Resolving the complexity of oviduct biology and pathology by comprehensive single-cell sequencing across species



Laura Santana González

Word count: 43.000 words

Nuffield Department of Women's and Reproductive Health

Balliol College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

November 2021

A mis padres, Loli y Franci, por su infinito amor, entrega y dedicación. Por darme mi vida y la suya. Por darme alas.

Noviembre 2021, Oxford

To my loving parents, Loli and Franci, who are the wind beneath my wings.

November 2021, Oxford

Resolving the complexity of oviduct biology and pathology by comprehensive single-cell sequencing across species

Laura Santana González

Abstract

The recent past has witnessed a substantial revolution in the omics fields due to the exponential development of powerful cutting-edge platforms and the optimisation of their protocols. Concretely, the ability to survey single transcriptomes composing a biological network through single-cell RNA sequencing (scRNA-seq) has translated into substantial innovation and exciting scientific discoveries in both basic and applied research. In this project, this technique was combined with a biological systems approach to gain deeper understanding of the identities and processes regulating the human fallopian tube and the mouse oviduct, two equivalent reproductive organs that play a key role in female health and fertilisation. The first part of this thesis focuses on establishing a comprehensive single-cell atlas of the mouse oviducts, which was one of the few mouse organs that had not been deeply characterised yet. To this end, an extensive and careful technical optimisation was undertaken, followed by an experimental phase and a subsequent analysis and initial validation of the high-throughput murine data. The second part of the work then focused on advancing our understanding of the fallopian tubes and establishing, for the first time, inter-species comparisons between the murine and human tubes. Furthermore, the results obtained by scRNA-seq were used to interrogate large bulk transcriptomic datasets from ovarian cancers, in search of putative biological mechanisms that under homeostatic conditions maintain the physiological status of this organ, but under unfavourable conditions could play a role in these pathological conditions.

The initial technical optimisation enabled the design of the most effective cell isolation protocol, specifically tailored for this organ. This optimised protocol maximised cell recovery, improved cell viability and allowed the isolation of a balanced and representative oviduct cell population. This protocol generated novel scRNA-seq data from the mouse oviducts, which along with the creation of a robust and consistent bioinformatics analytical pipeline, enabled the extraction of functional insights into the oviduct biology, some of which, were initially validated. Importantly, this bioinformatics analysis was able to capture not only the previously known cell lineages comprising the oviducts, but also, to identify novel ones, including new cellular subtypes and intermediate cell phenotypes requiring further investigation due to their potential importance in the organ differentiation and homeostasis. Moreover, the newly generated oviduct data allowed a comparison between the human fallopian tube versus the mouse oviduct populations and cell-cell interactions maintaining their biology for the first time. Results from this analysis highlighted the dominant presence of the adaptive immune system in the

fallopian tube in contrast to the oviduct situation, where the innate system was more represented. Furthermore, profound differences between the oviduct and fallopian tube cell identities were detected, even for those belonging to the same lineage and type. Despite these differences, the identified biological interactions among the cellular components and the molecular pathways underlying such interactions remained very similar across species. Finally, this project was also able to demonstrate how advanced insights into the cell identities and molecular processes characterising the healthy fallopian tubes and oviducts can serve the purpose of advancing our understanding of the diseased state and the mechanisms behind it. Indeed, scRNA-seq analysis of both species' organs enabled a series of bioinformatic analyses that highlighted pathways and cell populations potentially relevant for the pathogenesis of ovarian cancers. A finding that could open the door to new therapeutical avenues subject to the appropriate mechanistic validation of the data-driven evidence.

Acknowledgements

First, my most sincere and genuine gratitude goes to my mentor, my supervisor and my friend, Professor Ahmed Ahmed. I have learnt an endless list of things from him, but I especially thank the long hours of Science we have made and discussed together, the intellectual stimulation (and dreams) he awakened and his firm support. Furthermore, I am forever grateful to him and Abeer for welcoming me and being home.

