *ZFHX4* (Sakata et al., 2000). We reported homeodomain to be on the N-terminus of *LEUTX*, *TPRX1*, *TPRX2*, and *TPRX3*. The position of the homeodomain was central for *ARGFX* and *NOBOX*. As a protein with two homeodomains, *DUXA* had these domains at

each end of the protein. Our results demonstrating the conservation of the critical residues of homeodomains suggest their functional effect during early embryo development.

Next, we investigated whether the previously described nine amino acid transcription domain (9aaTAD), which is known to be implicated in transcription activation (Piskacek et al., 2007), was present in bovine PRDL TFs. We found 9aaTAD in three of the seven PRDL TFs using their protein sequences derived from the cDNA clones. Bovine LEUTX contained 9aaTAD at the C-terminus, similar to human LEUTX. A broader domain was described for LEUTX, referred to as the Leutx domain (Katayama et al., 2018). These domains at the C-terminus are thought to be the transcription activation arm of the protein, while the homeodomain at the N-terminus ensures DNA contact. A conformational experiment provided evidence for this by demonstrating the loss of interaction between LEUTX and EP300 and CBP in the absence of its 9aaTAD (Gawriyski et al., 2023). Bovine NOBOX had a central 9aaTAD overlapping with the homeodomain. Interestingly, two separate 9aaTAD was predicted for TPRX3, making it an interesting candidate as a factor involved in bovine EGA.

5.1.4 Ectopic expression of *TPRX3* confirms its transcription activator role

Given that bovine TPRX3 contained a 9aaTAD and was not tested before for its transcription activation capacity, we set out to overexpress *TPRX3* in bovine fetal fibroblast (BFF) cells to examine its effect on the transcriptome. We utilized a doxycycline-inducible system, in which the expression of *TPRX3* was stimulated only upon addition of doxycycline to the cell culture medium. Since our molecular cloning experiment yielded two distinct *TPRX3* isoforms, we opted for the longer *TPRX3* isoform, which encodes a homeodomain, can potentially bind DNA, participates in transcription activation. We verified the expression of *TPRX3* in induced BFF cells by the GFP signal, as our construct contained the GFP coding sequence downstream of *TPRX3*, separated by the IRES sequence.

We performed the RNA-seq analysis using two replicates that qualified for downstream analysis. The total number of raw and processed reads, as well as the alignment rates, were summarized in Table 2. The minimum and maximum alignment rates were approximately 80.5% and 90%, respectively. The doxycycline treatment was reflected in the Principal Component Analysis (PCA), which showed separation based on the *TPRX3* overexpression along the second principal component (PC2). Differential gene expression analysis revealed 87 and six significantly up- and downregulated genes, respectively. Visualization by volcano plot showed the prominent expression of *TPRX3* and its transcriptional effect. Because our method of choice for RNA-seq did not only enrich poly(A)-tailed RNA, we detected coding and noncoding RNAs induced by *TPRX3* overexpression. These included splicing machinery-related small nuclear RNA (snRNA), RNA modification-related small nucleolar RNAs (snoRNAs), and ribosomal

RNAs (rRNAs), which are fundamental structural and functional components of ribosomes. Gene ontology (GO) analysis suggested the biological processes these biomolecules are involved in.

Table 2. Read numbers obtained from RNA sequencing of the *TPRX3*-overexpressed cells.

Cell line	Number of raw reads	Read numbres after trimming and filtering	Alignment rate	Dox	Final cDNA amounts pooled (ng)
Replicate 1	52881806	51673529	0.8662	-	47.9
	49143393	47409233	0.8692	+	47.9
Replicate 2	45449211	44500066	0.8976	-	47.9
	62036456	59887867	0.8041	+	47.9

To support our results on the transcriptional effect of *TPRX3*, we compared our dataset with those derived from early bovine embryos (Graf et al., 2014; Paulson et al., 2022), to find commonly regulated genes. Common genes upregulated during the 16-cell-to-blastocyst transition and in *TPRX3*-overexpressed cells included mitochondrially encoded genes such as ND1-ND5, CYTB, COX1-COX3, ATP6, which are components of complexes I, III, IV, and V, respectively. It suggests that *TPRX3* potentially activates genes involved in the shift from low oxygen, pyruvate-based metabolism to high oxygen, mitochondrial-based respiration. Several noncoding RNAs were also identified as commonly regulated genes with statistical significance. Similarly, we observed that 12 genes upregulated by *TPRX3*-overexpressing cells were also upregulated by *LEUTX* overexpression in an independent dataset (T. D. Lewin et al., 2022).

5.2 Study II: Identification of embryonic genome activation regulators through transcriptome profiling of bovine embryo developmental stages

5.2.1 5'-end mRNA profiling of bovine oocytes and preimplantation embryos

To gain insights into gene regulation during bovine early development, we performed RNA sequencing on several bovine developmental stages using a 5'-end-focused method that captures the 5' ends of mRNAs, thereby identifying transcription start sites and guiding the annotation of putative promoter regions, while minimizing reads originating from the 3' ends of mRNAs that might be undergoing degradation. We performed STRT-N RNA-seq (Boskovic et al., 2023) after sampling single oocytes or single embryos from the following stages: GV and MII stage oocytes and 2-cell, 4-cell, 8-cell, 16-cell, and blastocyst stage

embryos. In total, we sequenced cDNA libraries of 48 samples, two of which were negative controls without any RNA. Out of the remaining 46 samples, 41 qualified for the downstream analyses. The average number of reads of samples selected for downstream analyses was 3,364,498.

In this study, we considered the first exons as the minimal genomic unit for gene expression quantification, as opposed to using entire genes, which are conventionally used as the unit of quantification. This quantification unit was previously defined as the transcript far 5'-end (TFE), named as such because it represents the ultimate end or transcription start site of a gene (Töhönen et al., 2015). Therefore, our approach is unique in that it provides a TFE-based transcriptome analysis of bovine early development, which is enabled by the 5'-end RNA-seq method STRT-N (Boskovic et al., 2023). Our analysis revealed a total number of 122,688 TFEs. It is noteworthy to mention that several TFEs can correspond to a single gene, suggesting distinct TSSs. We detected 88,357 TFEs corresponding to 18,501 annotated bovine genes. We also detected 34,331 TFEs that were found in genomic regions with no known annotation.

Pairwise correlation using the top 20,000 fluctuating TFEs identified two clusters in the heatmap, each of which contained stages either before or after bovine EGA. Almost half of these TFEs were annotated, corresponding to 6,695 known genes. The heatmap showed the formation of distinct groups by MII oocyte, 16-cell, and blastocyst stages. Cleavage stages of pre-EGA stages, namely 2-, 4-, and 8-cell embryos demonstrated similarity. These results were also reflected in the UMAP plot, in which the blastocyst stage stood apart from the other clusters, whereas the oocyte and cleavage stages formed their own cluster, each separated from one another. Within the cleavage stage cluster, 2-, 4-, and 8-cell embryos were more closely grouped, with 16-cell embryos positioned slightly apart, reflecting the onset of bovine major EGA. These results highlight the changes in the transcriptome in early embryos as a result of MZT and bovine EGA.

5.2.2 Changes of poly(A)-tailed RNAs and the percentages of aligned reads assigned to each genomic feature

We also demonstrated MZT by calculating the relative changes of the poly(A)-tailed RNA. Since the STRT-N protocol includes exogenously added defined amounts of spike-in RNAs that serve as constant reference entities for normalization, we reliably demonstrated that oocyte stages had similar levels of poly(A)-tailed RNA, with a significant drop observed at the 2-cell stage, which is an initial and significant sign of MZT (Kojima et al., 2024). The decrease in poly(A)-tailed RNA continued gradually until the blastocyst stage, indicating ongoing maternal mRNA clearance. We observed an increase in poly(A)-tailed RNA at the blastocyst stage.

Next, we calculated percentages of aligned reads corresponding to the following genomic features: upstream & 5'-UTR of coding genes, CDS & 3'-UTR of coding genes, upstream & first exon of noncoding genes, other than first exon

of noncoding genes, intron, and unannotated. Aligned reads are assigned to individual TFEs. We therefore present the overall percentages of aligned reads falling under TFEs, classified by their corresponding genomic features. The overall percentages were summarized in Table 3. The proportion of aligned reads assigned to unannotated TFEs was at least 16% across stages, indicating a high percentage of bovine genomic features that have not been characterized yet. Similarly, aligned reads assigned to intronic TFEs accounted for an average of 8.4% across all stages, suggesting the presence of distinct transcription start sites during early development. Most aligned reads were found in the upstream of 5'-UTR of the coding genes in all stages, consistent with the 5'-based STRT-N method. Another thing to note is the increased share of aligned reads assigned to CDS and 3'-UTR in the blastocyst stage. This might indicate distinct transcription initiation events, although it is possible that some of these reads were derived from degraded transcripts.

Table 3. Stage-averaged percentages of aligned reads across TFE annotations.

Annotation	GV oocyte	MII oocyte	2-cell	4-cell	8-cell	16-cell	Blastocyst
CDS and 3'-UTR of coding	14.23	13.40	14.26	13.74	14.02	12.61	28.15
Intron	8.49	7.70	8.65	8.78	9.89	9.67	5.71
Noncoding_other-exon	0.15	0.11	0.11	0.12	0.12	0.12	0.19
Unannotated	17.02	24.15	16.49	16.39	15.96	18.77	19.60
Upstream and 1st exon of noncoding	2.05	2.19	3.03	1.90	1.78	1.78	2.76
Upstream and 5'-UTR of coding	58.08	52.45	57.46	59.07	58.23	57.05	43.60

5.2.3 Differential expression of TFEs in consecutive bovine developmental stages

We performed differential expression analysis by considering TFEs as genes and compared the expression of TFEs at each stage to that of the preceding developmental stage. Detected significantly up- and downregulated genes and TFEs are summarized in Table 4.

Our results showed the highest downregulation following fertilization during the MII-oocyte-to-2-cell transition. This was followed by the gradual downregulation of TFEs during cleavage stages. Another prominent downregulation was observed during the 8-to-16-cell transition, which is in line with the previous report suggesting that maternal mRNA clearance is required for EGA (Bogliotti et al., 2020). Progression from the 8-cell to 16-cell stage was also accompanied by upregulation of the highest number of TFEs after fertilization, marking the major bovine EGA.

Table 4. Number of up- and downregulated genes and TFEs detected at given developmental transitions.

	Upregulated		Downregulated	
	TFE	Gene	TFE	Gene
GV-to-MII	8902	2733	6508	3618
MII-to-2-cell	211	149	21918	7400
2-cell-to-4-cell	29	21	122	87
4-cell-to-8-cell	4	1	92	70
8-cell-to-16-cell	1621	731	20461	8536
16-cell-to-blastocyst	12989	6501	2057	1301

We also examined the annotations of the up- and downregulated TFEs. The majority of the upregulated TFE identified during oocyte maturation were unannotated, suggesting that characterization and improved annotation of genomic features are needed, especially using materials derived from oocytes. Gene ontology (GO) analysis for upregulated TFEs pointed to biological processes such as cell cycle, which is expected for the completion of meiosis (R. Li & Albertini, 2013). GO terms for downregulated TFEs included ribosome and translation-related processes.

Upregulated TFEs during the MII-oocyte-to-2-cell transition were mostly coding genes. The downregulated TFEs during the same developmental transition included a high percentage of unannotated TFEs. These genomic features are to be elucidated. GO analysis revealed chromatin remodeling and nucleosome organization for upregulated TFEs. GO terms for downregulated TFEs included DNA repair and DNA damage.

Almost half of the upregulated TFEs were coding, whereas downregulated TFEs were much higher during the transition from the 8-cell to 16-cell stage. Furthermore, unannotated TFEs made up close to 40% of the upregulated TFEs during bovine EGA, highlighting the need for further research for the characterization of these genomic features. GO terms for upregulated TFEs pointed to events like RNA transcription, processing, and splicing. GO terms for downregulated TFEs included processes such as protein transport and post-translational protein modification.

From the 16-cell to blastocyst stage, upregulated TFEs were the highest we detected. GO terms for these TFEs included peptide metabolic process, ribo-protein complex biogenesis, amide biosynthetic process, and cellular response to stress. Cellular stress was also reported for prolonged culturing of embryos (Puscheck et al., 2015) and those that were cultured under low oxygen (Boskovic et al., 2025).

5.2.4 Temporal expression of PRD-like homeobox genes

We optimized the STRT-N pipeline to be compatible with Ensembl gene annotation files, into which we integrated cDNA cloning-defined gene definitions reported in our earlier study (Yaşar et al., 2024), to examine TFE-based dynamic expression of these PRDL homeobox genes.

We detected a single TFE for *TPRX3*, and it was located slightly upstream of our cDNA-based *TPRX3* structure, indicating a new site for transcription initiation. Expression levels observed in our current dataset and in another dataset (Graf et al., 2014) that we also used to report *TPRX3* expression differed. The former showed the highest expression at the 16-cell stage, whereas in the latter it occurred at the 8-cell stage. Similarly, we showed that *LEUTX* was expressed at its peak at the 16-cell stage, whereas its highest expression was previously reported to be at the 8-cell stage (Maeso et al., 2016). These results reflect the effect of the method of choice on the expression level quantification. We detected four distinct TFEs belonging to *LEUTX*, three of which were located at the 5' end of our cDNA-confirmed *LEUTX* structure, but no TFEs were found around the first exon of the unique Ensembl *LEUTX* transcript model ENSBTAT00000065536 belonging to the gene model ENSBTAG00000046425.

Since TFE analysis has the potential to capture possible novel sites for transcription initiation, we searched for TFE peaks that could indicate previously unannotated sites. We detected increased TFE signals across all developmental stages between the first and the second exon of *DPPA3* (ENSBTAG00000046609). Further investigation of this site could help characterize this gene further, which was shown to be required for normal development of bovine embryos (Bakhtari & Ross, 2014).

ARGFX showed two TFEs, one of which showed very low to zero expression, whereas the other one showed steady levels of low expression until the 16-cell stage. At this stage, elevated expression was observed with a sharp decrease at the blastocyst stage.

DUXA had three distinct TFEs, one of which was at the 5' end and two of which were located downstream. Like the other PRDL homeobox genes, *DUXA* also elevated its expression following bovine EGA. Notably, in this study, log₂ normalized value -10 was defined as the cutoff to assess whether a gene is expressed or not. One of the TFEs located at the 3' end of the gene (TFE45905) showed almost zero expression in all embryonic stages except for the 16-cell stage, which could have a potential regulatory role.

One of the two TFEs detected for *TPRX1* was located around the promoter of our previously reported *TPRX1* isoforms, rather than that of the Ensembl-predicted gene ENSBTAG00000057182. The other TFE for *TPRX1* was located close to the 3' end of gene models in this locus. Expression of TFEs identified for both *TPRX1* and *TPRX2* was observed most prominently at the 16-cell stage. The TFE identified at the 5' end of our *TPRX2* gene structure obtained through molecular cloning confirmed that TFE analysis is effective for capturing the ultimate 5' ends of genes, thereby providing an opportunity to define their promoters and investigate potential regulatory mechanisms.

As opposed to the abovementioned PRDL homeobox genes, *NOBOX* is another PRDL homeobox gene that is primarily expressed in oocytes. This was reflected in our TFE approach, which also highlighted several sites with identified TFEs.

To investigate the dynamic changes of PRDL homeobox genes further, we included several more genes that were categorized under this gene class (Bürglin, 2011). Previously, PRDL homeobox genes were classified as zygotic, maternal, and mixed based on the stages during which they were predominantly expressed during human early development (Töhönen et al., 2015). Similarly, to categorize these genes in the same manner, we plotted a heatmap and defined maternal and zygotic PRDL homeobox genes expressed during bovine early development. We, therefore, defined *ALX1*, *ALX3*, *ARX*, *HESX1*, *HOPX*, *MIXL1*, *NOBOX*, *PHOX2A*, *PITX3*, *SHOX*, *SHOX2*, *UNCX*, *VSX1*, and *VSX2* as maternal PRDL homeobox genes. The zygotic genes included *ARGFX*, *DUXA*, *LEUTX*, *OTX2*, *PHOX2B*, *TPRX1*, *TPRX2*, and *TPRX3*.

5.2.5 Identification of embryonic stage-specific gene expression

To identify TFEs exclusive to each developmental stage, we selected TFEs previously identified as upregulated between consecutive developmental stages and then applied a filter requiring that each gene be expressed exclusively at the latter stage and not any other stage analyzed. Since the number of significantly upregulated genes was very few in the 2-, 4-, and 8-cell stage, and because their transcriptional profiles resembled each other based on the UMAP plot, we excluded these stages from our search for stage-specific TFE expression.

Our analysis revealed 588 and 730 stage-specific TFEs for GV and MII oocytes, respectively. Out of the 588 TFEs, 229 corresponded to known genes, whereas MII oocyte-specific TFEs belonged to 155 annotated genes. The number of TFEs that are activated only during EGA was 197, with 75 of them corresponding to known genes. For the blastocyst stage, this number was 2,379, and 1400 TFEs belonged to the genes annotated in the bovine genome.

To highlight the TFEs of TF genes that are expressed in a stage-specific manner, we utilized the bovine TF database (W. K. Shen et al., 2023) and marked these TF genes in the heatmap. GV oocyte-specific TFEs belonged to *RFXANK*, *ZNF414*, *ZNF354A*, *ARX*, *FIGLA*, *ETV5*, *CEBPB*, *XBPI*, *TCF25*, *HESX1*, *RARB*, *IRF7*, *ZNF691* and *GBX2*. Some of the MII oocyte-specific TFEs were of the following TF genes: *PROX2*, *BCL2*, *TSHZ2*, *BSX*, *ZNF304*, *SP6*, *NR1D1*, *ZNF750*, *BAZ2B*, *ZEB2*, and *TCF4*. TFEs specific to the 16-cell stage included two of the previously discussed PRDL homeobox genes, *DUXA* and *LEUTX*. Furthermore, there were also *DPPA2*, *TCF3*, *TFAP2C*, *ID1*, *ZNF217*, *MYC*, *IRX5*, *ENSBTAG00000049688*, *HOXD11*, *EOMES*, *ZNF12*, *AHR*. Some of the blastocyst-specific TFEs were of *TEAD4*, *NFX1*, *MTA3*, *TBX3*, *ETS2*, *EBTF*, and *OTX2*. It is noteworthy to mention that abovementioned genes have at least one TFE that is expressed exclusively at a single stage, but it does not necessarily indicate that all transcripts or TFEs of those genes are expressed in a stage-specific fashion. One gene could have more than one TFE.

5.2.6 Putative regulators of bovine embryonic genome activation

To study bovine EGA, we aimed to focus on TFEs that are activated as a result of bovine EGA at the 16-cell stage. Our list of stage-specific TFEs served as a useful resource for this purpose. Identifying the factors that activated these 16-cell stage-specific TFEs gives insight into the regulation of bovine EGA. For this purpose, we extracted the promoter regions of these TFEs and performed *de novo* motif analysis to identify putative factors with binding motifs similar to the discovered motif, under the assumption that these factors may bind these promoters and regulate bovine EGA.

Among the candidate proteins whose binding motifs were similar to the *de novo* motifs, we selected proteins encoded in the bovine genome. DUXA was one of the proteins showing a close match to the motif we discovered (TAAYCYAATCA). A putative binder of the same motif was identified as PHOX2A, a maternal PRDL TF. Other maternal PRDL TFs included ALX1 and UNCX. Another *de novo* motif, TCTCYTCCTSCYYCT, showed a close match to the binding motif of GABPA. The third motif we identified, GCCACCAGGGAAGCC, had EBF1 as the putative binder. CTGGARAAT CCCATG matched the binding to another motif of RELA, a subunit of NF- κ B. KLF factors, such as KLF4 and KLF10, showed binding to another *de novo* motif, GVGGGGCGGGGCBGG.

Although OTX2 showed lower levels of expression in earlier stages, the last *de novo* motif discovered, CTTTAATCCAATTTT, showed a close match with the binding motif of the PRDL class protein OTX2. While we focused on putative binders whose genes were expressed at higher levels in pre-EGA stages, it is possible that proteins with low expression, or those predominantly expressed in oocytes, might also contribute to bovine EGA if they remain stable until the time of EGA. One study in mice showed that proteins detected in oocytes were stably observed until the blastocyst stage (H. Zhang et al., 2023).

5.3 Study III: Characterization of bovine *DUXC*

5.3.1 *DUXC* is a putative binder of the promoters of the genes activated at the 16-cell stage

We identified a *de novo* motif enriched in the promoters of a subset of upregulated genes at the 16-cell bovine embryos in our previous work (Yaşar et al., 2025). A putative binder of this motif, TAAYCYAATCA, is DUX4, a PRDL class protein that is not encoded in the bovine genome. However, a homologue of this protein has been suggested to be *DUXC*, which also contains a double homeo-domain similar to DUX4 and shares a conserved C-terminal domain with DUX4 (Clapp et al., 2007; Hewitt et al., 1994; Leidenroth et al., 2012). Therefore, it is likely that TAAYCYAATCA could show a match to the binding motif of *DUXC*. However, *DUXC* annotation in the bovine genome is only based on compu-

tational predictions. For example, Ensembl gene annotations for bovine contain predictions for *DUXC* without an official gene symbol, limiting its interpretability in downstream bovine genome-wide analyses (Dyer et al., 2025). For this reason, we aimed to derive the full-length *DUXC* transcript from early bovine embryos by performing molecular cDNA cloning and to report it to enable accurate gene annotation.