Second, I would like to acknowledge the nice bunch of people the lab has had since I joined. How I miss those days! Yiyan, Mo, Kenta, Nina, Zhiyuan, Nuria, Doni, Matteo, Tingyan, Mara, and, of course, the two very dear and beloved friends that the lab gave me, Salma and Kay. Special thanks to Mara as well, for her friendship and for proofreading and supporting this thesis.

My luck did not only stay in the lab, but expanded to the WIMM, which gave me one of my most special friends and family, Ioanna (and her extensions, Ale and Carlo). Her example, not only as a scientist, but also as the superwoman that she is, guided and inspired me throughout these years. Furthermore, my most sincere thanks to my friends Andrew and Héctor, who have been an absolute gift in my life. Also, how to forget the nice people that shared that time with me in the WIMM: Margarida, Alice, Erdinc, Max, Falk, Merve, Ivan, Mirian, Athena, Oana, Stefano, Hanifi. Of course, huge thanks to my favourite WIMM facility ever, the Flow Cytometry facility, whose members (especially Kevin!) are not only the most helpful, but the nicest people by far. Many thanks as well to Leticia for her patience, her diligence and her help with immunohistochemistry.

To my Oxford friends, thank you for sharing and contributing to some of the most beautiful years of my life, which I have lived with real passion. Alba, Santiago, Alejandro, Rodrigo, Viviana, Homero, Peter, Miren, Ferdinando, Tika, Sofia, Tatiana, Veronica, Ana, Giorgio, Teresa, Laura, Oscar, Sangseraima, Rob, Cleo, Karolina, Olivia, Flo, Matt, Gonzalo, Tharindi, Lucía, Dan, Alice, Dom.

Moreover, my infinite gratitude to the people with some of the most special places in my heart. My two grandmothers, Vicenta and Pepa, one gone, the other here, but both eternal. Thank you for everything you have taught me with your example and all the love you have given me. In their absolute divergent, but complementary ways, they have been fierce modern women living ahead of their difficult times. To their products, my beloved parents, who have been a source of inspiration, happiness and support. Who have raised, loved and educated their children to their best of their possibilities, and beyond. Whom I miss every single day of my life far away from them. Thanks to my brother, Jaime, who has the purest of hearts and didn't stop loving his big sister since he was born (although it took a bit longer for the opposite to happen) (I wasn't too sure of losing my royal rights to be the only spoiled sassy queen). Thank you to my tita Inma, a strong and wonderful woman that leads with her example. To my cousins, Cristina, Carlos, Juanfri, Claudia, Lorena and Nuria, who share beautiful memories with me, after a whole life growing up together. To my uncles and aunts, Jesús, Rufi, Juanfra, Arantxa and Domi, for the amazing people they have always been to me.

Furthermore, this thesis would have never happened without my sisters, my big women, my partners in crime. Alicia, Julia, Lorena and Victoria, who have always blindly loved, believed and supported me through every chapter of my life (and every single of my occurrences) (this one is nice, though). Also, huge thanks to a big part of myself, my life and home: Elena (and her extension, Roy), Victoria (and her extension, Candela), Ángela, Cristina, María, Belén (and her extension, Danielle), Remi, Jaime and Guille. Last, but not least, my deepest gratitude to my dear friends, Celia, Redu, Chabeli, Isa, María Jesús, and Ana Belén, who shared the dream and golden years of university and who are some of the most important people to me.

A special feeling of gratitude to my loving partner, Massimo, who has relentlessly held my hand through any situation. Thank you for your endless support and sharing your life with me. I cannot wait to continue opening and closing adventures together (next one coming!). I would also like to extend my thanks to the sweet Masieri: Mari, Lino, Luca, Anna and, the purest and most beautiful creatures on Earth, Thomas and Gioia (nostri bibinis) (blank space if new more coming).

Finally, my most sincere gratitude to the free and public Spanish and Andalusian educational system, which enabled me to get here. To my land and home, Lepe. To all my primary and secondary school teachers, who I hold very dear and think of frequently, Irene, Mariceli, Pilar Montes, Juan Redondo, Isabel Bueno, Ainhoa, Almudena, Lourdes, Eduardo, etc. Also, to the UK educational system, BBSRC, Esther and Gail, the University of Oxford, the NDWRH department (especially Delphine, Danielle, Simon, Lesley and Tom) and Sonia, who offered me this opportunity.