5.3.2 Molecular cDNA cloning of *DUXC* from the 8-cell stage bovine embryos

In order to perform molecular cloning of *DUXC*, we set out to design PCR primers, which required having template sequences of *DUXC* gene structures. Therefore, we initially sought to find available *DUXC* gene predictions. To find the *DUXC* locus, we utilized the pig *DUXC* sequence reported earlier and performed a tblastn search (Nip et al., 2023). Since *DUXC* is a multi-copy gene, at this locus we observed several Ensembl transcript predictions located in tandem. Many of these tandem transcripts contain the same sequence, as this locus is an array of repeated sequences. For this reason, we selected transcript predictions with distinct sequences to design primers for molecular cloning: ENSBTAT00000081668, ENSBTAT00000070667, ENSBTAT00000080304, ENSBTAT00000111847. These transcripts were from Ensembl bovine gene annotation versions 104 and 112.

Since our previously generated dataset contained potential transcription start site information, we reanalyzed this dataset and discovered TFEs that are found upstream of the 5' end of the predicted transcript ENSBTAT00000111847. This prompted us to design primers for this potential novel 5' end of *DUXC*. For the rest of the transcript models, we used the given structures as a model to design PCR primers. The expected fragment lengths for ENSBTAT00000081668, ENSBTAT00000070667, and ENSBTAT00000080304 were 733 nucleotides (nt), 813 nt, and 1382 nt, respectively. For ENSBTAT00000111847, we designed 4 distinct primer pairs with the expected fragment lengths of 2620 nt, 2067 nt, 1670 nt, and 1622.

We performed our initial experiments using Phusion polymerase and failed to amplify *DUXC* cDNA. The addition of enhancer reagents in the PCR reactions did not improve the results. Since the average GC content of our templates is 66.7%, we considered using a different polymerase that can work better with such GC content. A previous study, which showed successful amplification of a DNA fragment of 63% GC content using Q5 but not Phusion polymerase, prompted us to opt for Q5 polymerase (Assal & Lin, 2021). Although Q5 polymerase alone also failed, the addition of betaine resulted in the amplification of the gene model that was based on the TFE we discovered. Our molecular cloning revealed a novel first exon for bovine *DUXC*. In addition, the exon structures were slightly different, with exon 3 and 4 being longer as compared to the computationally predicted *DUXC* models. We also confirmed the presence of a double homeobox.

Furthermore, 9aaTAD was encoded in the 3' end of the *DUXC* cDNA sequence we identified. In this study, we report the first early embryo-derived full-length sequence of *DUXC*, while full-length sequences of the human counterpart of *DUXC*, namely, *DUX4*, have previously been reported from non-embryonic sources.

5.3.3 Clearance of *DUXC* mRNA depends on minor bovine EGA

With our updated *DUXC* annotation, we sought to reanalyze RNA-seq datasets (Yaşar et al., 2025; J. Zhang et al., 2025) derived from several bovine embryo development stages to observe the temporal expression of *DUXC*, which usually may be ignored because of its multi-copy nature. Some mappers may select one mapped read if they align to multiple positions (Kim et al., 2019), and some quantification tools may completely discard the signal that comes from multi-mapped reads (Deschamps-Francoeur et al., 2020). For this reason, we performed our reanalysis by considering these concerns and quantified *DUXC* expression across different stages and conditions.

The first dataset we utilized not only contains normal IVF embryo developmental stages but also those that were treated with α -amanitin or 5,6-dichlorobenzimidazole 1- β -D-ribofuranoside (DRB). The former directly binds to Pol II, causing its arrest and degradation, thereby inhibiting transcription (Brueckner & Cramer, 2008), while the latter indirectly inhibits transcription elongation (Chodosh et al., 1989). Because DRB-mediated transcriptional inhibition is reversible (Efrat & Kaempfer, 1984), embryos were incubated with DRB until the 2-cell stage to block minor EGA and subsequently washed to allow transcription. Embryos subjected to this condition were referred to as DRB wash (DRBW). A more direct approach to transcription inhibition was achieved using α -amanitin, producing 2-, 4-, and 8-cell embryos in which embryonic transcription was inhibited. These embryos were also referred to as transcriptionally-blocked embryos (TBE).

Our reanalysis of this first dataset showed that *DUXC* transcript levels increased immediately after fertilization. This increase compared to oocyte stages was also true for the 2-, 4-, 6-to-8-, and 8-cell stages. Notably, we performed the reanalysis by incorporating the cDNA-confirmed *DUXC* gene in the gene annotation file, and we validated that RNA-seq reads mapped to the novel first exon we identified. Interestingly, the signal for *DUXC* increased at the 4-cell stage.

Comparison of control and α -amanitin treatment groups revealed that expressed *DUXC* levels do not depend on the zygotic transcription, as we observed similar levels between groups for the 2- and 4-cell stage. A similar observation was made for only 8-cell bovine embryos using the same α -amanitin-mediated transcription inhibition (C. Zhou et al., 2023).

For the 8-cell embryos, one control replicate showed similar levels with the replicates of TBEs, whereas two replicates showed lower levels in the control group. Reduced *DUXC* transcript levels were observed for later stages, coinciding with the 8-cell stage and subsequent stages, indicating that the lower levels

observed in the control groups are probably due to the degradation of *DUXC* transcripts that needs to occur at this point of development. The differences among control 8-cell stage embryos are likely due to the slight differences of developmental time point. Although the expression of *DUXC* was reduced to the levels observed at the oocyte stages, samples of inner cell mass showed slightly higher levels of *DUXC* mRNA.

Transcriptionally inhibited 8-cell embryos showed higher *DUXC* levels as compared to the control 8-cell stage embryos, indicating the need for zygotic transcription for the clearance of *DUXC* mRNA levels later in preimplantation embryo development. This phenomenon was more clearly observed between control and DRBW embryos, the latter of which are embryos with minor EGA inhibition. The highest relative expression of *DUXC* was observed in the 8-cell embryos, in which minor EGA was inhibited. Transcripts that were derived from the bovine minor EGA were probably involved in the degradation of *DUXC* transcripts.

Along with *DUXC*, we also demonstrated the relative expressions of *ARGFX*, *DUXA*, *LEUTX*, *NOBOX*, *TPRX1*, *TPRX2*, and *TPRX3*, for which we had cDNA-confirmed gene annotations. As expected, *NOBOX* expression was mostly confined to the oocyte stage, and its transcript levels were increased at TBE embryos, highlighting the link between maternal mRNA clearance and embryonic transcription. *DUXA*, *LEUTX*, *TPRX1*, *TPRX2*, and *TPRX3* exhibited similar temporal expression patterns. The relative expression of these PRDL homeobox genes, along with *ARGFX* were increased at the 4-cell stage in the absence of α -amanitin, suggesting that their expression is of zygotic genome origin. Interestingly, *ARGFX* expression at the 8-cell stage was elevated in control DRBW embryos, pointing to a specific factor derived from minor EGA needed for its enhanced activation. Furthermore, we utilized our recent dataset (Yaşar et al., 2025) that allows spike-in normalization for a more reliable quantification. We showed that *DUXC* mRNAs were at peak levels at the 2-cell stage, the earliest stage after fertilization in this dataset.

5.3.4 Genomic organization of bovine *DUXC* locus

We set out to characterize the genomic organization of bovine *DUXC*, which was previously shown to be in a pericentromeric region (Leidenroth et al., 2012). These earlier experiments utilized hybridization-based experiments to show the chromosomal location of *DUXC*. The resolution at the repetitive genomic loci, including pericentromeric regions, has been improved in newer bovine genome assemblies. For this reason, we examined the *DUXC* locus in the latest version of the bovine reference genome (ARS-UCD2.0) as well as in an independently assembled recent bovine genome called NCBA_BosT1.0 (T. T. Li et al., 2023) using our experimentally-derived full-length *DUXC* sequence. Since the former and latter were Hereford and Holstein breed, respectively, we referred to them by their breed for the sake of simplicity.

Our initial approach was to perform *in silico* PCR by designing primers that could anneal all repeat units. This approach allowed us to identify six repeat units for the Hereford genome, whereas 18 repeat units were revealed in the Holstein genome. The length of the fragments generated by *in silico* PCR showed that the median of repeat units was greater in Holstein breed bovine. We also observed differences in the lengths of individual repeat units, with some of the repeat units being around 5500 bp, while the median of the Holstein repeat units was 5900 bp.

Apart from the complete repeats, at each end of the array in both breeds, there were incomplete repeats. We were unable to detect these incomplete repeats among the *in silico* PCR-amplified fragments, as they did not have the sequences for primers to anneal to. For convenience, we called the incomplete repeat at the telomere end Unit 1, with the upstream complete repeat unit called Unit 2.

We depicted the *DUXC* locus by overlaying our *DUXC* gene structure, similar to the approach used for *DUX4* in the human D4Z4 macrosatellite (Lemmers et al., 2010), thereby providing a detailed genomic map of this region. Based on our *DUXC* gene structure, the second half of the last exon was lost in Unit 1. However, in the genomic region where the second half of the *DUXC* last exon would be located, we identified a putative polyadenylation signal both in Holstein and Hereford bovine genomes. This polyadenylation signal (ATTAAA) was identical to the one observed at the 3' end of *DUX4*, which is thought to give *DUX4* transcripts stability in FSHD cells (Dixit et al., 2007). This signal is commonly found in the 3' UTRs of mRNAs (Beaudoing et al., 2000). Similarly, we observed less common polyadenylation signals (AATACA, AAGAAA) in the internal repeat units, positioned downstream of the last exon.

To assess the variation among the repeat units more systematically, we performed multiple sequence alignments using the repeat units for each breed and created a tree to show the relationship among them. Hereford internal units were highly similar but still showed minor differences reflected by the branches on the tree. The incomplete unit stood out with much longer branches. Most of the Holstein internal units were similar to each other, but some had much longer branches. Two of these units formed a separate cluster with the longest branch, except for the incomplete unit at the non-telomere end. In summary, we utilized our cDNA-derived *DUXC* sequence in combination with *in silico* PCR and manual inspection of the *DUXC* genomic region to characterize its genomic organization and generate a genomic map for this locus.

5.3.5 Identifying the repeat unit origin of *DUXC* mRNA

We set out to determine whether expressed *DUXC* mRNAs were derived from the internal or distal units. Although the second half of *DUXC* exon 5 had a deviated sequence, it contained a polyadenylation signal that human *DUX4* was shown to utilize. Since we needed to assess the expression of the 3' end of *DUXC*, we searched for RNA-seq datasets with bovine early embryonic stages and selected a dataset (L. Zhu et al., 2022) that was generated with a method utilizing resulting in high 3' end coverage (Picelli et al., 2013).

We utilized two indicators to determine whether the internal or distal *DUXC* copy is expressed. One of these indicators was the second half of the exon 5, which differed between the internal and distal copy. The other indicator was the sequence found in the exon-intron junction of exon 5. This 51 bp sequence was absent in the distal copy but present in the internal copy in the Holstein genome. For this reason, we initially set out to search for reads spanning exon 4 and 5. We identified reads spanning exon 4 and 5, which indicated that the internal *DUXC* is expressed, as this specific region RNA-seq reads aligned to was missing in the distal copy. We found these reads in both the 2-cell and 8-cell stage embryos. As per the other indicator, we also concluded that expressed *DUXC* transcripts were derived from the internal copies. This is because we identified reads aligning to the portion of the exon 5 found only in the internal but not the distal copies.

Although we could not identify RNA-seq reads split between exon 4 and exon 5 in the distal copy, we wanted to investigate the possibility of a further downstream exon for the distal *DUXC* copy. Therefore, we searched for reads split between the exon 4 and any downstream sequence until the telomere end. However, prior to our search, we wanted to ensure that the Holstein bovine we used had an assembled telomere. By using a computational tool (Brown et al., 2025), we verified the presence of the telomere in chromosome 7 in this genome assembly. We extracted a genomic region spanning the telomere end and the distal *DUXC* copy, excluding the internal *DUXC* copies, to which we aligned RNA-seq reads using a mapper (Kim et al., 2019), allowing a maximum intron length of 500,000 bp. Notably, the genomic distance between the distal *DUXC* and telomere is much shorter than 500,000 bp. Our search did not result in any reads between the exon 4 of the distal *DUXC* copy and any potential downstream exon.

Using our *DUXC* cDNA, we annotated *DUXC* in the distal and internal regions. Based on our approach, the distal *DUXC* exhibited a shorter last exon, whereas the internal *DUXC* contained a longer last exon. This indicates that the *DUXC* cDNA we isolated was also derived from the internal *DUXC* copies. An alternative approach we took for annotating *DUXC* was by transcript assembly that utilized RNA-seq reads from the same dataset. This approach revealed the same *DUXC* structure we experimentally confirmed. The same gene structure was obtained for both 2-cell and 8-cell stage embryos. Although we applied the same analysis for other developmental stages, including GV and MII oocytes as well as the morula and blastocyst stage embryos, the number of reads was very low for these stages. One exception is the blastocyst stage, where tens of reads mapped to the *DUXC* locus. However, these reads not only aligned to *DUXC* exons but also to introns as well as exon-intron junctions. This likely represents the expression of a distinct genomic feature. Furthermore, transcript assembly at the blastocyst stage did not recapitulate previously defined *DUXC* gene structure but instead yielded a single-exon unit, supporting the presence of a separate genomic feature at this locus. Taken together, we demonstrated lines of evidence to state that the distal *DUXC* is not expressed, and the observed *DUXC* mRNAs are derived from one or more internal *DUXC* copies.

5.3.6 Genomic organization of the *DUXC* locus across eight bovine breeds

We searched for publicly available genome assemblies to analyze the *DUXC* locus in different bovine breeds. To analyze the genomic organization of the *DUXC* locus, we first extracted this locus from each genome assembly and then examined its composition in terms of *DUXC* repeat units. For this purpose, we used Holstein unit 2, which we produced via *in silico* PCR, and aligned it to chromosome 7, allowing defining the boundaries of the *DUXC* locus. After extracting these loci, we performed pairwise alignment and provided *DUXC* locus-wide comparison across breeds. Next, we aligned the same unit 2 to each extracted *DUXC* locus to reveal how *DUXC* repeat units were organized in the genome assembly of each breed.

We showed that the Wagyu genome assembly contained the longest *DUXC* locus, in which we were able to annotate 28 copies of *DUXC*. However, we were unable to extract the *DUXC* locus from the Yunling breed, probably because the resolution of the Yunling genome assembly was not sufficient to cover this genomic region. Jersey, Hanwoo, Holstein, Hereford, and Tibetan showed close similarity to each other, with most regions showing sequence identity higher than 80% while some regions reached up to 98%. Since Jersey had a relatively small *DUXC* locus in its assembly, it only allowed annotation of one *DUXC* copy. The extracted genomic locus from the Mongolian assembly was the shortest among all, with the shared sequence identity between this region and the *DUXC* repeat unit being 77–78%. No *DUXC* was annotated in this genomic locus, whereas all other breeds had at least one *DUXC* gene annotated using our *DUXC* cDNA sequence.

5.4 Study IV: Effect of oxygen levels on bovine early embryo development

5.4.1 Blastocyst formation rates of embryos grown under normoxia, hypoxia and ultrahypoxia

To study the effect of oxygen levels on *in vitro* produced bovine embryos, pre-implantation bovine embryos were exposed to three distinct conditions. Two of these conditions included performing the entire embryo culture either under 20% atmospheric O₂ or 6% physiological O₂ levels, referred to as normoxia and hypoxia, respectively. The third condition, which was referred to as ultrahypoxia, was identical to the hypoxia condition except for maintaining the culture under 2% O₂ after the 16-cell stage. In the experiment performed to calculate cleavage and blastocyst formation rates, each condition had at least two replicates, with at least two of the replicates containing 50 zygotes at the beginning of embryo culture. Similar cleavage rates were observed for each condition, with normoxia, hypoxia, and ultrahypoxia showing around 85%, 85%, and 79% cleavage rates,

respectively. The formation of blastocyst rates at day 7 were 8%, 33%, and 2.1%, and at day 8, it was 13%, 36%, and 4.6% for normoxia, hypoxia, and ultrahypoxia conditions, respectively. This showed that the normoxia and ultrahypoxia conditions decreased the proportions of embryos that reached the blastocyst stage. The finding relating to the normoxia condition is in line with an earlier report showing that the efficiency of human IVF procedures is lower when embryos are cultured under normoxia conditions (Bontekoe S & Mastenbroek, 2012). However, the finding relating to the ultrahypoxia condition is in contrast to the earlier studies conducted on human and mouse embryos (De Munck et al., 2019; Kaskar et al., 2019; Y. Yang et al., 2016).

5.4.2 Transcriptome profiling of embryos grown under normoxia, hypoxia and ultrahypoxia

To investigate the effect of oxygen levels on the transcriptome, STRT-N RNA-seq was performed using zygote, 4-cell, 8-cell, 16-cell stage, and blastocyst stage for the normoxia and hypoxia conditions, whereas for the ultrahypoxia condition, only the blastocyst stage was used. To assess the quality of the RNA-seq experiment, temporal expression levels of known genes were plotted. As expected, expression levels of pluripotency markers *NANOG* and *POU5F1* (*OCT4*) increased at the blastocyst stage (Canizo et al., 2019; Khan et al., 2012; Kuijk et al., 2008), and the *GDF9* expression was observed in oocytes (Lonergan et al., 2003; Pennetier et al., 2004; Sendai et al., 2001). The expression of *SLC34A2*, a cell surface antigen specific to 8-cell-like cells (Yoshihara et al., 2022), was observed in post-EGA stages in bovine embryos. UMAP analysis showed clear separation of pre- and post-EGA stages.

Initially, differential gene expression analysis was performed for normoxia and hypoxia conditions by comparing the transcriptome of the 8- and 16-cell stage to that of the zygote stage. The first comparison revealed downregulation of 47 and 732 genes for normoxia and hypoxia. Maternal-to-zygotic transition (MZT), which is characterized by not only the burst of zygotic gene transcription but also maternal mRNA clearance, is an integral part of embryo development. Embryos grown under atmospheric O₂ levels might have partially failed to clear maternal transcripts.

The second comparison, which included the zygote-to-16-cell transition, showed a similar number of upregulated genes, with 253 genes upregulated in hypoxia and 260 genes upregulated in normoxia. The number of commonly upregulated genes in both conditions was 167. The number of upregulated genes exclusive to hypoxia was 86, including the genes *AHR*, *EOMES*, *MYC*, and *TFAP2C*, whose 16-cell stage-specific expression was reported (Yaşar et al., 2025). The number of upregulated genes identified as normoxia-specific was 93. Gene ontology (GO) analysis of upregulated genes in the 16-cell stage for both conditions showed similar general terms relating to transcription and translation. A high number of downregulation was observed during the zygote-to-16-cell tran-

sition, with 5,785 and 4,829 genes detected for hypoxia and normoxia conditions, respectively. KEGG pathway analysis of the downregulated genes showed similar terms for both conditions.

Later, differentially expressed genes were identified for all conditions. In order to compare the effect of oxygen levels across conditions, transcriptome profiles of blastocyst stage embryos were compared to those of 16-cell stage embryos grown under normoxia, hypoxia, or ultrahypoxia conditions. The number of upregulated genes during the 16-cell-to-blastocyst transition was 1,991 and 1,967 under hypoxia and ultrahypoxia conditions, respectively, whereas only about a quarter of this number was observed under normoxia. This may indicate an inefficient activation of hundreds of genes under normoxic conditions at the blastocyst stage. On the other hand, during the same developmental transition, the numbers of downregulated genes were 222, 1338, and 663 under hypoxia, ultrahypoxia, and normoxia, respectively.

KEGG pathway analysis for the upregulated genes identified the oxidative phosphorylation pathway. However, it also revealed that cellular processes involved in TCA cycle genes were detected under hypoxia and ultrahypoxia, but not under normoxia conditions. The metabolism of early embryos before compaction primarily depends on the oxidation of pyruvate, lactate, and amino acids, whereas following compaction, at stages including the blastocyst stage, oxygen consumption increases and metabolism dependency shifts towards glucose, requiring the activity of components involved in the TCA cycle (Gardner & Wale, 2013). These results show that oxygen levels under which embryos are cultured have a primary effect on embryo metabolism at post-compaction stages. The low oxygen levels thought to be present inside the uterus (Burton et al., 2021) therefore, have an impact on establishing the proper metabolic state required for blastocyst development.

6 CONCLUSIONS

In conclusion, the studies included in this thesis primarily elaborate on the identification, putative roles, and characterization of PRD-like homeobox genes in bovine preimplantation embryos. Conclusions drawn from these studies can be summarized as follows:

- Several PRD-like homeobox genes, predominantly active in human pre-implantation embryos, are also active in bovine preimplantation embryos.
- Molecular cloning is an efficient approach to initially determine the accurate structure of genes before their functional characterization.
- Bovine TPRX3 is a homeodomain transcription factor with a gene activation capacity.
- Regulatory mechanisms during early development are species-specific, with distinct sets of genes carrying out particular roles in each organism.
- 5'-targeted RNA-seq approaches help identify novel transcription start sites and *de novo* regulatory motifs, as well as regulatory proteins that modulate gene regulatory effects through such motifs.
- Bovine *DUXC* has a gene structure distinct from the proposed gene models.
- Genomic organization of the *DUXC* array in eight distinct cattle breeds is similar to that of human *DUX4*.
- Unlike the *DUX4* transcripts that are derived from the distal unit in human FSHD phenotypes, *DUXC* transcripts in early bovine embryos are derived from the internal units.
- The detection of *DUXC* mRNAs in early bovine embryos is unaffected by the inhibition of embryonic transcription.
- Clearance of *DUXC* mRNAs at the 8-cell stage requires minor EGA.
- Oxygen levels to which embryos are exposed during *in vitro* culture change their metabolism, specifically causing slightly delayed EGA in normoxia (20% O₂).

SUMMARY IN ESTONIAN

Täiustatud kromosoomanalüüsi meetodid embrüo kvaliteedi ja loote tervise hindamiseks

Inimese mudelorganismide nagu hiire, ja veise embrüonaalse arengu esimene nädal peale viljastumist on oluline arenguaken, kuna just sel perioodil käivituvad mitmed protsessid ja mehhanismid (näiteks emapoolsete mRNA-de translatsioon ja embrüo genoomi aktivatsioon), mis suunavad organismi arengut kuni selle täieliku väljakujunemiseni. Nende mehhanismide parem mõistmine aitab täiustada inimeste viljatusravis kasutatavaid abistava reproduktsiooni tehnoloogiaid ning parandada loomakasvatuse reproduktiivtehnoloogiate tõhusust. Käesolev doktoritöö keskendub veise embrüo preimplantatsioonilisele e pesastumiseelsele arenguperioodile, pöörates erilist tähelepanu PRDL-klassi homeoboksgeenidele, mis kodeerivad transkriptsioonifaktoreid. PRDL geenid tuvastati esmakordselt inimese embrüotes nende varajasele arengujärgule iseloomuliku aktiivsuse põhjal.

Töö hüpoteesideks on, et veise varajane embrüonaalne areng on reguleeritud PRDL geenide poolt kodeeritud transkriptsiooni- ja kofaktorite võrgustiku kaudu, näiteks DUXC, ning embrüonaalset arengut mõjutavad nii rakusisesed transkriptsioonilised programmid kui ka embrüote kasvatamisel kasutatavad kultiveerimistingimused, sealhulgas hapniku kontsentratsioon. Töö eesmärkideks oli tuvastada ja iseloomustada olulistel pesastumiseelsetel etappidel avalduvaid PRDL geene, täpsemalt uurida DUXC avaldumist ja võimalikku regulatoorset rolli. Samuti kaardistada veise varajase embrüonaalse arengu transkriptoomi dünaamikat, leida veise embrüo genoomi aktivatsiooni (EGA) regulaatoreid ning hinnata, kuidas *in vitro* kultiveerimistingimused mõjutavad embrüo transkriptoomi.

Doktoritöös käsitletud uurimuste esimeses osas kirjeldati seitsme PRDL homeoboksgeeni veise ortoloogi struktuuri ning nende ekspressiooni veise varajases embrüonaalses arengus. Lisaks näidati, et TPRX3 (ingl *Tetrapeptide Repeat Homeobox 3*) toimib transkriptsiooni aktivaatorina.

Teises uuringus kaardistati veise arenevates munarakkudes ja varajastes embrüotes oletatavad transkriptsiooni alguskohad. Kasutades valitud meetodi võimet välja selgitada potentsiaalseid promootorialasid, õnnestus DNA motiivi-analüüsi abil tuvastada mitmeid transkriptsioonifaktoreid, mis võivad osaleda veise suuremahulises embrüonaalse genoomi aktivatsiooni (EGA) käivitamises.

Kolmas uuring keskendus veise DUXC lookuse põhjalikule iseloomustamisele. Töös analüüsiti 8-rakulistest *in vitro* viljastatud veise embrüotest kloonitud DUXC cDNA-d. Saadud tulemused toetavad varasemaid evolutsioonilisi käsitusi ning võimaldavad täpsustada DUXC geeni annotatsiooni. Tuginedes kaheksa erineva veisetõu genoomiandmetele, näidati, et DUXC geeniklastri genoomne ülesehitus sarnaneb DUX4 omaga. Samuti selgu, et DUXC mRNA tuvastamine varajastes veise embrüotes ei sõltu embrüonaalse transkriptsiooni pärssimisest, samas kui DUXC mRNA taseme langus 8-rakulise embrüo staadiumis eeldab väiksemahulise EGA toimumist.

Doktoritöö viimases osas käsitleti hapnikutaseme mõju varajasele embrüonaalsele arengule. Tulemused näitasid, et normoksilised kultiveerimistingimused põhjustavad viivituse veise suuremahulises EGA-s.

Kokkuvõtteks aitavad käesolevas doktoritöös esitatud uuringud süvendada arusaamist PRDL-faktorite rollist emapoolselt ehk maternaalselt geeniekspressioonilt sügoodi geeniekspressioonile üleminekul ning veise suuremahulise EGA regulatsioonil.

REFERENCES

- Abugessaisa, I., Hasegawa, A., Noguchi, S., Cardon, M., Watanabe, K., Takahashi, M., Suzuki, H., Katayama, S., Kere, J., & Kasukawa, T. (2022). SkewC: Identifying cells with skewed gene body coverage in single-cell RNA sequencing data. *IScience*, 25(2), 103777. <https://doi.org/10.1016/J.ISCI.2022.103777>
- Adamson, G. D., Creighton, P., de Mouzon, J., Zegers-Hochschild, F., Dyer, S., & Chambers, G. M. (2025). How many infants have been born with the help of assisted reproductive technology? *Fertility and Sterility*, 124(1), 40–50. <https://doi.org/10.1016/J.FERTNSTERT.2025.02.009>
- Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10), 720–731. <https://doi.org/10.1038/nrg3293>
- Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltmann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/NAR/GKY379>
- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., Lehrach, H., Mardis, E. R., McVean, G. A., Nickerson, D. A., ... Peterson, J. L. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), 1061–1073. <https://doi.org/10.1038/nature09534>
- Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics*, 2(2), 130–138. <https://doi.org/10.1038/35052563>
- Arends, T., Tsuchida, H., Adeyemi, R. O., & Tapscott, S. J. (2024). DUX4-induced HSATII transcription causes KDM2A/B–PRC1 nuclear foci and impairs DNA damage response. *Journal of Cell Biology*, 223(5), e202303141. <https://doi.org/10.1083/JCB.202303141>
- Asami, M., Lam, B. Y. H., Ma, M. K., Rainbow, K., Braun, S., VerMilyea, M. D., Yeo, G. S. H., & Perry, A. C. F. (2022). Human embryonic genome activation initiates at the one-cell stage. *Cell Stem Cell*, 29(2), 209–216.e4. <https://doi.org/10.1016/J.STEM.2021.11.012>
- Assal, N., & Lin, M. (2021). PCR procedures to amplify GC-rich DNA sequences of *Mycobacterium bovis*. *Journal of Microbiological Methods*, 181, 106121. <https://doi.org/10.1016/J.MIMET.2020.106121>
- Bahn, J. H., Lee, J. H., Li, G., Greer, C., Peng, G., & Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research*, 22(1), 142–150. <https://doi.org/10.1101/GR.124107.111>
- Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. <https://doi.org/10.1093/NAR/GKV416>
- Bakhtari, A., & Ross, P. J. (2014). DPPA3 prevents cytosine hydroxymethylation of the maternal pronucleus and is required for normal development in bovine embryos. *Epigenetics*, 9(9), 1271–1279. <https://doi.org/10.4161/EPI.32087>
- Barnes, F. L., & First, N. L. (1991). Embryonic transcription in in vitro cultured bovine embryos. *Molecular Reproduction and Development*, 29(2), 117–123. <https://doi.org/10.1002/MRD.1080290205>

- Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., & Gautheret, D. (2000). Patterns of Variant Polyadenylation Signal Usage in Human Genes. *Genome Research*, 10(7), 1001–1010. <https://doi.org/10.1101/GR.10.7.1001>
- Beerli, R. R., Segal, D. J., Dreier, B., & Barbas, C. F. (1998). Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), 14628–14633. <https://doi.org/10.1073/PNAS.95.25.14628>
- Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The Mammalian Epigenome. *Cell*, 128(4), 669–681. <https://doi.org/10.1016/J.CELL.2007.01.033>
- Betts, D. H., & Madan, P. (2008). Permanent embryo arrest: molecular and cellular concepts. *Molecular Human Reproduction*, 14(8), 445–453. <https://doi.org/10.1093/MOLEHR/GAN035>
- Bogliotti, Y. S., Chung, N., Paulson, E. E., Chitwood, J., Halstead, M., Kern, C., Schultz, R. M., & Ross, P. J. (2020). Transcript profiling of bovine embryos implicates specific transcription factors in the maternal-to-embryo transition. *Biology of Reproduction*, 102(3), 671–679. <https://doi.org/10.1093/biolre/ioz209>
- Bogliotti, Y. S., Wu, J., Vilarino, M., Okamura, D., Soto, D. A., Zhong, C., Sakurai, M., Sampaio, R. V., Suzuki, K., Izpisua Belmonte, J. C., & Ross, P. J. (2018). Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9), 2090–2095. <https://doi.org/10.1073/PNAS.1716161115>
- Bontekoe S, M. E. van W. M. S. S. R. S., & Mastenbroek, S. (2012). Low oxygen concentrations for embryo culture in assisted reproductive technologies. *Cochrane Database of Systematic Reviews*, (7), CD008950. <https://doi.org/10.1002/14651858.CD008950.pub2>
- Booth, H. A. F., & Holland, P. W. H. (2007). Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. *Gene*, 387(1–2), 7–14. <https://doi.org/10.1016/J.GENE.2006.07.034>
- Boroviak, T., Stirparo, G. G., Dietmann, S., Hernando-Herraez, I., Mohammed, H., Reik, W., Smith, A., Sasaki, E., Nichols, J., & Bertone, P. (2018). Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development*, 145(21), dev167833. <https://doi.org/10.1242/dev.167833>
- Boskovic, N., Ivask, M., Yazgeldi Gunaydin, G., Yaşar, B., Katayama, S., Salumets, A., Org, T., Kurg, A., Lundin, K., Tuuri, T., Daub, C. O., & Kere, J. (2025). Oxygen level alters energy metabolism in bovine preimplantation embryos. *Scientific Reports*, 15(1), 11327. <https://doi.org/10.1038/s41598-025-95990-z>
- Boskovic, N., Yazgeldi, G., Ezer, S., Tervaniemi, M. H., Inzunza, J., Deligiannis, S. P., Yaşar, B., Skoog, T., Krjutškov, K., Katayama, S., & Kere, J. (2023). Optimized single-cell RNA sequencing protocol to study early genome activation in mammalian preimplantation development. *STAR Protocols*, 4(3), 102357. <https://doi.org/10.1016/j.xpro.2023.102357>
- Bosnakovski, D., Choi, S. H., Strasser, J. M., Toso, E. A., Walters, M. A., & Kyba, M. (2014). High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. *Skeletal Muscle*, 4(1), 4. <https://doi.org/10.1186/2044-5040-4-4>
- Bosnakovski, D., Daughters, R. S., Xu, Z., Slack, J. M. W., & Kyba, M. (2009). Biphasic Myopathic Phenotype of Mouse DUX, an ORF within Conserved FSHD-Related Repeats. *PLOS ONE*, 4(9), e7003. <https://doi.org/10.1371/JOURNAL.PONE.0007003>

- Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., & Kyba, M. (2008). DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Experimental Neurology*, 214(1), 87–96. <https://doi.org/10.1016/J.EXPNEUROL.2008.07.022>
- Bosnakovski, D., Toso, E. A., Ener, E. T., Gearhart, M. D., Yin, L., Lüttmann, F. F., Magli, A., Shi, K., Kim, J., Aihara, H., & Kyba, M. (2023). Antagonism among DUX family members evolved from an ancestral toxic single homeodomain protein. *IScience*, 26(10), 107823. <https://doi.org/10.1016/J.ISCI.2023.107823>
- Bosnakovski, D., Xu, Z., Ji Gang, E., Galindo, C. L., Liu, M., Simsek, T., Garner, H. R., Agha-Mohammadi, S., Tassin, A., Coppée, F., Belayew, A., Perlingeiro, R. R., & Kyba, M. (2008). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO Journal*, 27(20), 2766–2779. <https://doi.org/10.1038/emboj.2008.201>
- Brauckmann, S. (2015). The Many Spaces of Karl Ernst von Baer. *Biological Theory*, 3(1), 85–89. <https://doi.org/10.1162/BIOT.2008.3.1.85>
- Braude, P., Bolton, V., & Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature*, 332(6163), 459–461. <https://doi.org/10.1038/332459a0>
- Broad Institute. (2019). *Picard Tools*. <http://broadinstitute.github.io/picard/>
- Brown, M. R., Manuel Gonzalez de La Rosa, P., & Blaxter, M. (2025). tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, 41(2), btaf049. <https://doi.org/10.1093/BIOINFORMATICS/BTAF049>
- Brueckner, F., & Cramer, P. (2008). Structural basis of transcription inhibition by α -amanitin and implications for RNA polymerase II translocation. *Nature Structural & Molecular Biology*, 15(8), 811–818. <https://doi.org/10.1038/nsmb.1458>
- Burglin, T. R. (1994). A comprehensive classification of homeobox genes. In *Guidebook to the Homeobox Genes* (pp. 27–71). Oxford University Press/Oxford. <https://doi.org/10.1093/OSO/9780198599395.003.0003>
- Bürglin, T. R. (2011). Homeodomain Subtypes and Functional Diversity. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (pp. 95–122). Springer Netherlands. https://doi.org/10.1007/978-90-481-9069-0_5
- Bürglin, T. R., & Affolter, M. (2016). Homeodomain proteins: an update. *Chromosoma*, 125(3), 497–521. <https://doi.org/10.1007/s00412-015-0543-8>
- Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G., & Noll, M. (1989). Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO Journal*, 8(4), 1183–1190. <https://doi.org/10.1002/J.1460-2075.1989.TB03490.X>
- Burton, G. J., Cindrova-Davies, T., Yung, H. W., & Jauniaux, E. (2021). HYPOXIA AND REPRODUCTIVE HEALTH: Oxygen and development of the human placenta. *Reproduction*, 161(1), F53–F65. <https://doi.org/10.1530/REP-20-0153>
- Calhaz-Jorge, C., De Geyter, C., Kupka, M. S., De Mouzon, J., Erb, K., Mocanu, E., Motrenko, T., Scaravelli, G., Wyns, C., Goossens, V., Gliozheni, O., Strohmer, H., Petrovskaya, E., Tishkevich, O., Wyns, C., Bogaerts, K., Antonova, I., Vrcic, H., Ljiljak, D., ... Baranowski, R. (2016). Assisted reproductive technology in Europe, 2012: results generated from European registers by ESHRE. *Human Reproduction*, 31(8), 1638–1652. <https://doi.org/10.1093/HUMREP/DEW151>
- Camous, S., Kopecný, V., & Fléchon, J. E. (1986). Autoradiographic detection of the earliest stage of [3H]-uridine incorporation into the cow embryo. *Biology of the Cell*, 58(3), 195–200. <https://doi.org/10.1111/J.1768-322X.1986.TB00506.X>

- Campbell, A. E., Dyle, M. C., Albanese, R., Matheny, T., Sudheendran, K., Cortázar, M. A., Forman, T., Fu, R., Gillen, A. E., Caruthers, M. H., Floor, S. N., Calviello, L., & Jagannathan, S. (2023). Compromised nonsense-mediated RNA decay results in truncated RNA-binding protein production upon DUX4 expression. *Cell Reports*, 42(6), 112642. <https://doi.org/10.1016/J.CELREP.2023.112642>
- Campbell, A. E., Shadle, S. C., Jagannathan, S., Lim, J. W., Resnick, R., Tawil, R., Van Der Maarel, S. M., & Tapscott, S. J. (2018). NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *ELife*, 7, e31023. <https://doi.org/10.7554/ELIFE.31023>
- Canizo, J. R., Ynsaurrealde Rivolta, A. E., Vazquez Echegaray, C., Suvá, M., Alberio, V., Aller, J. F., Guberman, A. S., Salamone, D. F., Alberio, R. H., & Alberio, R. (2019). A dose-dependent response to MEK inhibition determines hypoblast fate in bovine embryos. *BMC Developmental Biology*, 19(1), 13. <https://doi.org/10.1186/S12861-019-0193-9>
- Carlson, M. (2019). org.Bt.eg.db: Genome wide annotation for Bovine. In *R package version 3.8.2*.
- Chen, X., Ke, Y., Wu, K., Zhao, H., Sun, Y., Gao, L., Liu, Z., Zhang, J., Tao, W., Hou, Z., Liu, H., Liu, J., & Chen, Z. J. (2019). Key role for CTCF in establishing chromatin structure in human embryos. *Nature*, 576(7786), 306–310. <https://doi.org/10.1038/s41586-019-1812-0>
- Chen, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., & Xu, Y. (2021). Structures of the human Mediator and Mediator-bound preinitiation complex. *Science*, 372(6546), eabg0635. <https://doi.org/10.1126/SCIENCE.ABG0635>
- Chen, Z., Xie, Z., & Zhang, Y. (2021). DPPA2 and DPPA4 are dispensable for mouse zygotic genome activation and pre-implantation development. *Development*, 148(24), dev200178. <https://doi.org/10.1242/DEV.200178>
- Chen, Z., & Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nature Genetics*, 51(6), 947–951. <https://doi.org/10.1038/s41588-019-0418-7>
- Chew, G. L., Campbell, A. E., De Neef, E., Sutliff, N. A., Shadle, S. C., Tapscott, S. J., & Bradley, R. K. (2019). DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Developmental Cell*, 50(5), 658–671.e7. <https://doi.org/10.1016/J.DEVCEL.2019.06.011>
- Chitwood, J. L., Burrue, V. R., Halstead, M. M., Meyers, S. A., & Ross, P. J. (2017). Transcriptome profiling of individual rhesus macaque oocytes and preimplantation embryos. *Biology of Reproduction*, 97(3), 353–364. <https://doi.org/10.1093/biolre/iox114>
- Chitwood, J. L., Rincon, G., Kaiser, G. G., Medrano, J. F., & Ross, P. J. (2013). RNA-seq analysis of single bovine blastocysts. *BMC Genomics*, 14(1), 350. <https://doi.org/10.1186/1471-2164-14-350>
- Chodosh, L. A., Fire, A., Samuels, M., & Sharp, P. A. (1989). 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole Inhibits Transcription Elongation by RNA Polymerase II in Vitro. *Journal of Biological Chemistry*, 264(4), 2250–2257. [https://doi.org/10.1016/S0021-9258\(18\)94169-4](https://doi.org/10.1016/S0021-9258(18)94169-4)
- Choi, S. H., Gearhart, M. D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N., & Kyba, M. (2016). DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Research*, 44(11), 5161–5173. <https://doi.org/10.1093/NAR/GKW141>

- Clapp, J., Mitchell, L. M., Bolland, D. J., Fantes, J., Corcoran, A. E., Scotting, P. J., Armour, J. A. L., & Hewitt, J. E. (2007). Evolutionary Conservation of a Coding Function for D4Z4, the Tandem DNA Repeat Mutated in Facioscapulohumeral Muscular Dystrophy. *The American Journal of Human Genetics*, 81(2), 264–279. <https://doi.org/10.1086/519311>
- Cox, C. M., Thoma, M. E., Tchangalova, N., Mburu, G., Bornstein, M. J., Johnson, C. L., & Kiarie, J. (2022). Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis. *Human Reproduction Open*, 2022(4), hoac051. <https://doi.org/10.1093/HROPEN/HOAC051>
- Cuthbert, J. M., Russell, S. J., White, K. L., & Benninghoff, A. D. (2019). The maternal-to-zygotic transition in bovine in vitro-fertilized embryos is associated with marked changes in small non-coding RNAs. *Biology of Reproduction*, 100(2), 331–350. <https://doi.org/10.1093/BIOLRE/IOY190>
- Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M. N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P. J., Coote, D., ... Hayes, B. J. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46(8), 858–865. <https://doi.org/10.1038/NG.3034>
- Dainat, J. (2024). *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format* (V1.4.1). Zenodo. <https://doi.org/10.5281/zenodo.3552717>
- Das, S., & Chadwick, B. P. (2016). Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLOS ONE*, 11(7), e0160022. <https://doi.org/10.1371/JOURNAL.PONE.0160022>
- De Iaco, A., Coudray, A., Duc, J., & Trono, D. (2019). DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Reports*, 20(5), EMBR201847382. <https://doi.org/10.15252/embr.201847382>
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics*, 49(6), 941–945. <https://doi.org/10.1038/ng.3858>
- De Iaco, A., Verp, S., Offner, S., Grun, D., & Trono, D. (2020). DUX is a non-essential synchronizer of zygotic genome activation. *Development*, 147(2), dev177725. <https://doi.org/10.1242/DEV.177725>
- De Munck, N., Janssens, R., Segers, I., Tournaye, H., Van De Velde, H., & Verheyen, G. (2019). Influence of ultra-low oxygen (2%) tension on in-vitro human embryo development. *Human Reproduction*, 34(2), 228–234. <https://doi.org/10.1093/HUMREP/DEY370>
- Deng, Q., Ramsköld, D., Reinius, B., & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*, 343(6167), 193–196. <https://doi.org/10.1126/science.1245316>
- Derelle, R., Lopez, P., Guyader, H. Le, & Manuel, M. (2007). Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evolution and Development*, 9(3), 212–219. <https://doi.org/10.1111/J.1525-142X.2007.00153.X>
- Deschamps-Francoeur, G., Simoneau, J., & Scott, M. S. (2020). Handling multi-mapped reads in RNA-seq. *Computational and Structural Biotechnology Journal*, 18, 1569–1576. <https://doi.org/10.1016/j.csbj.2020.06.014>
- Deutekom, J. C. T. va., Wljmenga, C., Tlenhoven, E. A. E. va., Gruter, A. M., Hewitt, J. E., Padberg, G. W., van Ommen, G. J. B., Hofker, M. H., & Fronts, R. R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Human Molecular Genetics*, 2(12), 2037–2042. <https://doi.org/10.1093/hmg/2.12.2037>