In summary, I am blessed.

Declaration

I declare that no part of this thesis has been accepted or submitted for any degree, diploma, certificate or other qualification in this University or elsewhere. The results presented herein are my own. Cases where help has been received are appropriately specified.

First-author publications

Santana Gonzalez, L., Rota, I. A., Artibani, M., Morotti, M., Hu, Z., Wietek, N., Alsaadi, A., Albukhari, A., Sauka-Spengler, T. and Ahmed, A. A. (2021). Mechanistic drivers of Müllerian duct development and differentiation into the oviduct. Frontiers in Cell and Developmental Biology. 9:605301.

Santana González, L., Artibani, M. and Ahmed, A. A. (2021). Studying Müllerian duct anomalies - from cataloguing phenotypes to discovering causation. Diseases Models & Mechanisms. 1;14(6):dmm047977.

Table of Contents

Table of Contents **10**

List of Acronyms **15**

List of Figures **18**

List of Tables **20**

I Introduction **21**

1 Embryonic development of the mouse and human Müllerian ducts . . 22

1.1 Progenitors' differentiation 22

1.2 Progenitors' invagination 24

1.3 Müllerian caudal elongation 25

2 Müllerian duct differentiation 27

3 Oviduct formation 28

4 The oviduct and the reproductive function 32

4.1 General transport mechanisms in the oviduct 32

4.2 Ovulation, egg pickup and oviduct-egg communication 33

4.3 The sperm journey and interactions with the female reproductive tract 35

4.4 Fertilisation and early development of the pre-implantation embryo 38

5 The oviduct as the origin and site of diseases 40

6 Prospects in the field and the role of this project 42

II Aims and objectives **44**

III Methods **45**

1 Technical optimisation of oviduct single-cell isolation 45

1.1 Mouse work 45

1.2 Sample processing and tissue dissociation 45
1.3 Fuorescence-activated cell sorting (FACS) 45
1.4 RNA extraction and quantitative reverse transcription polymerase chain reaction (RT-qPCR) for the quantification of immediate early genes (IEGs) expression 46
2 Single-cell transcriptomic atlas of the mouse oviducts and data validation 47
2.1 Nuclei extraction from mouse oviducts for single-nucleus RNA-Seq (snRNA-seq) 47
2.2 10X Genomics 48
2.3 Pre-processing and initial filtering of scRNA-seq data 48
2.4 Doublet identification and removal 48
2.5 scRNA-seq integration, clustering and differential-expression (DE) analysis 48
2.6 Visualisation of the gene expression and associated metrics . . 49
2.7 Study of biological terms enriched in specific clusters 49
2.8 Generation of 3D mouse oviduct organoids 50
2.9 Passage of 3D mouse oviduct organoids 50
2.10 Immunofluorescence staining 50
2.11 Immunohystochemistry (IHC) 51
3 Cross-species comparison of oviduct populations composition and interactions in mouse and human 51
3.1 Publicly available normal human fallopian tube scRNA-seq datasets 51
3.2 Patient samples 52
3.3 Human 10X Genomics and data analysis 52
3.4 Statistical correlation analysis and generation of the correlation matrix 52
3.5 Cell-cell ligand-receptor analysis with CellPhoneDB 52

4 RNA-Seq and correlation analyses . . . 53
4.1 Download of high-grade serous ovarian cancer (HGSOC) transcriptomic data . . . 53
4.2 Generation of dot plots, violin plots and scatter plots for the correlation analysis . . . 53
4.3 Statistical correlation analysis and p-value calculation . . . 53