- Dixit, M., Anseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Mattéotti, C., Van Acker, A. M., Leo, O., Figlewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppée, F., & Chen, Y. W. (2007). DUX4, a candidate gene of facio-scapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), 18157–18162. <https://doi.org/10.1073/PNAS.0708659104>
- Dobson, A. T., Raja, R., Abeyta, M. J., Taylor, T., Shen, S., Haqq, C., & Reijo Pera, R. A. (2004). The unique transcriptome through day 3 of human preimplantation development. *Human Molecular Genetics*, 13(14), 1461–1470. <https://doi.org/10.1093/hmg/ddh157>
- Dong, J., Hu, Y., Fan, X., Wu, X., Mao, Y., Hu, B., Guo, H., Wen, L., & Tang, F. (2018). Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biology*, 19(1), 31. <https://doi.org/10.1186/S13059-018-1416-2>
- Duan, J. E., Jiang, Z. C., Alqahtani, F., Mandoiu, I., Dong, H., Zheng, X., Marjani, S. L., Chen, J., & Tian, X. C. (2019). Methylome dynamics of bovine gametes and in vivo early embryos. *Frontiers in Genetics*, 10, 512. <https://doi.org/10.3389/fgene.2019.00512>
- Duncan, F. E., Schindler, K., Schultz, R. M., Blengini, C. S., Stein, P., Stricker, S. A., Wessel, G. M., & Williams, C. J. (2020). Unscrambling the oocyte and the egg: clarifying terminology of the female gamete in mammals. *Molecular Human Reproduction*, 26(11), 797–800. <https://doi.org/10.1093/molehr/gaaa066>
- Dyer, S. C., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Barrera-Enriquez, V. P., Becker, A., Bennett, R., Beracochea, M., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., ... Yates, A. D. (2025). Ensembl 2025. *Nucleic Acids Research*, 53(D1), D948–D957. <https://doi.org/10.1093/NAR/GKAE1071>
- Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., & Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes & Development*, 33(3–4), 194–208. <https://doi.org/10.1101/GAD.321174.118>
- Edwards, R. G. (1965). MATURATION IN VITRO OF HUMAN OVARIAN OOCYTES. *The Lancet*, 286(7419), 926–929. [https://doi.org/10.1016/S0140-6736\(65\)92903-X](https://doi.org/10.1016/S0140-6736(65)92903-X)
- Edwards, R. G., Bavister, B. D., & Steptoe, P. C. (1969). Early Stages of Fertilization in vitro of Human Oocytes Matured in vitro. *Nature*, 221(5181), 632–635. <https://doi.org/10.1038/221632a0>
- Efrat, S., & Kaempfer, R. (1984). Control of biologically active interleukin 2 messenger RNA formation in induced human lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 81(9), 2601–2605. <https://doi.org/10.1073/PNAS.81.9.2601>
- Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Muzny, D. M., Weinstock, G. M., Adelson, D. L., Eichler, E. E., Elnitski, L., Guigó, R., Hamernik, D. L., Kappes, S. M., Lewin, H. A., Lynn, D. J., Nicholas, F. W., Raymond, A., Rijnkels, M., Skow, L. C., Zdobnov, E. M., ... Zhao, F. Q. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science*, 324(5926), 522–528. <https://doi.org/10.1126/SCIENCE.1169588>
- Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819), 154–156. <https://doi.org/10.1038/292154a0>

- Ezer, S., Yoshihara, M., Katayama, S., Daub, C., Lohi, H., Krjutskov, K., & Kere, J. (2021). Generation of RNA sequencing libraries for transcriptome analysis of globin-rich tissues of the domestic dog. *STAR Protocols*, 2(4), 100995. <https://doi.org/10.1016/J.XPRO.2021.100995>
- Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S. M., Tapscott, S. J., & Bradley, R. K. (2015). A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *ELife*, 4, e04996. <https://doi.org/10.7554/ELIFE.04996>
- Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J. Z., Ne Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., & Andrau, J.-C. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Research*, 22, 2399–2408. <https://doi.org/10.1101/gr.138776.112>
- Frei, R. E., Schultz, G. A., & Church, R. B. (1989). Qualitative and quantitative changes in protein synthesis occur at the 8–16-cell stage of embryogenesis in the cow. *Reproduction*, 86(2), 637–641. <https://doi.org/10.1530/JRF.0.0860637>
- Fu, G. K., Hu, J., Wang, P. H., & Fodor, S. P. A. (2011). Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), 9026–9031. <https://doi.org/10.1073/PNAS.1017621108>
- Gabriëls, J., Beckers, M. C., Ding, H., De Vriese, A., Plaisance, S., Van Der Maarel, S. M., Padberg, G. W., Frants, R. R., Hewitt, J. E., Collen, D., & Belayew, A. (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene*, 236(1), 25–32. [https://doi.org/10.1016/S0378-1119\(99\)00267-X](https://doi.org/10.1016/S0378-1119(99)00267-X)
- Gal-mark, N., Schwartz, S., & Ast, G. (2008). Alternative splicing of Alu exons – two arms are better than one. *Nucleic Acids Research*, 36(6), 2012–2023. <https://doi.org/10.1093/NAR/GKN024>
- Galán, A., Montaner, D., Póo, M. E., Valbuena, D., Ruiz, V., Aguilar, C., Dopazo, J., & Simón, C. (2010). Functional Genomics of 5- to 8-Cell Stage Human Embryos by Blastomere Single-Cell cDNA Analysis. *PLOS ONE*, 5(10), e13615. <https://doi.org/10.1371/JOURNAL.PONE.0013615>
- Galliot, B., De Vargas, C., & Miller, D. (1999). Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. *Development Genes and Evolution*, 209, 186–197. <https://doi.org/10.1007/s004270050243>
- Gao, F., Niu, Y., Sun, Y. E., Lu, H., Chen, Y., Li, S., Kang, Y., Luo, Y., Si, C., Yu, J., Li, C., Sun, N., Si, W., Wang, H., Ji, W., & Tan, T. (2017). De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Research*, 27(4), 526–539. <https://doi.org/10.1038/CR.2017.25>
- Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., Tian, Y., Liu, J., Chen, Z. J., & Liu, J. (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell*, 173(1), 248–259.e15. <https://doi.org/10.1016/J.CELL.2018.02.028>
- Gardner, D. K., & Wale, P. L. (2013). Analysis of metabolism to select viable human embryos for transfer. *Fertility and Sterility*, 99(4), 1062–1072. <https://doi.org/10.1016/J.FERTNSTERT.2012.12.004>
- Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro, A., Sciaiini, Marco, Scherer, & Cédric. (2024). *viridis(Lite)* – Colorblind-Friendly Color Maps for R. <https://doi.org/10.5281/zenodo.4679423>

- Gawriyski, L., Jouhilahti, E. M., Yoshihara, M., Fei, L., Weltner, J., Airenne, T. T., Trokovic, R., Bhagat, S., Tervaniemi, M. H., Murakawa, Y., Salokas, K., Liu, X., Miettinen, S., Bürglin, T. R., Sahu, B., Otonkoski, T., Johnson, M. S., Katayama, S., Varjosalo, M., & Kere, J. (2023). Comprehensive characterization of the embryonic factor LEUTX. *IScience*, 26(3), 106172. <https://doi.org/10.1016/j.isci.2023.106172>
- Gawriyski, L., Tan, Z., Liu, X., Chowdhury, I., Malaymar Pinar, D., Zhang, Q., Weltner, J., Jouhilahti, E. M., Wei, G. H., Kere, J., & Varjosalo, M. (2024). Interaction network of human early embryonic transcription factors. *EMBO Reports*, 25(3), 1589–1622. <https://doi.org/10.1038/S44319-024-00074-0>
- Gehring, W. J., Affolter, M., & Bürglin, T. (1994). Homeodomain proteins. *Annual Review of Biochemistry*, 63, 487–526. <https://doi.org/10.1146/annurev.bi.63.070194.002415>
- Geng, L. N., Yao, Z., Snider, L., Fong, A. P., Cech, J. N., Young, J. M., vanderMaarel, S. M., Ruzzo, W. L., Gentleman, R. C., Tawil, R., & Tapscott, S. J. (2012). DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Developmental Cell*, 22(1), 38–51. <https://doi.org/10.1016/J.DEVCEL.2011.11.013>
- Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., Gill, C. A., Green, R. D., Hamernik, D. L., Kappes, S. M., Lien, S., Matukumalli, L. K., McEwan, J. C., Nazareth, L. V., Schnabel, R. D., Weinstock, G. M., Wheeler, D. A., Ajmone-Marsan, P., Barends, W., Boettcher, P. J., ... Dodds, K. G. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926), 528–532. <https://doi.org/10.1126/SCIENCE.1167936>
- Gilbert, I., Scantland, S., Sylvestre, E. L., Gravel, C., Laflamme, I., Sirard, M. A., & Robert, C. (2009). The dynamics of gene products fluctuation during bovine pre-hatching development. *Molecular Reproduction and Development*, 76(8), 762–772. <https://doi.org/10.1002/mrd.21030>
- Göke, J., Lu, X., Chan, Y. S., Ng, H. H., Ly, L. H., Sachs, F., & Szczerbinska, I. (2015). Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. *Cell Stem Cell*, 16(2), 135–141. <https://doi.org/10.1016/J.STEM.2015.01.005>
- Graf, A., Krebs, S., Zakhartchenko, V., Schwalb, B., Blum, H., & Wolf, E. (2014). Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), 4139–4144. <https://doi.org/10.1073/pnas.1321569111>
- Grant, C. E., & Bailey, T. L. (2021). XSTREME: Comprehensive motif analysis of biological sequence datasets. *BioRxiv*, 2021.09.02.458722. <https://doi.org/10.1101/2021.09.02.458722>
- Grigoriadis, D., Perdikopanis, N., Georgakilas, G. K., & Hatzigeorgiou, A. G. (2022). DeepTSS: multi-branch convolutional neural network for transcription start site identification from CAGE data. *BMC Bioinformatics*, 23(Suppl 2), 395. <https://doi.org/10.1186/S12859-022-04945-Y>
- Gross, N., Strillacci, M. G., Peñagaricano, F., & Khatib, H. (2019). Characterization and functional roles of paternal RNAs in 2–4 cell bovine embryos. *Scientific Reports*, 9(1), 20347. <https://doi.org/10.1038/s41598-019-55868-3>
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., Martin, L., Ware, C. B., Blish, C. A., Chang, H. Y., Pera, R. A. R., & Wysocka, J. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*, 522(7555), 221–225. <https://doi.org/10.1038/nature14308>

- Grow, E. J., Weaver, B. D., Smith, C. M., Guo, J., Stein, P., Shadle, S. C., Hendrickson, P. G., Johnson, N. E., Butterfield, R. J., Menafrá, R., Kloet, S. L., van der Maarel, S. M., Williams, C. J., & Cairns, B. R. (2021). p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nature Genetics*, 53, 1207–1220. <https://doi.org/10.1038/s41588-021-00893-0>
- Guan, T., Guo, J., Lin, R., Liu, J., Luo, R., Zhang, Z., Pei, D., & Liu, J. (2024). Single-cell analysis of preimplantation embryonic development in guinea pigs. *BMC Genomics*, 25(1), 911. <https://doi.org/10.1186/s12864-024-10815-z>
- Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., Plath, K., & Smith, A. (2017). Epigenetic resetting of human pluripotency. *Development*, 144(15), 2748–2763. <https://doi.org/10.1242/DEV.146811>
- Guo, M., Zhang, Y., Zhou, J., Bi, Y., Xu, J., Xu, C., Kou, X., Zhao, Y., Li, Y., Tu, Z., Liu, K., Lin, J., Yang, P., Gao, S., & Wang, Y. (2019). Precise temporal regulation of Dux is important for embryo development. *Cell Research*, 29(11), 956–959. <https://doi.org/10.1038/s41422-019-0238-4>
- Guo, Q., Xu, F., Song, S., Kong, S., Zhai, F., Xiu, Y., Liu, D., Li, M., Lian, Y., Ding, L., Liu, Q., Yang, M., Du, Z., Wang, N., Long, C., Wang, X., Wang, Y., Yan, Z., Qiao, J., ... Yuan, P. (2025). Allelic transcriptomic profiling identifies the role of PRD-like homeobox genes in human embryonic-cleavage-stage arrest. *Developmental Cell*, 60(9), 1290–1303.e6. <https://doi.org/10.1016/j.devcel.2024.12.031>
- Guo, Y., Kitano, T., Inoue, K., Murano, K., Hirose, M., Li, T. D., Sakashita, A., Ishizu, H., Ogonuki, N., Matoba, S., Sato, M., Ogura, A., & Siomi, H. (2024). Obox4 promotes zygotic genome activation upon loss of Dux. *ELife*, 13, e95856. <https://doi.org/10.7554/ELIFE.95856>
- Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. In *Methods in Molecular Biology* (Vol. 1418, pp. 335–351). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3578-9_16
- Halstead, M. M., Ma, X., Zhou, C., Schultz, R. M., & Ross, P. J. (2020). Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *Nature Communications*, 11(1), 4654. <https://doi.org/10.1038/s41467-020-18508-3>
- Hamatani, T., Carter, M. G., Sharov, A. A., & Ko, M. S. H. (2004). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Developmental Cell*, 6(1), 117–131. [https://doi.org/10.1016/S1534-5807\(03\)00373-3](https://doi.org/10.1016/S1534-5807(03)00373-3)
- Hamatani, T., Ko, M. S., Yamada, M., Kuji, N., Mizusawa, Y., Shoji, M., Hada, T., Asada, H., Maruyama, T., & Yoshimura, Y. (2006). Global gene expression profiling of preimplantation embryos. *Human Cell*, 19(3), 98–117. <https://doi.org/10.1111/J.1749-0774.2006.00018.X>
- Hamm, D. C., Paatela, E. M., Bennett, S. R., Wong, C. J., Campbell, A. E., Wladyka, C. L., Smith, A. A., Jagannathan, S., Hsieh, A. C., & Tapscott, S. J. (2023). The transcription factor DUX4 orchestrates translational reprogramming by broadly suppressing translation efficiency and promoting expression of DUX4-induced mRNAs. *PLOS Biology*, 21(9), e3002317. <https://doi.org/10.1371/JOURNAL.PBIO.3002317>
- Hanes, S. D., & Brent, R. (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell*, 57(7), 1275–1283. [https://doi.org/10.1016/0092-8674\(89\)90063-9](https://doi.org/10.1016/0092-8674(89)90063-9)

- Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports*, 2(3), 666–673. <https://doi.org/10.1016/j.CELREP.2012.08.003>
- Hendrickson, P. G., Doráis, J. A., Grow, E. J., Whiddon, J. L., Lim, J. W., Wike, C. L., Weaver, B. D., Pflueger, C., Emery, B. R., Wilcox, A. L., Nix, D. A., Peterson, C. M., Tapscott, S. J., Carrell, D. T., & Cairns, B. R. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERV1 retrotransposons. *Nature Genetics*, 49(6), 925–934. <https://doi.org/10.1038/ng.3844>
- Herquel, B., Ouararhni, K., Khetchoumian, K., Ignat, M., Teletin, M., Mark, M., Béchade, G., Van Dorsselaer, A., Sanglier-Cianfèrani, S., Hamiche, A., Cammas, F., Davidson, I., & Losson, R. (2011). Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), 8212–8217. <https://doi.org/10.1073/PNAS.1101544108>
- Hewitt, J. E., Lyle, R., Clark, L. N., Valleley, E. M., Wright, T. J., Wijmenga, C., Van Deutekom, J. C. t., Francis, F., Sharpe, P. T., Hofker, M., Frants, R. R., & Williamson, R. (1994). Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Human Molecular Genetics*, 3(8), 1287–1295. <https://doi.org/10.1093/hmg/3.8.1287>
- Holland, P. W. H. (2013). Evolution of homeobox genes. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(1), 31–45. <https://doi.org/10.1002/WDEV.78>
- Holland, P. W. H., Booth, H. A. F., & Bruford, E. A. (2007). Classification and nomenclature of all human homeobox genes. *BMC Biology*, 5, 47. <https://doi.org/10.1186/1741-7007-5-47>
- Hosseini, S. M., Dufort, I., Caballero, J., Moulavi, F., Ghanaei, H. R., & Sirard, M. A. (2015). Transcriptome profiling of bovine inner cell mass and trophectoderm derived from in vivo generated blastocysts. *BMC Developmental Biology*, 15(1), 49. <https://doi.org/10.1186/S12861-015-0096-3>
- Howards, S. S. (1997). Antoine van Leeuwenhoek and the discovery of sperm. *Fertility and Sterility*, 67(1), 16–17. [https://doi.org/10.1016/S0015-0282\(97\)81848-1](https://doi.org/10.1016/S0015-0282(97)81848-1)
- Hu, B., Jin, H., Shi, Y., Yu, H., Wu, X., Wang, S., & Zhang, K. (2024). Single-cell RNA-Seq reveals the earliest lineage specification and X chromosome dosage compensation in bovine preimplantation embryos. *The FASEB Journal*, 38(4), e23492. <https://doi.org/10.1096/FJ.202302035RR>
- Hu, Z., Tan, D. E. K., Chia, G., Tan, H., Leong, H. F., Chen, B. J., Lau, M. S., Tan, K. Y. S., Bi, X., Yang, D., Ho, Y. S., Wu, B., Bao, S., Wong, E. S. M., & Tee, W. W. (2020). Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nature Cell Biology*, 22(2), 175–186. <https://doi.org/10.1038/s41556-019-0453-8>
- Huang, K., Maruyama, T., & Fan, G. (2014). The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. *Cell Stem Cell*, 15(4), 410–415. <https://doi.org/10.1016/j.STEM.2014.09.014>
- Huang, L., Lou, C. H., Chan, W., Shum, E. Y., Shao, A., Stone, E., Karam, R., Song, H. W., & Wilkinson, M. F. (2011). RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. *Molecular Cell*, 43(6), 950–961. <https://doi.org/10.1016/j.MOLCEL.2011.06.031>
- Huang, W., & Khatib, H. (2010). Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. *BMC Genomics*, 11(1), 711. <https://doi.org/10.1186/1471-2164-11-711>

- Huntriss, J., Hinkins, M., & Picton, H. M. (2006). cDNA cloning and expression of the human NOBOX gene in oocytes and ovarian follicles. *Molecular Human Reproduction*, 12(5), 283–289. <https://doi.org/10.1093/MOLEHR/GAL035>
- Hyttel, P., Viuff, D., Avery, B., Laurincik, J., & Greve, T. (1996). Transcription and cell cycle-dependent development of intranuclear bodies and granules in two-cell bovine embryos. *Reproduction*, 108(2), 263–270. <https://doi.org/10.1530/JRF.0.1080263>
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, 21(7), 1160–1167. <https://doi.org/10.1101/gr.110882.110>
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P., & Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nature Protocols*, 7(5), 813–828. <https://doi.org/10.1038/nprot.2012.022>
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., & Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2), 163–166. <https://doi.org/10.1038/nmeth.2772>
- Jagannathan, S., Ogata, Y., Gafken, P. R., Tapscott, S. J., & Bradley, R. K. (2019). Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *ELife*, 8, e41740. <https://doi.org/10.7554/ELIFE.41740>
- Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., & Amit, I. (2014). Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*, 343(6172), 776–779. <https://doi.org/10.1126/SCIENCE.1247651>
- Jang, J., Jung, J., Lee, Y. H., Lee, S., Baik, M., & Kim, H. (2023). Chromosome-level genome assembly of Korean native cattle and pangenome graph of 14 *Bos taurus* assemblies. *Scientific Data*, 10(1), 560. <https://doi.org/10.1038/S41597-023-02453-Z>
- Ji, S., Chen, F., Stein, P., Wang, J., Zhou, Z., Wang, L., Zhao, Q., Lin, Z., Liu, B., Xu, K., Lai, F., Xiong, Z., Hu, X., Kong, T., Kong, F., Huang, B., Wang, Q., Xu, Q., Fan, Q., ... Xie, W. (2023). OBOX regulates mouse zygotic genome activation and early development. *Nature*, 620(7976), 1047–1053. <https://doi.org/10.1038/s41586-023-06428-3>
- Jiang, Z., Sun, J., Dong, H., Luo, O., Zheng, X., Obergfell, C., Tang, Y., Bi, J., O'Neill, R., Ruan, Y., Chen, J., & Tian, X. (Cindy). (2014). Transcriptional profiles of bovine in vivo pre-implantation development. *BMC Genomics*, 15(1), 756. <https://doi.org/10.1186/1471-2164-15-756>
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36(Issue suppl_2), W5–W9. <https://doi.org/10.1093/NAR/GKN201>
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J. M., Vincentelli, R., Luscombe, N. M., Hughes, T. R., Lemaire, P., Ukkonen, E., Kivioja, T., & Taipale, J. (2013). DNA-binding specificities of human transcription factors. *Cell*, 152(1–2), 327–339. <https://doi.org/10.1016/j.cell.2012.12.009>
- Jouhilahti, E.-M., Madissoon, E., Vesterlund, L., Tökönen, V., Krjutškov, K., Plaza Reyes, A., Petropoulos, S., Månsson, R., Linnarsson, S., Bürglin, T., Lanner, F., Hovatta, O., Katayama, S., & Kere, J. (2016). The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development*, 143(19), 3459–3469. <https://doi.org/10.1242/DEV.134510>

- Jukam, D., Shariati, S. A. M., & Skotheim, J. M. (2017). Zygotic Genome Activation in Vertebrates. *Developmental Cell*, 42(4), 316–332. <https://doi.org/10.1016/j.devcel.2017.07.026>
- Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics*, 10(1), 19–31. <https://doi.org/10.1038/nrg2487>
- Kalbfleisch, T. S., McKay, S. D., Murdoch, B. M., Adelson, D. L., Almansa-Villa, D., Becker, G., Beckett, L. M., Benítez-Galeano, M. J., Biase, F., Casey, T., Chuong, E., Clark, E., Clarke, S., Cockett, N., Couldrey, C., Davis, B. W., Elsik, C. G., Faraut, T., Gao, Y., ... Rosen, B. D. (2024). The Ruminant Telomere-to-Telomere (RT2T) Consortium. *Nature Genetics*, 56(8), 1566–1573. <https://doi.org/10.1038/S41588-024-01835-2>
- Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*, 53(D1), D672–D677. <https://doi.org/10.1093/NAR/GKAE909>
- Karolchik, D., Hinricks, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, 32(Issue suppl_1), D493–D496. <https://doi.org/10.1093/NAR/GKH103>
- Kaskar, K., Cochran, R., Hamilton, D. P., David, A., Henkel, R., Gibbons, W., & Blesson, C. S. (2019). Effect of ultra-low oxygen (2%) environment on mouse embryo morphokinetics and blastocyst development. *Fertility and Sterility*, 112(3), e270–e271. <https://doi.org/10.1016/J.FERTNSTERT.2019.07.802>
- Katayama, S., Ranga, V., Jouhilahti, E.-M., Airenne, T. T., Johnson, M. S., Mukherjee, K., Bürglin, T. R., & Kere, J. (2018). Phylogenetic and mutational analyses of human LEUTX, a homeobox gene implicated in embryogenesis. *Scientific Reports*, 8(1), 17421. <https://doi.org/10.1038/s41598-018-35547-5>
- Kawamura-Saito, M., Yamazaki, Y., Kaneko, K., Kawaguchi, N., Kanda, H., Mukai, H., Gotoh, T., Motoi, T., Fukayama, M., Aburatani, H., Takizawa, T., & Nakamura, T. (2006). Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Human Molecular Genetics*, 15(13), 2125–2137. <https://doi.org/10.1093/HMG/DDL136>
- Kent, W. J. (2002). BLAT – the BLAST-like alignment tool. *Genome Research*, 12(4), 656–664. <https://doi.org/10.1101/gr.229202>
- Kepkova, K. V., Vodicka, P., Toralova, T., Lopatarova, M., Cech, S., Dolezel, R., Havlicek, V., Besenfelder, U., Kuzmany, A., Sirard, M. A., Laurincik, J., & Kanka, J. (2011). Transcriptomic analysis of in vivo and in vitro produced bovine embryos revealed a developmental change in cullin 1 expression during maternal-to-embryonic transition. *Theriogenology*, 75(9), 1582–1595. <https://doi.org/10.1016/j.theriogenology.2010.12.019>
- Khan, D. R., Dubé, D., Gall, L., Peynot, N., Ruffini, S., Laffont, L., Le Bourhis, D., Degrelle, S., Jouneau, A., & Duranthon, V. (2012). Expression of Pluripotency Master Regulators during Two Key Developmental Transitions: EGA and Early Lineage Specification in the Bovine Embryo. *PLOS ONE*, 7(3), e34110. <https://doi.org/10.1371/JOURNAL.PONE.0034110>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, 9(1), 72–74. <https://doi.org/10.1038/nmeth.1778>
- Kojima, M. L., Hoppe, C., & Giraldez, A. J. (2024). The maternal-to-zygotic transition: reprogramming of the cytoplasm and nucleus. *Nature Reviews Genetics*, 26, 245–267. <https://doi.org/10.1038/s41576-024-00792-0>
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4), 610–620. <https://doi.org/10.1016/j.molcel.2015.04.005>
- Kong, Q., Yang, X., Zhang, H., Liu, S., Zhao, J., Zhang, J., Weng, X., Jin, J., & Liu, Z. (2020). Lineage specification and pluripotency revealed by transcriptome analysis from oocyte to blastocyst in pig. *FASEB Journal*, 34(1), 691–705. <https://doi.org/10.1096/fj.201901818RR>
- Kowaljew, V., Marcowycz, A., Anseau, E., Conde, C. B., Sauvage, S., Mattéotti, C., Arias, C., Corona, E. D., Nuñez, N. G., Leo, O., Wattiez, R., Figlewicz, D., Laoudj-Chenivesse, D., Belayew, A., Coppée, F., & Rosa, A. L. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscular Disorders*, 17(8), 611–623. <https://doi.org/10.1016/J.NMD.2007.04.002>
- Kravchenko, P., & Tachibana, K. (2024). Rise and SINE: roles of transcription factors and retrotransposons in zygotic genome activation. *Nature Reviews Molecular Cell Biology*, 26, 68–79. <https://doi.org/10.1038/s41580-024-00772-6>
- Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J., & Schmitz, J. (2007). Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends in Genetics*, 23(4), 158–161. <https://doi.org/10.1016/j.tig.2007.02.002>
- Krjutškov, K., Katayama, S., Saare, M., Vera-Rodriguez, M., Lubenets, D., Samuel, K., Laisk-Podar, T., Teder, H., Einarsdottir, E., Salumets, A., & Kere, J. (2016). Single-cell transcriptome analysis of endometrial tissue. *Human Reproduction*, 31(4), 844–853. <https://doi.org/10.1093/HUMREP/DEW008>
- Kuijk, E. W., Du Puy, L., Van Tol, H. T. A., Oei, C. H. Y., Haagsman, H. P., Colenbrander, B., & Roelen, B. A. J. (2008). Differences in early lineage segregation between mammals. *Developmental Dynamics*, 237(4), 918–927. <https://doi.org/10.1002/DVDY.21480>
- Labuda, D., Sinnett, D., Richer, C., Deragon, J. M., & Striker, G. (1991). Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. *Journal of Molecular Evolution*, 32(5), 405–414. <https://doi.org/10.1007/BF02101280>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lanzuolo, C., & Orlando, V. (2012). Memories from the Polycomb Group Proteins. *Annual Review Genetics*, 46, 561–89. <https://doi.org/10.1146/annurev-genet-110711-155603>
- Lavagi, I., Krebs, S., Simmet, K., Beck, A., Zakhartchenko, V., Wolf, E., & Blum, H. (2018). Single-cell RNA sequencing reveals developmental heterogeneity of blastomeres during major genome activation in bovine embryos. *Scientific Reports*, 8(1), 4071. <https://doi.org/10.1038/s41598-018-22248-2>
- Leidenroth, A., Clapp, J., Mitchell, L. M., Coneyworth, D., Dearden, F. L., Iannuzzi, L., & Hewitt, J. E. (2012). Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma*, 121(5), 489–497. <https://doi.org/10.1007/S00412-012-0380-Y>

- Leidenroth, A., & Hewitt, J. E. (2010). A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evolutionary Biology*, 10(1), 364. <https://doi.org/10.1186/1471-2148-10-364>
- Lemmers, R. J. L. F., de Kievit, P., Sandkuijl, L., Padberg, G. W., van Ommen, G. J. B., Frants, R. R., & van der Maarel, S. M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nature Genetics*, 32(2), 235–236. <https://doi.org/10.1038/ng999>
- Lemmers, R. J. L. F., Van Der Vliet, P. J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J. G., Snider, L., Straasheijm, K. R., Van Ommen, G. J., Padberg, G. W., Miller, D. G., Tapscott, S. J., Tawil, R., Frants, R. R., & Van Der Maarel, S. M. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*, 329(5999), 1650–1653. <https://doi.org/10.1126/science.1189044>
- Leng, L., Sun, J., Huang, J., Gong, F., Yang, L., Zhang, S., Yuan, X., Fang, F., Xu, X., Luo, Y., Bolund, L., Peters, B. A., Lu, G., Jiang, T., Xu, F., & Lin, G. (2019). Single-Cell Transcriptome Analysis of Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation Development. *Cell Stem Cell*, 25(5), 697–712.e6. <https://doi.org/10.1016/j.stem.2019.09.004>
- Lepikhov, K., Zakhartchenko, V., Hao, R., Yang, F., Wrenzycki, C., Niemann, H., Wolf, E., & Walter, J. (2008). Evidence for conserved DNA and histone H3 methylation reprogramming in mouse, bovine and rabbit zygotes. *Epigenetics & Chromatin*, 1(1), 8. <https://doi.org/10.1186/1756-8935-1-8>
- Lewin, H. A. (2009). It's a bull's market. *Science*, 324(5926), 478–479. <https://doi.org/10.1126/SCIENCE.1173880>
- Lewin, T. D., Blagrove, J. R., & Holland, P. W. H. (2023). Rapid Evolution of the Embryonically Expressed Homeobox Gene LEUTX within Primates. *Genome Biology and Evolution*, 15(6), evad097. <https://doi.org/10.1093/GBE/EVAD097>
- Lewin, T. D., Fouladi-Nashta, A. A., & Holland, P. W. H. (2022). PRD-Class Homeobox Genes in Bovine Early Embryos: Function, Evolution, and Overlapping Roles. *Molecular Biology and Evolution*, 39(5), msac098. <https://doi.org/10.1093/MOLBEV/MSAC098>
- Lewin, T. D., Royall, A. H., & Holland, P. W. H. (2021). Dynamic Molecular Evolution of Mammalian Homeobox Genes: Duplication, Loss, Divergence and Gene Conversion Sculpt PRD Class Repertoires. *Journal of Molecular Evolution*, 89(6), 396–414. <https://doi.org/10.1007/S00239-021-10012-6>
- Li, G., & Holland, P. W. (2010). The origin and evolution of ARGFX homeobox loci in mammalian radiation. *BMC Evolutionary Biology*, 10(1), 182. <https://doi.org/10.1186/1471-2148-10-182>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/BIOINFORMATICS/BTY191>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/BIOINFORMATICS/BTP352>
- Li, L., Guo, F., Gao, Y., Ren, Y., Yuan, P., Yan, L., Li, R., Lian, Y., Li, J., Hu, B., Gao, J., Wen, L., Tang, F., & Qiao, J. (2018). Single-cell multi-omics sequencing of human early embryos. *Nature Cell Biology*, 20(7), 847–858. <https://doi.org/10.1038/s41556-018-0123-2>

- Li, R., & Albertini, D. F. (2013). The road to maturation: Somatic cell interaction and self-organization of the mammalian oocyte. *Nature Reviews Molecular Cell Biology*, 14(3), 141–152. <https://doi.org/10.1038/nrm3531>
- Li, T. T., Xia, T., Wu, J. Q., Hong, H., Sun, Z. L., Wang, M., Ding, F. R., Wang, J., Jiang, S., Li, J., Pan, J., Yang, G., Feng, J. N., Dai, Y. P., Zhang, X. M., Zhou, T., & Li, T. (2023). De novo genome assembly depicts the immune genomic characteristics of cattle. *Nature Communications*, 14(1), 6601. <https://doi.org/10.1038/s41467-023-42161-1>
- Li, Y., Sun, J., Ling, Y., Ming, H., Chen, Z., Fang, F., Liu, Y., Cao, H., Ding, J., Cao, Z., Zhang, X., Bondioli, K., Jiang, Z., & Zhang, Y. (2020). Transcription profiles of oocytes during maturation and embryos during preimplantation development in vivo in the goat. *Reproduction Fertility and Development*, 32(7), 714–725. <https://doi.org/10.1071/RD19391>
- Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., & Gao, X. (2023). Repetitive DNA sequence detection and its role in the human genome. *Communications Biology*, 6(1), 954. <https://doi.org/10.1038/S42003-023-05322-Y>
- Liu, B., Xu, Q., Wang, Q., Feng, S., Lai, F., Wang, P., Zheng, F., Xiang, Y., Wu, J., Nie, J., Qiu, C., Xia, W., Li, L., Yu, G., Lin, Z., Xu, K., Xiong, Z., Kong, F., Liu, L., ... Xie, W. (2020). The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature*, 587(7832), 139–144. <https://doi.org/10.1038/s41586-020-2847-y>
- Liu, L., Leng, L., Liu, C., Lu, C., Yuan, Y., Wu, L., Gong, F., Zhang, S., Wei, X., Wang, M., Zhao, L., Hu, L., Wang, J., Yang, H., Zhu, S., Chen, F., Lu, G., Shang, Z., & Lin, G. (2019). An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nature Communications*, 10(1), 364. <https://doi.org/10.1038/s41467-018-08244-0>
- Liu, Y., Zhao, H., Shao, F., Zhang, Y., Nie, H., Zhang, J., Li, C., Hou, Z., Chen, Z. J., Wang, J., Zhou, B., Wu, K., & Lu, F. (2023). Remodeling of maternal mRNA through poly(A) tail orchestrates human oocyte-to-embryo transition. *Nature Structural & Molecular Biology*, 30(2), 200–215. <https://doi.org/10.1038/s41594-022-00908-2>
- Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91(7), 2757–2761. <https://doi.org/10.1073/PNAS.91.7.2757>
- Lonergan, P., Gutiérrez-Adán, A., Rizos, D., Pintado, B., La Fuente, J. de, & Boland, M. P. (2003). Relative messenger RNA abundance in bovine oocytes collected in vitro or in vivo before and 20 hr after the preovulatory luteinizing hormone surge. *Molecular Reproduction and Development*, 66(3), 297–305. <https://doi.org/10.1002/MRD.10357>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. <https://doi.org/10.1186/S13059-014-0550-8>
- Lu, X., Zhang, Y., Wang, L., Wang, L., Wang, H., Xu, Q., Xiang, Y., Chen, C., Kong, F., Xia, W., Lin, Z., Ma, S., Liu, L., Wang, X., Ni, H., Li, W., Guo, Y., & Xie, W. (2021). Evolutionary epigenomic analyses in mammalian early embryos reveal species-specific innovations and conserved principles of imprinting. *Science Advances*, 7(48), eabi6178. <https://doi.org/10.1126/SCIADV.ABI6178>
- Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology*, 16(11), 665–677. <https://doi.org/10.1038/NRM4063>

- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., & Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, *487*(7405), 57–63. <https://doi.org/10.1038/nature11244>
- Machug, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P., & Bradley, D. G. (1997). Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (*Bos taurus* and *Bos indicus*). *Genetics*, *146*(3), 1071–1086. <https://academic.oup.com/genetics/article/146/3/1071/6053945>
- Madisson, E., Damdimopoulos, A., Katayama, S., Krjutškov, K., Einarsdottir, E., Mamia, K., De Groef, B., Hovatta, O., Kere, J., & Damdimopoulou, P. (2019). Pleomorphic Adenoma Gene 1 Is Needed For Timely Zygotic Genome Activation and Early Embryo Development. *Scientific Reports*, *9*(1), 8411. <https://doi.org/10.1038/s41598-019-44882-0>
- Madisson, E., Jouhilahti, E.-M., Vesterlund, L., Tökönen, V., Krjutškov, K., Petropoulos, S., Einarsdottir, E., Linnarsson, S., Lanner, F., Månsson, R., Hovatta, O., Bürglin, T. R., Katayama, S., & Kere, J. (2016). Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Scientific Reports*, *6*(1), 28995. <https://doi.org/10.1038/srep28995>
- Maeso, I., Dunwell, T. L., Wyatt, C. D. R., Marlétaz, F., Veto, B., Bernal, J. A., Quah, S., Irimia, M., & Holland, P. W. H. (2016). Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. *BMC Biology*, *14*(1), 45. <https://doi.org/10.1186/s12915-016-0267-0>
- Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T., & Turner, J. M. A. (2020). A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature*, *586*(7830), 612–617. <https://doi.org/10.1038/s41586-020-2629-6>
- Mao, C. A. N., Gan, L., & Klein, W. H. (1994). Multiple Otx Binding Sites Required for Expression of the Strongylocentrotus purpuratus Spec2a Gene. *Developmental Biology*, *165*(1), 229–242. <https://doi.org/10.1006/DBIO.1994.1249>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. <https://doi.org/10.14806/EJ.17.1.200>
- Mazid, M. A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., Liu, H., Fu, L., Yang, Y., Ibañez, D. P., Lai, J., Wei, X., An, J., Guo, P., Yuan, Y., ... Esteban, M. A. (2022). Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature*, *605*(7909), 315–324. <https://doi.org/10.1038/s41586-022-04625-0>
- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., & Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature*, *308*(5958), 428–433. <https://doi.org/10.1038/308428a0>
- McMahon, S. B. (2014). MYC and the Control of Apoptosis. *Cold Spring Harbor Perspectives in Medicine*, *4*(7), a014407. <https://doi.org/10.1101/CSHPERSPECT.A014407>
- Memili, E., Dominko, T., & First, N. L. (1998). Onset of Transcription in Bovine Oocytes and Preimplantation Embryos. *Genetics, Gene Regulation, and Expression*, *51*, 36–42. [https://doi.org/10.1002/\(SICI\)1098-2795\(199809\)51:1<36::AID-MRD4>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1098-2795(199809)51:1<36::AID-MRD4>3.0.CO;2-X)
- Memili, E., & First, N. L. (1999). Control of Gene Expression at the Onset of Bovine Embryonic Development. *Biology of Reproduction*, *61*(5), 1198–1207. <https://doi.org/10.1095/BIOLREPROD61.5.1198>

- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/MOLBEV/MSAA015>
- Misirlioglu, M., Page, G. P., Sagirkaya, H., Kaya, A., Parrish, J. J., First, N. L., & Memili, E. (2006). Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50), 18905–18910. <https://doi.org/10.1073/PNAS.0608247103>
- Mitsuhashi, H., Mitsuhashi, S., Lynn-jones, T., Kawahara, G., & Kunkel, L. M. (2013). Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Human Molecular Genetics*, 22(3), 568–577. <https://doi.org/10.1093/HMG/DDS467>
- Morinaga, T., Yasuda, H., Hashimoto, T., Higashio, K., & Tamaoki, T. (1991). A Human α -Fetoprotein Enhancer-Binding Protein, ATBF1, Contains Four Homeodomains and Seventeen Zinc Fingers. *Molecular and Cellular Biology*, 11(12), 6041–6049. <https://doi.org/10.1128/MCB.11.12.6041-6049.1991>
- Moya-Jódar, M., Ullate-Agote, A., Barlabé, P., Rodríguez-Madoz, J. R., Abizanda, G., Barreda, C., Carvajal-Vergara, X., Vilas-Zornoza, A., Romero, J. P., Garate, L., Agirre, X., Coppiello, G., Prósper, F., & Aranguren, X. L. (2023). Revealing cell populations catching the early stages of human embryo development in naive pluripotent stem cell cultures. *Stem Cell Reports*, 18(1), 64–80. <https://doi.org/10.1016/J.STEMCR.2022.11.015>
- Murphy, P. J., Wu, S. F., James, C. R., Wike, C. L., & Cairns, B. R. (2018). Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. *Cell*, 172(5), 993–1006.e13. <https://doi.org/10.1016/J.CELL.2018.01.022>
- Negrón-Pérez, V. M., Zhang, Y., & Hansen, P. J. (2017). Single-cell gene expression of the bovine blastocyst. *Reproduction*, 154(5), 627–644. <https://doi.org/10.1530/REP-17-0345>
- Nichols, J., & Smith, A. (2009). Naive and Primed Pluripotent States. *Cell Stem Cell*, 4(6), 487–492. <https://doi.org/10.1016/J.STEM.2009.05.015>
- Nip, Y., Bennett, S. R., Smith, A. A., Jones, T. I., Jones, P. L., & Tapscott, S. J. (2023). Human DUX4 and porcine DUXC activate similar early embryonic programs in pig muscle cells: implications for preclinical models of FSHD. *Human Molecular Genetics*, 32(11), 1864–1874. <https://doi.org/10.1093/HMG/DDAD021>
- Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H., & Nakatani, Y. (1996). The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. *Cell*, 87(5), 953–959. [https://doi.org/10.1016/S0092-8674\(00\)82001-2](https://doi.org/10.1016/S0092-8674(00)82001-2)
- Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., & Arima, T. (2014). Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLOS Genetics*, 10(12), e1004868. <https://doi.org/10.1371/JOURNAL.PGEN.1004868>
- Olbrich, T., Vega-Sendino, M., Tillo, D., Wu, W., Zolnerowich, N., Pavani, R., Tran, A. D., Domingo, C. N., Franco, M., Markiewicz-Potoczny, M., Pegoraro, G., FitzGerald, P. C., Kruhlak, M. J., Lazzerini-Denchi, E., Nora, E. P., Nussenzweig, A., & Ruiz, S. (2021). CTCF is a barrier for 2C-like reprogramming. *Nature Communications*, 12(1), 4856. <https://doi.org/10.1038/s41467-021-25072-x>
- Ong, C. T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics*, 15(4), 234–246. <https://doi.org/10.1038/nrg3663>