IV Results . . . 54

1 Chapter 1. Technical optimisation of oviduct single-cell isolation . . . 54
1.1 Introduction . . . 54
1.2 Enzymatic dissociation followed by FACS enrichment performs efficiently for the population-balanced isolation of viable oviduct single cells . . . 58
1.3 Multiple consecutive rounds of single-cell dissociation prove to be the best strategy to maximise the number of isolated epithelial cells . . . 63
1.4 A confounding stress response dominated by the expression of immediate early genes is detected in oviduct cells after enzymatic single-cell dissociation . . . 65
1.5 Conclusion . . . 68
2 Chapter 2. Single-cell transcriptomic atlas of the mouse oviducts . . . 69
2.1 Introduction . . . 69
2.2 Single-cell data analysis setup and workflow . . . 72
2.3 Single-cell RNA-Seq of the mouse oviduct faithfully captures the main cellular lineages previously reported in this organ . . . 76
2.4 scRNA-seq analysis of the oviduct epithelium repertoire reveals diverse secretory and ciliated subtypes marked by distinctive biological processes . . . 79

2.5 Low or absent gene and protein expression of the major histocompatibility complex (MHC) Type I is identified in oviduct epithelial cells 86

2.6 In-depth characterisation of the four distinct secretory subpopulations, the two ciliated subpopulations and the two intermediate secretory-ciliated clusters 89

2.7 An alternative single-nucleus RNA sequencing method validates and improves epithelial resolution and bypasses the stress response and the need for fresh samples 97

2.8 Further validation is accomplished by 3D organoid culture, immunohistochemistry and fluorescence microscopy 101

2.9 The mucosal stroma is mainly comprised of mesenchymal, vascular, smooth-muscle and mesothelial cell populations 103

2.10 The single-cell immune landscape in the adult murine oviducts reveals a heterogeneous population of macrophages, dendritic cells, T and B lymphocytes and natural killer (NK) cells . . . 106

2.11 Conclusion 110

3 Chapter 3. Cross-species comparison of cell populations and interactions in the oviduct and fallopian tube 113

3.1 Introduction 113

3.2 The physiological single-cell transcriptomic landscape of the human fallopian tubes 114

3.3 Inter-species comparisons of the oviduct and fallopian tube cell types 117

3.4 Characterisation of the complex cell-to-cell interactions network in the oviduct and fallopian tube 120

3.5 Conclusion 124

4 Chapter 4. scRNA-seq, microarray and bulk RNA-Seq analyses discover a TGFβ signalling signature conserved in oviducts, fallopian tubes and high-grade serous ovarian cancer 125
4.1 Introduction 125
4.2 TGFβ interactions are significantly enriched in oviduct and fallopian tube homeostasis 126
4.3 TGFβ1 and TGFβ3 are differentially expressed in EMT-high HGSOCs and secreted by immune and mesenchymal cells, respectively, in the normal fallopian tube and oviduct 131
4.4 The expression of TGFβ ligands positively correlates with the expression of markers of their proposed secreting cell types in HGSOCs and the same cell-type markers show higher expression in EMT-enriched tumours 133
4.5 Conclusion 137

V Conclusion and future directions 139
1 Overview and discussion 139
2 Conclusion 141
3 Future directions 142

VI Bibliography 143

List of Acronyms

AOCS australian ovarian cancer study

AR acrosome reaction

BM basal membrane

BMP bone morphogenetic protein

BSA bovine serum albumin

CBF ciliary beat frequency

CBRG computational biology research group

cDNA complementary deoxyribonucleic acid

COC cumulus-oocyte complex

CG chorionic gonadotropin

DE differential expression or differentially expressed

DEG differentially expressed gene

DNA deoxyribonucleic acid

DPBS dulbecco's phosphate-buffered saline

dsDNA double-stranded deoxyribonucleic acid

ECM extracellular matrix

EGF epidermal growth factor

EMT epithelial-mesenchymal transition

ERα oestrogen receptor α

ERK extracellular signal-related kinase

FACS fluorescence-activated cell sorting

FGF fibroblast growth factor

FSC forward scatter

FRT female reproductive tube

FSH follicle-stimulating hormone

GEO gene expression omnibus

GO gene ontology

HGSOC high-grade serous ovarian cancer

H&E hematoxilin and eosin

IEGs immediate early genes

IF immunofluorescence

IFN interferon

IGF insulin-like growth factor

IHC immunohistochemistry

ISG interferon-stimulated gene

LH luteinizing hormone

LPS lipopolysaccharides

MACS magnetic bead-based cell selection

MAPK mitogen-activated protein kinase

MHC major histocompatibility complex

MDA müllerian duct anomaly

mRNA messenger ribonucleic acid

MRKH mayer-rokitansky-küster-hauser

miRNA micro ribonucleic acid

NA not applicable

NCBI national center for biotechnology information

NGS next-generation sequencing

NK natural killer

NSB nuclei suspension buffer

PAMP pathogen-associated molecular pattern

PCA principal component analysis

PCP planar cell polarity

PCR polymerase chain reaction

PFA paraformaldehyde

PGE2 prostaglandin E2

PRR pattern-recognition receptor

PTS particle transport speed

qPCR quantitative polymerase chain reaction

RA retinoic acid

RBC red blood cell

RNA ribonucleic acid

RNA-Seq RNA sequencing

RT-PCR reverse transcription polymerase chain reaction

scRNA-seq single-cell RNA sequencing

SD standard deviation

snRNA-seq single-nucleus RNA sequencing

SM smooth muscle

SSC side scatter

ssDNA single-stranded DNA

TCGA the cancer genome atlas

TCR t-cell receptor

TF transcription factor

TGFβ transforming growth factor-β

TLR toll-like receptor

tSNE t-distributed stochastic neighbor embedding

UMAP uniform manifold approximation and projection

UMI unique molecular identifier

List of Figures

1 Female reproductive tube (FRT) morphology in murine and human species . . . 21
2 Müllerian duct development and differentiation in mouse . . 23
3 Müllerian duct development and FTR formation in humans . 26
4 Histological characterisation of the newborn murine oviducts 29
5 Histoanatomical characterisation of murine oviducts 31
6 The hormone concentration during the oestrus cycle of mice 34
7 Distribution of ciliated epithelial cells in the oviducts 34
8 Single-cell RNA sequencing experimental workflow with mouse oviducts or human fallopian tubes 55
9 Cell-cell and cell-matrix connections in the oviduct epithelium 57
10 Gating strategy for single-cell FACS sorting of mouse oviducts 61
11 Characterisation of the epithelial loss during oviduct dissociation and FACS sorting 62
12 A serial dissociation strategy to enhance epithelial recovery during oviduct single-cell harvesting 64
13 Detection of a stress response in oviduct cells following Liberase enzymatic dissociation 66
14 Evaluation of the alternative use of a cold-active protease for enzymatic dissociation of the oviducts 67
15 scRNA-seq 10X Chromium protocol 72
16 Bioinformatic analytical pipeline for the analysis of scRNA-seq oviduct data 74
17 Impact of the filtering strategy on data quality 75
18 Cell-lineage heterogeneity in the mouse oviduct 78
19 Cell-lineage genetic markers of the oviduct 79
20 Single-cell characterisation of the oviduct epithelium 81
21 The immune transcriptomic signature of the oviduct epithelium 85
22 The oviduct gene expression of the MHC Type I components 87
23 Gating strategy for the flow-cytometry analysis of oviduct epithelial cells expressing the MHC Type I complex 88
24 Differential gene expression profiles of the discrete oviduct epithelial clusters 91
25 Expression of ciliated and secretory genetic markers in the oviduct epithelial clusters 93
26 Gene expression of the molecular markers characterising the intermediate cluster Cil_Mapk10 95
27 Gene expression of hormone receptors in the oviduct epithelial subclusters 96
28 Distribution of read types per sample in each transcriptomic sequencing technique 98
29 Single-nucleus RNA-Seq characterisation of the oviduct cell-lineage abundances, validation of the scRNA-seq epithelial identities and absence of stress response 100
30 Immunofluorescence of oviduct epithelial organoids 102