- Oomen, M. E., Rodriguez-Terrones, D., Kurome, M., Zakhartchenko, V., Mottes, L., Simmet, K., Noll, C., Nakatani, T., Mourra-Diaz, C. M., Aksoy, I., Savatier, P., Göke, J., Wolf, E., Kaessmann, H., & Torres-Padilla, M.-E. (2025). An atlas of transcription initiation reveals regulatory principles of gene and transposable element expression in early mammalian development. *Cell*, 188(4), 1156–1174.e20. <https://doi.org/10.1016/J.CELL.2024.12.013>
- Org, T., Hensen, K., Kreevan, R., Mark, E., Sarv, O., Andreson, R., Jaakma, Ü., Salumets, A., & Kurg, A. (2019). Genome-wide histone modification profiling of inner cell mass and trophectoderm of bovine blastocysts by RAT-ChIP. *PLOS ONE*, 14(11), e0225801. <https://doi.org/10.1371/JOURNAL.PONE.0225801>
- Palermo, G., Joris, H., Devroey, P., & Van Steirteghem, A. C. (1992). Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *The Lancet*, 340(8810), 17–18. [https://doi.org/10.1016/0140-6736\(92\)92425-F](https://doi.org/10.1016/0140-6736(92)92425-F)
- Paloviita, P., Nykänen, S., Harjuhaahto, S., Grym, H., Santaniemi, R., Tyynismaa, H., Torregrosa-Muñumer, R., & Vuoristo, S. (2026). Integrated transcriptomic analysis reveals metabolic remodeling and gene expression networks related to human 8-cell-stage embryo-like cells. *Cell Reports*, 45(1), 116748. <https://doi.org/10.1016/J.CELREP.2025.116748>
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/BIOINFORMATICS/BTY633>
- Paulson, E. E., Fishman, E. L., Schultz, R. M., & Ross, P. J. (2022). Embryonic microRNAs are essential for bovine preimplantation embryo development. *Proceedings of the National Academy of Sciences of the United States of America*, 119(45), e2212942119. <https://doi.org/10.1073/pnas.2212942119>
- Pennetier, S., Uzbekova, S., Perreau, C., Papillier, P., Mermillod, P., & Dalbiès-Tran, R. (2004). Spatio-Temporal Expression of the Germ Cell Marker Genes MATER, ZAR1, GDF9, BMP15, and VASA in Adult Bovine Tissues, Oocytes, and Preimplantation Embryos. *Biology of Reproduction*, 71(4), 1359–1366. <https://doi.org/10.1095/BIOLREPROD.104.030288>
- Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., & Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*, 174(2), 391–405.e19. <https://doi.org/10.1016/J.CELL.2018.05.043>
- Perez, G., Barber, G. P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J. N., Hinrichs, A. S., Lee, C. M., Nassar, L. R., Raney, B. J., Speir, M. L., van Baren, M. J., Vaske, C. J., Haussler, D., Kent, W. J., & Haussler, M. (2025). The UCSC Genome Browser database: 2025 update. *Nucleic Acids Research*, 53(D1), D1243–D1249. <https://doi.org/10.1093/NAR/GKAE974>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295. <https://doi.org/10.1038/nbt.3122>
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., & Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*, 165(4), 1012–1026. <https://doi.org/10.1016/j.cell.2016.03.023>
- Picelli, S. (2017). Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biology*, 14(5), 637–650. <https://doi.org/10.1080/15476286.2016.1201618>

- Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, 10(11), 1096–1098. <https://doi.org/10.1038/NMETH.2639>
- Pineda, P. S., MacPhillamy, C., Ren, Y., Chen, T., Zhong, L., Adelson, D. L., Dessaix, C., Perez-Silva, J., Haggerty, L., Martin, F. J., Bottema, C. D. K., Pitchford, W. S., Rosen, B. D., Smith, T. P. L., & Low, W. Y. (2025). Insights into natural neocentromere evolution from a cattle T2T X chromosome. *Nature Communications*, 16(1), 10745. <https://doi.org/10.1038/s41467-025-65778-w>
- Piskacek, S., Gregor, M., Nemethova, M., Grabner, M., Kovarik, P., & Piskacek, M. (2007). Nine-amino-acid transactivation domain: Establishment and prediction utilities. *Genomics*, 89(6), 756–768. <https://doi.org/10.1016/J.YGENO.2007.02.003>
- Plante, L., L. Shepherd, D., Allan King, W., & Plante, C. (1994). Cleavage and 3H-uridine incorporation in bovine embryos of high in vitro developmental potential. *Molecular Reproduction and Development*, 39(4), 375–383. <https://doi.org/10.1002/MRD.1080390405>
- Porubsky, D., Guitart, X., Yoo, D. A., Dishuck, P. C., Harvey, W. T., & Eichler, E. E. (2025). SVbyEye: a visual tool to characterize structural variation among whole-genome assemblies. *Bioinformatics*, 41(6), btaf332. <https://doi.org/10.1093/BIOINFORMATICS/BTAF332>
- Puscheck, E. E., Awonuga, A. O., Yang, Y., Jiang, Z., & Rappolee, D. A. (2015). Molecular Biology of the Stress Response in the Early Embryo and its Stem Cells. *Advances in Experimental Medicine and Biology*, 843, 77–128. https://doi.org/10.1007/978-1-4939-2480-6_4
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/BIOINFORMATICS/BTQ033>
- Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., & Wright, P. E. (1997). Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. *Cell*, 91(6), 741–752. [https://doi.org/10.1016/S0092-8674\(00\)80463-8](https://doi.org/10.1016/S0092-8674(00)80463-8)
- Ramsköld, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebukova, I., Loring, J. F., Laurent, L. C., Schroth, G. P., & Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, 30(8), 777–782. <https://doi.org/10.1038/nbt.2282>
- Resnick, R., Wong, C. J., Hamm, D. C., Bennett, S. R., Skene, P. J., Hake, S. B., Henikoff, S., van der Maarel, S. M., & Tapscott, S. J. (2019). DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Reports*, 29(7), 1812–1820.e5. <https://doi.org/10.1016/j.celrep.2019.10.025>
- Reyes, J. M., Chitwood, J. L., & Ross, P. J. (2015). RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. *Molecular Reproduction and Development*, 82(2), 103–114. <https://doi.org/10.1002/MRD.22445>
- Rickard, A. M., Petek, L. M., & Miller, D. G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Human Molecular Genetics*, 24(20), 5901–5914. <https://doi.org/10.1093/HMG/DDV315>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/NBT.1754>

- Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms*, 1839(12), 1362–1372.
<https://doi.org/10.1016/j.bbagrm.2014.02.007>
- Royall, A. H., Maeso, I., Dunwell, T. L., & Holland, P. W. H. (2018). Mouse Obox and Crxos modulate preimplantation transcriptional profiles revealing similarity between paralogous mouse and human homeobox genes. *EvoDevo*, 9(1), 2.
<https://doi.org/10.1186/S13227-018-0091-4>
- Ruthenburg, A. J., Allis, C. D., & Wysocka, J. (2007). Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell*, 25(1), 15–30. <https://doi.org/10.1016/J.MOLCEL.2006.12.014>
- Ryan, J. F., Burton, P. M., Mazza, M. E., Kwong, G. K., Mullikin, J. C., & Finnerty, J. R. (2006). The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: Evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biology*, 7(7), R64.
<https://doi.org/10.1186/gb-2006-7-7-r64>
- Sakashita, A., Kitano, T., Ishizu, H., Guo, Y., Masuda, H., Ariura, M., Murano, K., & Siomi, H. (2023). Transcription of MERV1 retrotransposons is required for pre-implantation embryo development. *Nature Genetics*, 55(3), 484–495.
<https://doi.org/10.1038/s41588-023-01324-y>
- Sakata, N., Hemmi, K., Kawaguchi, M., Miura, Y., Noguchi, S., Ma, D., Sasahara, M., Kato, T., Hori, M., & Tamaoki, T. (2000). The Mouse ZFH-4 Protein Contains Four Homeodomains and Twenty-Two Zinc Fingers. *Biochemical and Biophysical Research Communications*, 273(2), 686–693.
<https://doi.org/10.1006/BBRC.2000.2990>
- Samorinha, C., Pereira, M., Machado, H., Figueiredo, B., & Silva, S. (2014). Factors associated with the donation and non-donation of embryos for research: a systematic review. *Human Reproduction Update*, 20(5), 641–655.
<https://doi.org/10.1093/humupd/dmu026>
- Sasagawa, Y., Danno, H., Takada, H., Ebisawa, M., Tanaka, K., Hayashi, T., Kurisaki, A., & Nikaido, I. (2018). Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biology*, 19(1), 29. <https://doi.org/10.1186/S13059-018-1407-3>
- Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., & Ueda, H. R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology*, 14(4), 3097. <https://doi.org/10.1186/GB-2013-14-4-R31>
- Schettini, G. P., Morozzyuk, M., & Biase, F. H. (2024). Identification of novel cattle (*Bos taurus*) genes and biological insights of their function in pre-implantation embryo development. *BMC Genomics*, 25(1), 775.
<https://doi.org/10.1186/S12864-024-10685-5>
- Schultz, R. M., Stein, P., & Svoboda, P. (2018). The oocyte-to-embryo transition in mouse: past, present, and future. *Biology of Reproduction*, 99(1), 160–174.
<https://doi.org/10.1093/biolre/iox013>
- Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., McKay, D. J., & Harrison, M. M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Research*, 25(11), 1715–1726.
<https://doi.org/10.1101/GR.192682.115>

- Schwartz, Y. B., & Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nature Reviews Genetics*, 8(1), 9–22. <https://doi.org/10.1038/nrg1981>
- Sendai, Y., Itoh, T., Yamashita, S., & Hoshi, H. (2001). Molecular Cloning of a cDNA Encoding a Bovine Growth Differentiation Factor-9 (GDF-9) and Expression of GDF-9 in Bovine Ovarian Oocytes and In Vitro-Produced Embryos. *Cloning*, 3(1), 3–10. <https://doi.org/10.1089/152045501300189286>
- Seo, S., Larkin, D. M., & Loores, J. J. (2013). Cattle genomics and its implications for future nutritional strategies for dairy cattle. *Animal*, 7(s1), 172–183. <https://doi.org/10.1017/S1751731111002588>
- Sha, Q. Q., Zheng, W., Wu, Y. W., Li, S., Guo, L., Zhang, S., Lin, G., Ou, X. H., & Fan, H. Y. (2020). Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. *Nature Communications*, 11(1), 4917. <https://doi.org/10.1038/s41467-020-18680-6>
- Shadle, S. C., Bennett, S. R., Wong, C. J., Karreman, N. A., Campbell, A. E., van der Maarel, S. M., Bass, B. L., & Tapscott, S. J. (2019). DUX4-induced bidirectional HSATII satellite repeat transcripts form intranuclear double-stranded RNA foci in human cell models of FSHD. *Human Molecular Genetics*, 28(23), 3997–4011. <https://doi.org/10.1093/HMG/DDZ242>
- Shadle, S. C., Zhong, J. W., Campbell, A. E., Conerly, M. L., Jagannathan, S., Wong, C. J., Morello, T. D., van der Maarel, S. M., & Tapscott, S. J. (2017). DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLOS Genetics*, 13(3), e1006658. <https://doi.org/10.1371/JOURNAL.PGEN.1006658>
- Shen, H., Yang, M., Li, S., Zhang, J., Peng, B., Wang, C., Chang, Z., Ong, J., & Du, P. (2021). Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell*, 184(11), 2843–2859.e20. <https://doi.org/10.1016/J.CELL.2021.04.020>
- Shen, W. K., Chen, S. Y., Gan, Z. Q., Zhang, Y. Z., Yue, T., Chen, M. M., Xue, Y., Hu, H., & Guo, A. Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Research*, 51(D1), D39–D45. <https://doi.org/10.1093/NAR/GKAC907>
- Shi, Z., Yan, Y., Zhu, R., Zhu, X., Hu, K., Yue, Y., Xu, W., Xuan, M., Gan, X., Yang, Z., Zhang, Y., & Cao, Z. (2025). DUXA regulates the early development of porcine cloned embryos. *Theriogenology*, 247, 117563. <https://doi.org/10.1016/J.THERIOGENOLOGY.2025.117563>
- Smit, A. F. A. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Research*, 21(8), 1863–1872. <https://doi.org/10.1093/NAR/21.8.1863>
- Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., & Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, 511(7511), 611–615. <https://doi.org/10.1038/NATURE13581>
- Snider, L., Asawachaicharn, A., Tyler, A. E., Geng, L. N., Petek, L. M., Maves, L., Miller, D. G., Lemmers, R. J. L. F., Winokur, S. T., Tawil, R., van der Maarel, S. M., Filippova, G. N., & Tapscott, S. J. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Human Molecular Genetics*, 18(13), 2414–2430. <https://doi.org/10.1093/HMG/DDP180>

- Snider, L., Geng, L. N., Lemmers, R. J. L. F., Kyba, M., Ware, C. B., Nelson, A. M., Tawil, R., Filippova, G. N., van der Maarel, S. M., Tapscott, S. J., & Miller, D. G. (2010). Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. *PLOS Genetics*, 6(10), e1001181. <https://doi.org/10.1371/JOURNAL.PGEN.1001181>
- Sokka, J., Yoshihara, M., Kvist, J., Laiho, L., Warren, A., Stadelmann, C., Jouhilahti, E. M., Kilpinen, H., Balboa, D., Katayama, S., Kyttälä, A., Kere, J., Otonkoski, T., Weltner, J., & Trokovic, R. (2022). CRISPR activation enables high-fidelity reprogramming into human pluripotent stem cells. *Stem Cell Reports*, 17(2), 413–426. <https://doi.org/10.1016/j.stemcr.2021.12.017>
- Sokkary, N., & Dietrich, J. E. (2018). Ovarian Embryology, Anatomy, and Physiology Including Normal Menstrual Physiology. In *Endocrine Surgery in Children* (pp. 319–326). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-54256-9_23
- Spens, A. E., Sutliff, N. A., Bennett, S. R., Campbell, A. E., & Tapscott, S. J. (2023). Human DUX4 and mouse Dux interact with STAT1 and broadly inhibit interferon-stimulated gene induction. *ELife*, 12, e82057. <https://doi.org/10.7554/ELIFE.82057>
- Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378), 490–495. <https://doi.org/10.1038/NATURE10716>
- Steenwyk, J. L., & Rokas, A. (2021). ggpubfigs: Colorblind-Friendly Color Palettes and ggplot2 Graphic System Extensions for Publication-Quality Scientific Figures. *Microbiology Resource Announcements*, 10(44), e0087121. <https://doi.org/10.1128/MRA.00871-21>
- Steptoe, P. C., & Edwards, R. G. (1978). Birth after the reimplantation of a human embryo. *Lancet*, 312(8085), 366. [https://doi.org/10.1016/s0140-6736\(78\)92957-4](https://doi.org/10.1016/s0140-6736(78)92957-4)
- Steptoe, P. C., Edwards, R. G., & Purdy, J. M. (1971). Human Blastocysts grown in Culture. *Nature*, 229(5280), 132–133. <https://doi.org/10.1038/229132a0>
- Storm, M. P., Kumpfmüller, B., Bone, H. K., Buchholz, M., Sanchez Ripoll, Y., Chaudhuri, J. B., Niwa, H., Tosh, D., & Welham, M. J. (2014). Zscan4 Is Regulated by PI3-Kinase and DNA-Damaging Agents and Directly Interacts with the Transcriptional Repressors LSD1 and CtBP2 in Mouse Embryonic Stem Cells. *PLOS ONE*, 9(3), e89821. <https://doi.org/10.1371/JOURNAL.PONE.0089821>
- Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, 8(1), 90. <https://doi.org/10.1186/1742-4690-8-90>
- Sun, Y., Nien, C. Y., Chen, K., Liu, H. Y., Johnston, J., Zeitlinger, J., & Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Research*, 25(11), 1703–1714. <https://doi.org/10.1101/GR.192542.115>
- Surani, M. A., Hayashi, K., & Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. *Cell*, 128(4), 747–762. <https://doi.org/10.1016/J.CELL.2007.02.010>
- Suzumori, N., Yan, C., Matzuk, M. M., & Rajkovic, A. (2002). Nobox is a homeobox-encoding gene preferentially expressed in primordial and growing oocytes. *Mechanisms of Development*, 111(1–2), 137–141. [https://doi.org/10.1016/S0925-4773\(01\)00620-7](https://doi.org/10.1016/S0925-4773(01)00620-7)

- Svarcova, O., Laurincik, J., Avery, B., Mlyncek, M., Niemann, H., & Maddox-Hyttel, P. (2007). Nucleolar development and allocation of key nucleolar proteins require de novo transcription in bovine embryos. *Molecular Reproduction and Development*, 74(11), 1428–1435. <https://doi.org/10.1002/MRD.20727>
- Svoboda, P. (2018). Mammalian zygotic genome activation. *Seminars in Cell and Developmental Biology*, 84, 118–126. <https://doi.org/10.1016/j.semcdb.2017.12.006>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., Lao, K., & Surani, M. A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols*, 5(3), 516–535. <https://doi.org/10.1038/nprot.2009.236>
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5), 377–382. <https://doi.org/10.1038/nmeth.1315>
- Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., & Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell*, 29(3), 449–459.e6. <https://doi.org/10.1016/j.stem.2022.01.014>
- Theunissen, T. W., Friedli, M., He, Y., Planet, E., O’Neil, R. C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., Duc, J., Cohen, M. A., Wert, K. J., Castanon, R., Zhang, Z., Huang, Y., Nery, J. R., Drotar, J., Lungjangwa, T., ... Jaenisch, R. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell*, 19(4), 502–515. <https://doi.org/10.1016/J.STEM.2016.06.011>
- Theunissen, T. W., Powell, B. E., Wang, H., Mitalipova, M., Faddah, D. A., Reddy, J., Fan, Z. P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y., Rangarajan, S., D’Alessio, A., Zhang, J., Gao, Q., Dawlaty, M. M., Young, R. A., ... Jaenisch, R. (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell*, 15(4), 471–487. <https://doi.org/10.1016/J.STEM.2014.07.002>
- Thiery, M. (2009). Reinier de Graaf (1641–1673) and the Graafian follicle. *Gynecological Surgery*, 6(2), 189–191. <https://doi.org/10.1007/S10397-009-0466-6>
- Thompson, B., Davidson, E. A., Liu, W., Nebert, D. W., Bruford, E. A., Zhao, H., Dermitzakis, E. T., Thompson, D. C., & Vasiliou, V. (2020). Overview of PAX gene family: analysis of human tissue-specific variant expression and involvement in human disease. *Human Genetics*, 140(3), 381–400. <https://doi.org/10.1007/S00439-020-02212-9>
- Thompson, P. J., Macfarlan, T. S., & Lorincz, M. C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Molecular Cell*, 62(5), 766–776. <https://doi.org/10.1016/j.molcel.2016.03.029>
- Thomson, J. A. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391), 1145–1147. <https://doi.org/10.1126/SCIENCE.282.5391.1145>
- Tökönen, V., Katayama, S., Vesterlund, L., Jouhilahti, E. M., Sheikhi, M., Madissoon, E., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Bürglin, T. R., Linnarsson, S., Hovatta, O., & Kere, J. (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature Communications*, 6, 8207. <https://doi.org/10.1038/ncomms9207>

- Töhönen, V., Katayama, S., Vesterlund, L., Sheikhi, M., Antonsson, L., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Linnarsson, S., Hovatta, O., & Kere, J. (2017). Transcription activation of early human development suggests DUX4 as an embryonic regulator. *BioRxiv*, 123208. <https://doi.org/10.1101/123208>
- Tripurani, S. K., Lee, K.-B., Wang, L., Wee, G., Smith, G. W., Lee, Y. S., Latham, K. E., & Yao, J. (2011). A Novel Functional Role for the Oocyte-Specific Transcription Factor Newborn Ovary Homeobox (NOBOX) during Early Embryonic Development in Cattle. *Endocrinology*, 152(3), 1013–1023. <https://doi.org/10.1210/EN.2010-1134>
- Turgut, A. Ç., & Kutlubay, R. (2023). Soranus of Ephesus (98–138 AD) and his contributions on human embryology, paediatrics and obstetric knowledge. *Child's Nervous System*, 40(4), 969–973. <https://doi.org/10.1007/S00381-023-06215-5>
- Uehara, T., & Yanagimachi, R. (1976). Microsurgical Injection of Spermatozoa into Hamster Eggs with Subsequent Transformation of Sperm Nuclei into Male Pronuclei. *Biology of Reproduction*, 15(4), 467–470. <https://doi.org/10.1095/BIOLREPROD15.4.467>
- University of Tartu. (2018). *UT Rocket*. share.neic.no. <https://doi.org/10.23673/PH6N-0144>
- Vallée, M., Dufort, I., Desrosiers, S., Labbe, A., Gravel, C., Gilbert, I., Robert, C., & Sirard, M. A. (2009). Revealing the bovine embryo transcript profiles during early in vivo embryonic development. *Reproduction*, 138(1), 95–105. <https://doi.org/10.1530/REP-08-0533>
- Van Geel, M., Dickson, M. C., Beck, A. F., Bolland, D. J., Frants, R. R., Van der Maarel, S. M., De Jong, P. J., & Hewitt, J. E. (2002). Genomic Analysis of Human Chromosome 10q and 4q Telomeres Suggests a Common Origin. *Genomics*, 79(2), 210–217. <https://doi.org/10.1006/GENO.2002.6690>
- Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., & Belmonte, J. C. I. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development*, 138(17), 3699–3709. <https://doi.org/10.1242/dev.064741>
- Vega-Sendino, M., Lüttmann, F. F., Olbrich, T., Chen, Y., Kuenne, C., Stein, P., Tillo, D., Carey, G. I., Zhong, J., Savy, V., Radonova, L., Lu, T., Saykali, B., Kim, K. P., Domingo, C. N., Schüler, L., Günther, S., Bentsen, M., Bosnakovski, D., ... Ruiz, S. (2024). The homeobox transcription factor DUXBL controls exit from totipotency. *Nature Genetics*, 56(4), 697–709. <https://doi.org/10.1038/s41588-024-01692-z>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Delser, P. M., Hare, A. J., Burger, J., Collins, M. J., Kehati, R., Hesse, P., Fulton, D., Sauer, E. W., Mohaseb, F. A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., ... Bradley, D. G. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*, 365(6449), 173–176. <https://doi.org/10.1126/SCIENCE.AAV1002>
- Viuff, D., Avery, B., Greve, T., King, W. A., & Hyttel, P. (1996). Transcriptional Activity in In Vitro Produced Bovine Two- and Four-Cell Embryos. *Molecular Reproduction and Development*, 43(2), 171–179. [https://doi.org/10.1002/\(SICI\)1098-2795\(199602\)43:2<171::AID-MRD6>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2795(199602)43:2<171::AID-MRD6>3.0.CO;2-O)
- Viuff, D., Hyttel, P., Avery, B., Vajta, G., Greve, T., Callesen, H., & Thomsen, P. D. (1998). Ribosomal Ribonucleic Acid Is Transcribed at the 4-Cell Stage in In Vitro Produced Bovine Embryos. *Biology of Reproduction*, 59(3), 626–631. <https://doi.org/10.1095/BIOLREPROD59.3.626>

- Vuoristo, S., Bhagat, S., Hydén-Granskog, C., Yoshihara, M., Gawrylski, L., Jouhilahti, E. M., Ranga, V., Tamirat, M., Huhtala, M., Kirjanov, I., Nykänen, S., Krjutškov, K., Damdimopoulos, A., Weltner, J., Hashimoto, K., Recher, G., Ezer, S., Paluoja, P., Paloviita, P., ... Kere, J. (2022). DUX4 is a multifunctional factor priming human embryonic genome activation. *IScience*, 25(4), 104137. <https://doi.org/10.1016/J.ISCI.2022.104137>
- Wallace, L. M., Garwick, S. E., Mei, W., Belayew, A., Coppee, F., Ladner, K. J., Guttridge, D., Yang, J., & Harper, S. Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Annals of Neurology*, 69(3), 540–552. <https://doi.org/10.1002/ANA.22275>
- Wang, Q., Zhang, C., Dang, Y., Sun, J., Zou, Z., Li, C., Ma, S., Li, Z., Liu, H., Ma, X., Yang, Z., Wang, L., Wu, K., Chen, Z. J., Xie, W., & Zhao, H. (2025). Maternal factor OTX2 regulates human embryonic genome activation and early development. *Nature Genetics*, 57, 2772–2784. <https://doi.org/10.1038/s41588-025-02350-8>
- Wang, T., Peng, J., Fan, J., Tang, N., Hua, R., Zhou, X., Wang, Z., Wang, L., Bai, Y., Quan, X., Wang, Z., Zhang, L., Luo, C., Zhang, W., Kang, X., Liu, J., Li, L., & Li, L. (2024). Single-cell multi-omics profiling of human preimplantation embryos identifies cytoskeletal defects during embryonic arrest. *Nature Cell Biology*, 26(2), 263–277. <https://doi.org/10.1038/s41556-023-01328-0>
- Wei, Q., Zhong, L., Zhang, S., Mu, H., Xiang, J., Yue, L., Dai, Y., & Han, J. (2017). Bovine lineage specification revealed by single-cell gene expression analysis from zygote to blastocyst. *Biology of Reproduction*, 97(1), 5–17. <https://doi.org/10.1093/biolre/iox071>
- Weltner, J., Balboa, D., Katayama, S., Bernal, M., Krjutškov, K., Jouhilahti, E.-M., Trokovic, R., Kere, J., & Otonkoski, T. (2018). Human pluripotent reprogramming with CRISPR activators. *Nature Communications*, 9(1), 2643. <https://doi.org/10.1038/s41467-018-05067-x>
- Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W., & Tapscott, S. J. (2017). Conservation and innovation in the DUX4-family gene network. *Nature Genetics*, 49(6), 935–940. <https://doi.org/10.1038/ng.3846>
- Whiddon, J. M., Yao, J. L., Kasinathan, Z., & Snider, B. (2013). DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis. *PLoS Genet*, 9(11), e1003947. <https://doi.org/10.1371/journal.pgen.1003947>
- Wiedemann, S. M., Mildner, S. N., Bönisch, C., Israel, L., Mäyser, A., Matheisl, S., Straub, T., Merkl, R., Leonhardt, H., Kremmer, E., Schermelleh, L., & Hake, S. B. (2010). Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *Journal of Cell Biology*, 190(5), 777–791. <https://doi.org/10.1083/JCB.201002043>
- Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., Gruter, A. M., Hofker, M. H., Moerer, P., Williamson, R., van Ommen, G. J. B., Padberg, G. W., & Frants, R. R. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nature Genetics*, 2(1), 26–30. <https://doi.org/10.1038/ng0992-26>
- Wilming, L. G., Boychenko, V., & Harrow, J. L. (2015). Comprehensive comparative homeobox gene annotation in human and mouse. *Database*, 2015, bav091. <https://doi.org/10.1093/DATABASE/BAV091>
- Wong, C. C., Loewke, K. E., Bossert, N. L., Behr, B., De Jonge, C. J., Baer, T. M., & Pera, R. A. R. (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnology*, 28(10), 1115–1121. <https://doi.org/10.1038/nbt.1686>

- Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., Li, W., Li, Y., Ma, J., Peng, X., Zheng, H., Ming, J., Zhang, W., Zhang, J., Tian, G., ... Xie, W. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*, 534(7609), 652–657.
<https://doi.org/10.1038/nature18606>
- Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S., Zhang, N., Duan, F., Ming, J., Zhang, X., Niu, W., Song, W., Jin, H., Guo, Y., Dai, S., ... Sun, Y. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature*, 557(7704), 256–260.
<https://doi.org/10.1038/s41586-018-0080-8>
- Wu, L., Su, P., Li, D., Zhou, K., Song, W., Shi, J., Wang, Q., Huang, Q., Hu, S., Zhang, X., Liu, X., Zhou, J., & Miao, Y.-L. (2025). LEUTX and OTX2 orchestrate totipotency-to-pluripotency transitions and heterogeneity in pEPSCs. *Cell Reports*, 44(10), 116385. <https://doi.org/10.1016/j.celrep.2025.116385>
- Wu, S. L., Tsai, M. S., Wong, S. H., Hsieh-Li, H. M., Tsai, T. S., Chang, W. T., Huang, S. L., Chiu, C. C., & Wang, S. H. (2010). Characterization of genomic structures and expression profiles of three tandem repeats of a mouse double homeobox gene: Duxbl. *Developmental Dynamics*, 239(3), 927–940. <https://doi.org/10.1002/DVDY.22210>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3), 100141.
<https://doi.org/10.1016/J.XINN.2021.100141>
- Xia, W., Xu, J., Yu, G., Yao, G., Xu, K., Ma, X., Zhang, N., Liu, B., Li, T., Lin, Z., Chen, X., Li, L., Wang, Q., Shi, D., Shi, S., Zhang, Y., Song, W., Jin, H., Hu, L., ... Sun, Y. P. (2019). Resetting histone modifications during human parental-to-zygotic transition. *Science*, 365(6451), 353–360. <https://doi.org/10.1126/SCIENCE.AAW5118>
- Xiang, Y., Qian, Y., Li, Z., Wang, J., Tian, R., Meng, W., Bu, J., Huang, F., Ai, Z., Wu, D., Chen, X., Wu, Y., Shen, L., Chan, Y.-S., Gao, Y., Ma, J., Liu, W., Gao, S., Zhang, D., & Liang, H. (2026). Endogenous retroviruses synthesize heterologous chimeric RNAs to reinforce human early embryo development. *Science*, 391(6783), eadv5257. <https://doi.org/10.1126/science.adv5257>
- Xie, S. Q., Leeke, B. J., Whilding, C., Wagner, R. T., Garcia-Llagostera, F., Low, Y. X., Chammas, P., Cheung, N. T. F., Dormann, D., McManus, M. T., & Percharde, M. (2022). Nucleolar-based Dux repression is essential for embryonic two-cell stage exit. *Genes and Development*, 36(5–6), 331–347.
<https://doi.org/10.1101/GAD.349172.121>
- Xu, J., Shu, Y., Yao, G., Zhang, Y., Niu, W., Zhang, Y., Ma, X., Jin, H., Zhang, F., Shi, S., Wang, Y., Song, W., Dai, S., Cheng, L., Zhang, X., Xie, W., Hsueh, A. J., & Sun, Y. (2021). Parental methylome reprogramming in human uniparental blastocysts reveals germline memory transition. *Genome Research*, 31(9), 1519.
<https://doi.org/10.1101/GR.273318.120/-/DC1>
- Xu, Q., & Xie, W. (2018). Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. *Trends in Cell Biology*, 28(3), 237–253.
<https://doi.org/10.1016/J.TCB.2017.10.008>
- Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., Xie, W., Wu, T., Xie, L., & Yu, G. (2024). Using clusterProfiler to characterize multiomics data. *Nature Protocols*, 19(11), 3292–3320.
<https://doi.org/10.1038/S41596-024-01020-Z>

- Xu, W., Rould, M. A., Jun, S., Desplan, C., & Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*, 80(4), 639–650. [https://doi.org/10.1016/0092-8674\(95\)90518-9](https://doi.org/10.1016/0092-8674(95)90518-9)
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J. Y., Horvath, S., & Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, 500(7464), 593–597. <https://doi.org/10.1038/nature12364>
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., Huang, J., Li, M., Wu, X., Wen, L., Lao, K., Li, R., Qiao, J., & Tang, F. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural and Molecular Biology*, 20(9), 1131–1139. <https://doi.org/10.1038/nsmb.2660>
- Yandlm, C., & Karakulah, G. (2019). Expression dynamics of repetitive DNA in early human embryonic development. *BMC Genomics*, 20(1), 439. <https://doi.org/10.1186/S12864-019-5803-1>
- Yang, L., Di, A., Song, L., Liu, X., Wu, D., Wang, S., Hao, Z., Bu, L., Bai, C., Su, G., Wei, Z., Zhang, L., Liu, Z., Gao, S., & Li, G. (2025). Generation of modified cows and sheep from spermatid-like haploid embryonic stem cells. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-025-02832-4>
- Yang, Y., Xu, Y., Ding, C., Khoudja, R. Y., Lin, M., Awonuga, A. O., Dai, J., Puscheck, E. E., Rappolee, D. A., & Zhou, C. (2016). Comparison of 2, 5, and 20 % O₂ on the development of post-thaw human embryos. *Journal of Assisted Reproduction and Genetics*, 33(7), 919–927. <https://doi.org/10.1007/S10815-016-0693-5>
- Yaşar, B., Boskovic, N., Ivask, M., Weltner, J., Jouhilahti, E. M., Vill, P., Skoog, T., Jaakma, Ü., Kere, J., Bürglin, T. R., Katayama, S., Org, T., & Kurg, A. (2024). Molecular cloning of PRD-like homeobox genes expressed in bovine oocytes and early IVF embryos. *BMC Genomics*, 25(1), 1048. <https://doi.org/10.1186/S12864-024-10969-W>
- Yaşar, B., Boskovic, N., Org, T., Kere, J., Kurg, A., & Katayama, S. (2025). RNA sequencing of mRNA 5'-ends reveals regulators of bovine embryonic genome activation. *BMC Genomics*, 26(1), 910. <https://doi.org/10.1186/S12864-025-12110-X>
- Yaşar, B., Org, T., Ivask, M., Gunaydin, G. Y., Boskovic, N., Jaakma, Ü., Kere, J., Kurg, A., & Katayama, S. (2026). Structure and genomic organization of the human DUX4 homologue bovine DUXC. *BioRxiv*, 2026.03.05.709741. <https://doi.org/10.64898/2026.03.05.709741>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1), 134. <https://doi.org/10.1186/1471-2105-13-134>
- Yoshihara, M., & Kere, J. (2023). Transcriptomic differences between human 8-cell-like cells reprogrammed with different methods. *Stem Cell Reports*, 18(8), 1621–1628. <https://doi.org/10.1016/j.stemcr.2023.06.009>
- Yoshihara, M., Kirjanov, I., Nykänen, S., Sokka, J., Weltner, J., Lundin, K., Gawriyski, L., Jouhilahti, E. M., Varjosalo, M., Tervaniemi, M. H., Otonkoski, T., Trokovic, R., Katayama, S., Vuoristo, S., & Kere, J. (2022). Transient DUX4 expression in human embryonic stem cells induces blastomere-like expression program that is marked by SLC34A2. *Stem Cell Reports*, 17(7), 1743–1756. <https://doi.org/10.1016/j.stemcr.2022.06.002>
- Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*, 69(1), e96. <https://doi.org/10.1002/CPBI.96>

- Yu, X., Liang, S., Chen, M., Yu, H., Li, R., Qu, Y., Kong, X., Guo, R., Zheng, R., Izsvák, Z., Sun, C., Yang, M., & Wang, J. (2022). Recapitulating early human development with 8C-like cells. *Cell Reports*, 39(12), 110994. <https://doi.org/10.1016/j.CELREP.2022.110994>
- Yuan, S., Zhan, J., Zhang, J., Liu, Z., Hou, Z., Zhang, C., Yi, L., Gao, L., Zhao, H., Chen, Z. J., Liu, J., & Wu, K. (2023). Human zygotic genome activation is initiated from paternal genome. *Cell Discovery*, 9(1), 13. <https://doi.org/10.1038/s41421-022-00494-z>
- Zalzman, M., Falco, G., Sharova, L. V., Nishiyama, A., Thomas, M., Lee, S. L., Stagg, C. A., Hoang, H. G., Yang, H. Te, Indig, F. E., Wersto, R. P., & Ko, M. S. H. (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature*, 464(7290), 858–863. <https://doi.org/10.1038/nature08882>
- Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., Zhang, Y., Wang, Q., Li, Y., Meng, A., & Xie, W. (2018). Widespread Enhancer Dememorization and Promoter Priming during Parental-to-Zygotic Transition. *Molecular Cell*, 72(4), 673–686.e6. <https://doi.org/10.1016/j.MOLCEL.2018.10.017>
- Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., Liu, W., Kou, X., Zhao, Y., He, W., Li, C., Chen, B., Li, Y., Wang, Q., Ma, J., ... Xie, W. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*, 537(7621), 553–557. <https://doi.org/10.1038/nature19361>
- Zhang, H., Ji, S., Zhang, K., Chen, Y., Ming, J., Kong, F., Wang, L., Wang, S., Zou, Z., Xiong, Z., Xu, K., Lin, Z., Huang, B., Liu, L., Fan, Q., Jin, S., Deng, H., & Xie, W. (2023). Stable maternal proteins underlie distinct transcriptome, translatome, and proteome reprogramming during mouse oocyte-to-embryo transition. *Genome Biology*, 24(1), 166. <https://doi.org/10.1186/S13059-023-02997-8>
- Zhang, J., Li, H., Li, L., Wu, J., Song, L., Liu, X., Pan, Z., Zhou, C., Li, W., Liu, Z., Jiao, M., Hu, M., Dong, Z., Zhang, H., Shi, B., Wang, Y., Wang, D., Carter, B., Zhao, S., ... Zhang, Y. (2025). Super RNA Pol II domains enhance minor ZGA through 3D interaction to ensure the integrity of major transcriptional waves in late-ZGA mammals. *Cell Genomics*, 5(6), 100856. <https://doi.org/10.1016/j.XGEN.2025.100856>
- Zhang, P., Zucchelli, M., Bruce, S., Hambiliki, F., Stavreus-Evers, A., Levkov, L., Skottman, H., Kerkelä, E., Kere, J., & Hovatta, O. (2009). Transcriptome Profiling of Human Pre-Implantation Development. *PLOS ONE*, 4(11), e7844. <https://doi.org/10.1371/JOURNAL.PONE.0007844>
- Zhang, X., Zhang, C., Zhou, D., Zhang, T., Chen, X., Ren, J., He, C., Meng, F., Zhou, Q., Yang, Q., Dai, C., Lin, G., Zeng, S., & Leng, L. (2023). Telomeres cooperate in zygotic genome activation by affecting DUX4/Dux transcription. *IScience*, 26(3), 106158. <https://doi.org/10.1016/j.isci.2023.106158>
- Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., Xu, F., & Xie, W. (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell*, 63(6), 1066–1079. <https://doi.org/10.1016/j.MOLCEL.2016.08.032>
- Zhong, Y. F., & Holland, P. W. H. (2011). The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. *BMC Evolutionary Biology*, 11(1), 169. <https://doi.org/10.1186/1471-2148-11-169>

- Zhou, C., Halstead, M. M., Elie Bonnet-Garnier, A., Schultz, R. M., & Ross, P. J. (2023). Histone remodeling reflects conserved mechanisms of bovine and human pre-implantation development. *EMBO Reports*, 24(3), e55726. <https://doi.org/10.15252/EMBR.202255726>
- Zhou, C., Wang, Y., Zhang, J., Su, J., An, Q., Liu, X., Zhang, M., Wang, Y., Liu, J., & Zhang, Y. (2019). H3K27me3 is an epigenetic barrier while KDM6A overexpression improves nuclear reprogramming efficiency. *The FASEB Journal*, 33(3), 4638–4652. <https://doi.org/10.1096/fj.201801887R>
- Zhou, K., Wang, T., Zhang, J., Zhang, J., Liu, X., Guan, J., Su, P., Wu, L., Yang, X., Hu, R., Sun, Q., Fan, Z., Yang, S., Chu, X., Song, W., Shang, Y., Zhou, S., Hao, X., Zhang, X., ... Miao, Y. L. (2024). LEUTX regulates porcine embryonic genome activation in somatic cell nuclear transfer embryos. *Cell Reports*, 43(6), 114372. <https://doi.org/10.1016/j.celrep.2024.114372>
- Zhu, L., Zhou, T., Iyyappan, R., Ming, H., Dvoran, M., Wang, Y., Chen, Q., Roberts, R. M., Susor, A., & Jiang, Z. (2022). High-resolution ribosome profiling reveals translational selectivity for transcripts in bovine preimplantation embryo development. *Development*, 149(21), dev200819. <https://doi.org/10.1242/DEV.200819>
- Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., Wang, X., Wei, Y., Liu, P., Yan, J., Ren, X., Yuan, P., Yuan, Y., Yan, Z., Wen, L., ... Tang, F. (2018). Single-cell DNA methylome sequencing of human pre-implantation embryos. *Nature Genetics*, 50(1), 12–19. <https://doi.org/10.1038/S41588-017-0007-6>
- Zou, Z., Wang, Q., Wu, X., Schultz, R. M., & Xie, W. (2024). Kick-starting the zygotic genome: licensors, specifiers, and beyond. *EMBO Reports*, 25(10), 4113–4130. <https://doi.org/10.1038/S44319-024-00223-5>
- Zou, Z., Zhang, C., Wang, Q., Hou, Z., Xiong, Z., Kong, F., Wang, Q., Song, J., Liu, B., Liu, B., Wang, L., Lai, F., Fan, Q., Tao, W., Zhao, S., Ma, X., Li, M., Wu, K., Zhao, H., ... Xie, W. (2022). Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science*, 378(6615), abo7923. <https://doi.org/10.1126/SCIENCE.ABO7923>

ACKNOWLEDGEMENTS

Throughout my PhD, many people supported and helped me along the way. Completing this journey would not have been possible without them.

I would like to thank my supervisors at the University of Tartu, Prof. Ants Kurg and Dr. Tõnis Org. Ants, you were the first person I met in person in Estonia, and I felt welcomed and supported from the moment I stepped foot in the country until this very day. I am grateful for your kindness and guidance through all the challenges I encountered over the years. I also appreciate the trust and opportunity you gave me to grow as a researcher. Tõnis, you have been my go-to person whenever I needed help with experiments in the lab, and you always made yourself available and generously offered your time and assistance. I could not have successfully completed the experiments without your support. I really appreciate your willingness to jump in to solve problems, not only related to the projects but also with general PhD-related struggles.

I would also like to thank my supervisors at the Karolinska Institutet Prof. Juha Kere and Dr. Shintaro Katayama. Juha, despite being supervised mostly remotely due to the nature of this joint PhD program, I always felt very included in your research team, which I really appreciate. You helped me gradually take on more responsibility and become more independent, which is what a PhD is all about. I am grateful for your trust and encouragement throughout my PhD, which was an important driving force to move forward. Shintaro, I am very grateful for the time and effort you devoted to supervising me. I started with no computational skills, and through your immense support and dedication, I have reached a level where I can now conduct analyses independently. I cannot overstate your contribution to my progress, which has shaped my development as a researcher.

I would like to thank Prof. Andres Salumets for making my doctoral project possible by writing the grant that brought together the members of the MATER Consortium and created a collaborative and stimulating PhD experience for me.

I would also like to thank my peers in the group, Dr. Nina Boskovic and Gamze Yazgeldi Günaydın. Nina, I appreciate your support during my PhD and am glad our close collaboration resulted in several completed projects. I am also happy that we always found a way to move forward productively. Gamze, I am grateful for your kindness and support with the computational work, and for always being available to help. I am happy to have had the chance to collaborate with you.

Thank you, Merle Külaots and Rita Kreevan. Merle, I greatly appreciate your help with the experiments and your company. Working with you has always been enjoyable. Your positivity and our chats made even the hardest days easier. Rita, walking this path with a fellow PhD student and a rant buddy for failures and PhD life has been essential, and I am happy we got to share it. I really appreciate your companionship throughout my PhD.

I would like to thank Dr. Thomas Bürglin for the time he invested in assisting me with the evolutionary analysis, and I am grateful to Dr. Marilin Ivask and

Prof. Ülle Jaakma for agreeing to collaborate and providing access to bovine material. I also extend my thanks to Dr. Eeva-Mari Jouhilahti and Dr. Jere Weltner for sharing their experience and insights, which greatly supported my progress in experimental work. I am especially thankful to Dr. Tiina Skoog for being the person I could reach out to at Karolinska, always offering help both remotely and during my secondment in Stockholm.

Thank you Kadri Raav and Gerly Põder for the excellent and very detailed administrative support you provided throughout the program. I am also grateful to Sulev Kuuse for his generous help at the beginning of my PhD, and to Annely Kukk for her support and kindness. Also, thank you Piibe Vill for being a great Master's student and for making my first teaching and supervision experience such a good one.

I would like to acknowledge the writing retreats I attended while working on this thesis, organized by the Institute of Genomics and Institute of Pharmacy at the University of Tartu, both held jointly by the Estonian Doctoral School. These retreats were highly beneficial for my writing. I would also like to acknowledge the support of the members of the UT HPC Center.

I am grateful to Karolinska Institutet Examination Board Members, Prof. Catharina Larsson, Dr. Jan-Bernd Stukenborg, Dr. Sergey Rodin, for investing their time in reviewing my work. I also appreciate the time Dr. Tambet Tõnissoo, Reviewer at the University of Tartu, devoted to providing valuable feedback on my thesis, which has greatly improved its quality and readability.