31 **Validation of the Fxyd4_Dach1 epithelial secretory subcluster by IHC** . . . 103
32 **Single-cell RNA-Seq characterisation of the oviduct mesenchymal lineage** . . . 105
33 **Identification of the two oviduct endothelial cell clusters** . . . 106
34 **Diagram of the detected oviduct immune cell repertoire** . . . 107
35 **Identification and further classification of the oviduct immune repertoire** . . . 108
36 **Analysis of the tissue-resident phenotype of oviduct macrophages** . . . 110
37 **External and internal fallopian tube scRNA-seq data contribution** . . . 114
38 **Pearson linear correlation tests between the transcriptomes of individual cell clusters** . . . 118
39 **Human-mouse cluster Spearman's correlation matrix based on the expression of the 50 most highly-variable genes** . . . 120
40 **Heatmap showing the total number of interactions between oviduct cell types and their associated lineages** . . . 122
41 **Heatmap showing the total number of interactions between fallopian tube cell types and their associated lineages** . . . 123
42 **Overview of statistically-significant TGFβ ligand-receptor interactions using CellPhoneDB on the oviduct scRNA-seq dataset** . . . 129
43 **Overview of statistically-significant TGFβ ligand-receptor interactions using CellPhoneDB on the fallopian tube scRNA-seq dataset** . . . 130
44 **Differential expression of TGFβ ligands in EMT-high versus EMT-low ovarian tumours** . . . 131
45 ***TGFβ1* and *TGFβ3* expression in the oviduct and the fallopian tube** . . . 132
46 **The positive correlation between the expression of *TGFβ1* and immune gene markers and *TGFβ3* and mesenchymal gene markers in ovarian tumours** . . . 134
47 **The positive correlation between the expression of shared murine and human macrophage gene markers and *TGFβ1* in ovarian tumours** . . . 135
48 **Differential expression of immune and mesenchymal gene markers in EMT-high versus EMT-low ovarian tumours** . . . 136

List of Tables

1 **Embryonic molecular markers of the Müllerian ducts.** 24
2 **The thermal cycler programme used for the cDNA reverse transcription** 47
3 **The thermal cycler programme used for the qPCR of immediate early genes (IEGs)** 47
4 **Technical metrics per single-cell RNA sample from adult mouse oviducts** 75
5 **Technical metrics per single-nucleus RNA sample from adult mouse oviducts** 97

I Introduction

The scope of this thesis encompasses multiple interdisciplinary areas of investigation focused on the mouse oviduct or fallopian tubes if referencing the human counterparts. Therefore, before embarking on the main body of this project, it is important to provide a broad and detailed introduction to understand the biology of this organ. The extent of this work is restricted to the murine and human species although others are mentioned when comparisons are required and when significant discoveries in the field have been achieved by studying species differing from the human and murine ones. To clarify and set a consensus on terminology for this project, the nomenclatures oviduct and uterine horn are applied to all species, with the exception of humans, where they are known as fallopian tubes and uterus, respectively.

The oviducts are tubular anatomical structures physically connecting the ovaries with the uterine horns in the female reproductive tract (Figure 1, pg 21). Nevertheless, the oviduct function goes beyond a mere bridging role due to their essential contribution to amniotes' reproduction and female health, which have earned them the symbolic title of "conduits of life". The first sections of this introduction chronologically focus on the embryonic genesis of these structures as the early Müllerian ducts from multipotent stem cells and how the Müllerian ducts transition into mature functional reproductive organs. Subsequently, this introduction dives into the key role of the adult oviducts during homeostasis, reproduction and female fertility. Lastly, I present a description of the different pathologies that can arise from the disturbance of the oviduct development and homeostasis, the latest scientific strategies developed to further expand our knowledge about the biology and dysfunction of this organ and a discussion about the role and relevance of this project within the broader research framework.