I would like to thank the people I have had great times with in Estonia: Natàlia, Spyros, Niels, Rita, Eric, Hanna, Keiu, Meruert, Seungbaek, Nina, Yan, Ayesh, Naila, Mike, Lindsay, and of course, Sooty, my favorite cat in Tartu. Spending time with all of you made this experience completely unique. I would also like to thank the other MATER students for the moments we shared throughout the program: Nidhi, Laura, Maria, Yiqun, Lena, and Kate. In addition, I am deeply grateful to all the kind, easygoing, and welcoming Estonian people I met during my time here.

I am also grateful to those whose support began well before my PhD and continued throughout. Beril, I cannot express how much I appreciate the support I got from you! Your infectious sense of humor has become my survival skill. You are a true friend who can turn bad days into good ones.

Special thanks to my brother Özgü, who has been a constant source of support in my life. I am lucky that this support has recently been doubled by my friend, former classmate and now sister-in-law, Yağmur. And very special thanks to my dearest niece, Talin, who made me a first-time uncle. You have brought so much joy and made everything feel lighter.

My deepest thanks go to my dearest mother and father, who have provided me with unconditional love and support since the very beginning, probably since I first expressed *DUX4*. Your presence has given me strength, and I am incredibly lucky to have you by my side in every endeavor.

PUBLICATIONS

CURRICULUM VITAE

Name: Barış Yaşar
Date of Birth: 06.06.1991
Address: Institute of Molecular and Cell Biology, University of Tartu,
Riia 23, Tartu, Estonia
E-mail: baris.yasar@ut.ee; baris.yasar@ki.se

Education

09/2020–04/2026 MSCA PhD fellow – Double PhD Degree Programme between the University of Tartu (Estonia) and Karolinska Institutet (Sweden)
10/2014–09/2017 MSc in Molecular Biosciences, Heidelberg University (Germany)
01/2010–06/2014 BSc in Molecular Biology and Genetics, Bilkent University (Turkey)

Employment

09/2020–08/2026 Junior Research Fellow in Biotechnology at the University of Tartu
09/2017–04/2018 Research Assistant at Heidelberg University Biochemistry Center

Conference Presentations

12/2020–12/2024 Oral and poster presentations at the joint annual conference of the Institute of Molecular and Cell Biology and the Institute of Genomics, University of Tartu in Tartu, Estonia
10/2023 Oral presentation at the Karl Ernst von Baer's seminar in Tartu, Estonia
03/2023 Poster presentation at the 36th International Mammalian Genome Conference in Tsukuba, Japan
10/2022 Oral presentation at MATER Consortium Dissemination Workshop in Maastricht, the Netherlands
06/2022 Poster presentation at the European Society of Human Genetics Annual Conference (ESHG) in Vienna, Austria

Teaching

Teaching Assistant, Practical Course in Biotechnology (2022)
Co-supervisor, Piibe Vill – Master's Thesis (2021–2022)

Awards and Scholarships

Best poster presentation award at the joint annual conference of the UT Institute of Molecular and Cell Biology and the UT Institute of Genomics (2022)

Merit-Based Scholarship awarded through national examination entrance examination Bilkent University (2011–2014)

Radboud University Summer School Scholarship (2013)

Bilkent University Honor/High-Honor Student (2010–2014)

Publications

Yaşar, B., Org, T., Ivask, M., Gunaydin, G. Y., Boskovic, N., Jaakma, Ü., Kere, J., Kurg, A., & Katayama, S. (2026). Structure and genomic organization of the human DUX4 homologue bovine DUXC. *BioRxiv*, 2026.03.05.709741. <https://doi.org/10.64898/2026.03.05.709741>

Yaşar, B.*, Boskovic, N.*, Org, T., Kere, J., Kurg, A., & Katayama, S. (2025). RNA sequencing of mRNA 5'-ends reveals regulators of bovine embryonic genome activation. *BMC Genomics*, 26(1), 910. <https://doi.org/10.1186/S12864-025-12110-X> PMID: 41087861.

Boskovic, N., Ivask, M., Yazgeldi Gunaydin, G., **Yaşar, B.**, Katayama, S., Salumets, A., Org, T., Kurg, A., Lundin, K., Tuuri, T., Daub, C. O., & Kere, J. (2025). Oxygen level alters energy metabolism in bovine preimplantation embryos. *Scientific Reports*, 15(1), 11327. <https://doi.org/10.1038/s41598-025-95990-z> PMID: 40175462.

Yaşar, B., Boskovic, N., Ivask, M., Weltner, J., Jouhilahti, E. M., Vill, P., Skoog, T., Jaakma, Ü., Kere, J., Bürglin, T. R., Katayama, S., Org, T., & Kurg, A. (2024). Molecular cloning of PRD-like homeobox genes expressed in bovine oocytes and early IVF embryos. *BMC Genomics*, 25(1), 1048. <https://doi.org/10.1186/S12864-024-10969-W> PMID: 39506635.

Boskovic, N., Yazgeldi, G., Ezer, S., Tervaniemi, M. H., Inzunza, J., Deligiannis, S. P., **Yaşar, B.**, Skoog, T., Krjutškov, K., Katayama, S., & Kere, J. (2023). Optimized single-cell RNA sequencing protocol to study early genome activation in mammalian preimplantation development. *STAR Protocols*, 4(3), 102357. <https://doi.org/10.1016/j.xpro.2023.102357> PMID: 37314922.

ELULOOKIRJELDUS

Nimi: Barış Yaşar
Sünniaeg: 06.06.1991
Aadress: Tartu Ülikooli molekulaar- ja rakubioloogia instituut, Riia 23, Tartu, Eesti
E-post: baris.yasar@ut.ee; baris.yasar@ki.se

Haridus

09/2020–04/2026 MSCA PhD tudeng – Tartu Ülikool (Eesti) ja Karolinska Instituut (Rootsi)
10/2014–09/2017 MSc (Molekulaarbioloogia), Heidelbergi Ülikool (Saksamaa)
01/2010–06/2014 BSc (Molekulaarbioloogia ja Geneetika), Bilkenti Ülikool (Türgi)

Teenistuskäik

09/2020–08/2026 Biotehnoloogia Nooremteadur, Tartu Ülikool
09/2017–04/2018 Teadusassistent, Heidelbergi Ülikooli Biokeemia keskus

Konverentsiettekanded

12/2020–12/2024 Suulised ja posterettekanded Tartu Ülikooli molekulaar- ja rakubioloogia instituudi ning genoomika instituudi aasta-konverentsil Tartus, Eestis
10/2023 Suuline ettekanne Karl Ernst von Baeri seminaril Tartus, Eestis
03/2023 Posterettekanne 36. rahvusvahelisel Imetajate Genoomi konverentsil (International Mammalian Genome Conference) Tsukubas, Jaapanis
10/2022 Suuline ettekanne MATER konsortsiumi levitamise töötoas Maastrichtis, Madalmaades
06/2022 Posterettekanne Euroopa Inimgeneetika Seltsi aastakonverentsil (ESHG) Viinis, Austrias

Õpetamiskogemus

Õppeassistent, Praktiline biotehnoloogia (2022)
Magistritöö kaasjuhendaja, Piibe Vill (2021–2022)

Auhinnad ja stipendiumid

Parima posteresitluse auhind Tartu Ülikooli molekulaar- ja rakubioloogia instituudi ning Tartu Ülikooli genoomika instituudi ühisel aastakonverentsil (2022)

Tulemuspõhine stipendium, määratud riikliku sisseastumiseksami tulemuste alusel, Bilkenti Ülikool (2011–2014)

Radboudi Ülikooli Suvekooli Stipendium (2013)

Bilkenti Ülikooli “honor”/“high honor” õpilane (2010–2014)

Teaduspublikatsioonid

Yaşar, B., Org, T., Ivask, M., Gunaydin, G. Y., Boskovic, N., Jaakma, Ü., Kere, J., Kurg, A., & Katayama, S. (2026). Structure and genomic organization of the human DUX4 homologue bovine DUXC. *BioRxiv*, 2026.03.05.709741. <https://doi.org/10.64898/2026.03.05.709741>

Yaşar, B.*, Boskovic, N.*, Org, T., Kere, J., Kurg, A., & Katayama, S. (2025). RNA sequencing of mRNA 5'-ends reveals regulators of bovine embryonic genome activation. *BMC Genomics*, 26(1), 910. <https://doi.org/10.1186/S12864-025-12110-X> PMID: 41087861.

Boskovic, N., Ivask, M., Yazgeldi Gunaydin, G., **Yaşar, B.**, Katayama, S., Salumets, A., Org, T., Kurg, A., Lundin, K., Tuuri, T., Daub, C. O., & Kere, J. (2025). Oxygen level alters energy metabolism in bovine preimplantation embryos. *Scientific Reports*, 15(1), 11327. <https://doi.org/10.1038/s41598-025-95990-z> PMID: 40175462.

Yaşar, B., Boskovic, N., Ivask, M., Weltner, J., Jouhilahti, E. M., Vill, P., Skoog, T., Jaakma, Ü., Kere, J., Bürglin, T. R., Katayama, S., Org, T., & Kurg, A. (2024). Molecular cloning of PRD-like homeobox genes expressed in bovine oocytes and early IVF embryos. *BMC Genomics*, 25(1), 1048. <https://doi.org/10.1186/S12864-024-10969-W> PMID: 39506635.

Boskovic, N., Yazgeldi, G., Ezer, S., Tervaniemi, M. H., Inzunza, J., Deligiannis, S. P., **Yaşar, B.**, Skoog, T., Krjutškov, K., Katayama, S., & Kere, J. (2023). Optimized single-cell RNA sequencing protocol to study early genome activation in mammalian preimplantation development. *STAR Protocols*, 4(3), 102357. <https://doi.org/10.1016/j.xpro.2023.102357> PMID: 37314922.

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 fluorescence-based pH-dependent fiber optic penicillin flowthrough 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 environmental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben.** Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes.** Respiration rhythms 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 temperate 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 micro-organisms 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, Keroplattidae, 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 O₃ and CO₂ on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptone-mal 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 invasions 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 Northern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn4652 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 N₂ 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 Roots.** 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 Tn4652 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 Tn4652 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 cerevisiae*. 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 green-finches – 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 CO₂ 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 Semtsenko.** 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 Säälük.** 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 Tn4652 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 Värvi.** Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Vääk.** 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 Uusküla.** 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 boreo-nemoral 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 Nagiraja.** 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 CO₂ and O₃ on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and inter-annual patterns. Tartu, 2014, 115 p.
263. **Küllli 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 CO₂ 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 (Hanse-nula) 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.
368. **Viktorija Kukuškina.** Understanding the mechanisms of endometrial receptivity through integration of ‘omics’ data layers. Tartu, 2020, 169 p.
369. **Martti Vasar.** Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur.** Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post.** Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.
372. **Kristi Krebs.** Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkiivi.** Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida*. Tartu, 2020, 154 p.
374. **Elin Soomets.** Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk.** Signals and responses of ColRS two-component system in *Pseudomonas putida*. Tartu, 2020, 133 p.
376. **Indrek Teino.** Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre.** The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp.** Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing.** Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent.** Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.

381. **Einar Kärgerberg.** Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere.** The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante.** Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.
384. **Ajai Kumar Pathak.** Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter.** Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.
386. **Burak Yelmen.** Characterization of ancient Eurasian influences within modern human genomes. Tartu, 2021, 134 p.
387. **Linda Ongaro.** A genomic portrait of American populations. Tartu, 2021, 182 p.
388. **Kairi Raime.** The identification of plant DNA in metagenomic samples. Tartu, 2021, 108 p.
389. **Heli Einberg.** Non-linear and non-stationary relationships in the pelagic ecosystem of the Gulf of Riga (Baltic Sea). Tartu, 2021, 119 p.
390. **Mickaël Mathieu Pihain.** The evolutionary effect of phylogenetic neighbourhoods of trees on their resistance to herbivores and climatic stress. Tartu, 2022, 145 p.
391. **Annika Joy Meitern.** Impact of potassium ion content of xylem sap and of light conditions on the hydraulic properties of trees. Tartu, 2022, 132 p.
392. **Elise Joonas.** Evaluation of metal contaminant hazard on microalgae with environmentally relevant testing strategies. Tartu, 2022, 118 p.
393. **Kreete Lüll.** Investigating the relationships between human microbiome, host factors and female health. Tartu, 2022, 141 p.
394. **Triin Kaasiku.** A wader perspective to Boreal Baltic coastal grasslands: from habitat availability to breeding site selection and nest survival. Tartu, 2022, 141 p.
395. **Meeli Alber.** Impact of elevated atmospheric humidity on the structure of the water transport pathway in deciduous trees. Tartu, 2022, 170 p.
396. **Ludovica Molinaro.** Ancestry deconvolution of Estonian, European and Worldwide genomic layers: a human population genomics excavation. Tartu, 2022, 138 p.
397. **Tina Saupe.** The genetic history of the Mediterranean before the common era: a focus on the Italian Peninsula. Tartu, 2022, 165 p.
398. **Mari-Ann Lind.** Internal constraints on energy processing and their consequences: an integrative study of behaviour, ornaments and digestive health in greenfinches. Tartu, 2022, 137 p.
399. **Markus Valge.** Testing the predictions of life history theory on anthropometric data. Tartu, 2022, 171 p.
400. **Ants Tull.** Domesticated and wild mammals as reservoirs for zoonotic helminth parasites in Estonia. Tartu, 2022, 152 p.

401. **Saleh Rahimlouye Barabi.** Investigation of diazotrophic bacteria association with plants. Tartu, 2022, 137 p.
402. **Farzad Aslani.** Towards revealing the biogeography of belowground diversity. Tartu, 2022, 124 p.
403. **Nele Taba.** Diet, blood metabolites, and health. Tartu, 2022, 163 p.
404. **Katri Pärna.** Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes. Tartu, 2022, 190 p.
405. **Silva Lilleorg.** Bacterial ribosome heterogeneity on the example of bL31 paralogs in *Escherichia coli*. Tartu, 2022, 189 p.
406. **Oliver Aasmets.** The importance of microbiome in human health. Tartu, 2022, 123 p.
407. **Henel Jürgens.** Exploring post-translational modifications of histones in RNA polymerase II-dependent transcription. Tartu, 2022, 147 p.
408. **Mari Tagel.** Finding novel factors affecting the mutation frequency: a case study of tRNA modification enzymes TruA and RluA. Tartu, 2022, 176 p.
409. **Marili Sell.** The impact of environmental change on ecophysiology of hemiboreal tree species – acclimation mechanisms in belowground. Tartu, 2022, 163 p.
410. **Kaarin Hein.** The hissing behaviour of Great Tit (*Parus major*) females reflects behavioural phenotype and breeding success in a wild population. Tartu, 2022, 96 p.
411. **Maret Gerz.** The distribution and role of mycorrhizal symbiosis in plant communities. Tartu, 2022, 206 p.
412. **Kristiina Nõomaa.** Role of invasive species in brackish benthic community structure and biomass changes. Tartu, 2023, 151 p.
413. **Anton Savchenko.** Taxonomic studies in Dacrymycetes: *Cerinomyces* and allied taxa. Tartu, 2023, 181 p.
414. **Ahto Agan.** Interactions between invasive pathogens and resident mycobio-
me in the foliage of trees. Tartu, 2023, 155 p.
415. **Diego Pires Ferraz Trindade.** Dark diversity dynamics linked to global change: taxonomic and functional perspective. Tartu, 2023, 134 p.
416. **Madli Jõks.** Biodiversity drivers in oceanic archipelagos and habitat fragments, explored by agent-based simulation models. Tartu, 2023, 116 p.
417. **Ciara Baines.** Adaptation to oncogenic pollution and natural cancer defences in the aquatic environment. Tartu, 2023, 164 p.
418. **Rain Inno.** Placental transcriptome and miRNome in normal and complicated pregnancies. Tartu, 2023, 145 p.
419. **Daniyal Gohar.** Diversity, genomics, and potential functions of fungus-inhabiting bacteria. Tartu, 2023, 138 p.
420. **Sirli Rosendahl.** Fitness effects of chromosomal toxin-antitoxin systems in *Pseudomonas putida*. Tartu, 2023, 154 p.
421. **Mathilde Frédérique E. André.** New Guinea, a hotspot for Human evolution: settlement history and adaptation in northern Sahul. Tartu, 2023, 202 p.

422. **Vlad-Julian Piljukov.** Biochemical characterization of Irc3 helicase. Tartu, 2023, 137 p.
423. **Gerli Albert.** Carbon use strategies of macrophyte communities in the northeastern Baltic Sea: implications for a high CO₂ environment. Tartu, 2023, 128 p.
424. **Mariann Koel.** The molecular interactions between trophoblast and endometrial cells in embryo implantation. Tartu, 2023, 171 p.
425. **Robin Gielen.** Diversity and ecological role of pathogenic fungi in insect populations. Tartu, 2023, 139 p.
426. **Kaspar Reier.** Quantity, stability and disparity of ribosomal components in *Escherichia coli* stationary phase. Tartu, 2023, 151 p.
427. **Linda Rusalepp.** The impact of environmental drivers and competition on phenolic metabolite profiles in hybrid aspen and silver birch. Tartu, 2023, 153 p.
428. **Eliisa Pass.** The effect of managed forest-wetland landscapes on forest grouse and nest predation. Tartu, 2023, 115 p.
429. **Sanni Färkkilä.** Methods for studying plant-fungal interactions – reflecting on the old, the new and the upcoming. Tartu, 2024, 147 p.
430. **Maarja Jõeloo.** Advances in microarray-based copy number variation discovery and phenotypic associations. Tartu, 2024, 209 p.
431. **Natàlia Pujol Gualdo.** Decoding genetic associations of female reproductive health traits. Tartu, 2024, 205 p.
432. **Sirelin Sillamaa.** The role of helicases Hmi1 and Irc3 in yeast mitochondrial DNA maintenance. Tartu, 2024, 189 p.
433. **Iris Reinula.** Genetic variation of grassland plants in changing landscapes. Tartu, 2024, 201 p.
434. **Vi Ngan Tran.** The cellular dynamics and epithelial morphogenesis in *Drosophila* wing development. Tartu, 2024, 158 p.
435. **Slendy Julieth Rodríguez Alarcón.** Intraspecific trait diversity in plants: characterizing effects of trait variation on community assembly and ecosystem functioning. Tartu, 2024, 129 p.
436. **Arun Kumar Devarajan.** Microbes and climate change: insights from plant-microbe interactions in rice phyllosphere and soil microbiomes in subarctic grasslands. Tartu, 2024, 224 p.
437. **Leonard Owuraku Opare.** Rearing density effects on a commercially important insect species. Tartu, 2024, 145 p.
438. **Siqiao Liu.** The effect of anthropogenic disturbance on soil fungal communities. Tartu, 2024, 172 p.
439. **Kertu Liis Krigul.** The gut microbiome at the interface of human health and disease. Tartu, 2024, 158 p.
440. **Danat Yermakovich.** The evolutionary history of complex traits: implications of archaic admixture. Tartu, 2024, 153 p.
441. **Yiming Meng.** Plant mycorrhizal type and status in the global flora. Tartu, 2024, 200 p.

442. **Iryna Yatsiuk.** Evolution, species delimitation and diversity in myxomycetes: *Arcyria* and allied genera. Tartu, 2024, 193 p.
443. **Daniela León Velandia.** Mycorrhizal trait distribution and composition in plant communities under natural gradients. Tartu, 2024, 121 p.
444. **Bruno Paganeli.** Dark diversity methods for prioritization of areas and species in nature conservation. Tartu, 2024, 155 p.
445. **Mario Reiman.** Placental transcriptome in normal and complicated pregnancies. Tartu, 2025, 167 p.
446. **Maarja Kõrkjas.** Dynamics of tree-related microhabitats in live forest trees and its links with biodiversity. Tartu, 2025, 134 p.
447. **Eleonora Beccari.** Mapping and exploring trait spaces across the tree of life. Tartu, 2025, 190 p.
448. **Jack R. Hall.** Dissolved organic carbon dynamics of Baltic Sea macroalgae: production, bioavailability and ecosystem effects. Tartu, 2025, 135 p.
449. **Artjom Stepanjuk.** Function of adhesion molecules and signalling pathways in human endometrial and embryonic models. Tartu, 2025, 247 p.
450. **Marianne Kivastik.** Heterostylous plants in an era of global change: the role of local, landscape and climatic actors. Tartu, 2025, 167 p.
451. **Yehor Yatsiuk.** Large tree-cavities as key structures for forest biodiversity. Tartu, 2025, 215 p.
452. **Ovidiu Copoț.** Relevance of eDNA, citizen science, and species distribution modelling for fungal conservation. Tartu, 2025, 198 p.
453. **Tarmo Puurand.** Human genome studies with k-mer frequencies. Tartu, 2025, 184 p.
454. **Stênio Ítalo Araújo Foerster.** Phylogenetic comparative studies of body size in insects and arachnids: from predictions to applications. Tartu, 2025, 171 p.
455. **Hanna Maria Kariis.** Improving pharmacotherapy outcomes in psychiatric and cardiovascular conditions. Tartu, 2025, 193 p.
456. **Elisabeth Prangel.** The impact of land-use change and ecological restoration on biodiversity and ecosystem service supply in semi-natural grasslands. Tartu, 2025, 233 p.
457. **Nidal Fetnassi.** Determinants of moth assemblages across human-modified landscapes of Estonia and Morocco. Tartu, 2025, 162 p.
458. **Vineesh Nedumpally.** Assembling the phylogenetic tree of northern European macroheteroceran moths. Tartu, 2025, 171 p.
459. **Ali Hakimzadeh.** Long-read metabarcoding: from available tools to reference databases. Tartu, 2026, 124 p.
460. **John Yangyuoru Kupagme.** Biodiversity of African soil fungi. Tartu, 2026, 162 p.