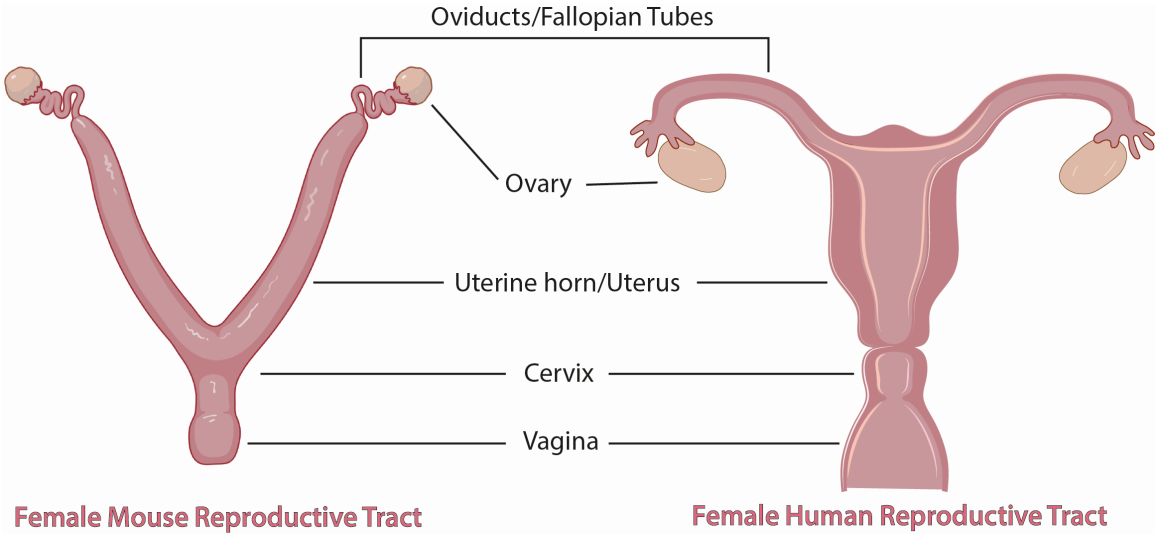


Figure 1. Female reproductive tube (FRT) morphology in murine and human species
The human female reproductive tract is represented on the right and the murine on the left.

1 Embryonic development of the mouse and human Müllerian ducts

The mouse oviducts and the human fallopian tubes share an analogous embryonic development starting as two rudimentary Müllerian ducts, one on the left and another on the right body side (for an extensive review of this process see: Mullen and Behringer 2014, Roly et al. 2018, Cunha et al. 2018, Zhao and Yao 2019, Santana Gonzalez et al. 2021). These anatomical precursors are the anlagen of not only the oviducts but also the uterine horns, the cervix and the two upper thirds of the vagina, named in a cranial-caudal axis order. When cranial coordinates are mentioned, they refer to the female reproductive tube (FRT) location in close proximity to the ovary whereas the caudal side represents the vaginal end. During the process of Müllerian duct development three distinct stages can be recognised: (1) progenitors' differentiation, (2) progenitors' invagination and (3) caudal tubal elongation. These events occur at species-specific times such as the embryonic days 11.75-13.5 in mouse (Kobayashi et al. 2004, Guioli et al. 2007, Orvis and Behringer 2007), Carnegie stages 18-23 in human (Hashimoto 2003) and Hamburger-Hamilton stages 16-30/32 in chicken (Guioli et al. 2007, Atsuta and Takahashi 2016) (Figure 2A, B, C, pg 23).

1.1 Progenitors' differentiation

The beginning of this organogenetic process is first detected at the progenitors' differentiation stage and results into an epithelial thickening known as placode or ridge (Jacob et al. 1999, Atsuta and Takahashi 2016, Guioli et al. 2007, Orvis and Behringer 2007, Dohr and Tarmann 1984, Gruenwald 1941). This structure is a proliferative stem cell reservoir and locates on the cranial part of the urogenital mesonephros, one of the three proposed embryonic kidneys. Concretely, this reservoir is found in the coelomic epithelium, the outermost layer surrounding the mesonephros, from which it buds out forming the Müllerian duct towards the mesonephric caudal axis. Curiously, both female and male (Wolffian ducts) precursor reproductive organs concomitantly develop in the embryo regardless of the genetic sex and it is not until later stages that this situation is resolved by regressing the gender-unmatching tubes and promoting the differentiation of the sex-complementary tube set (Figure 2D, pg 23).

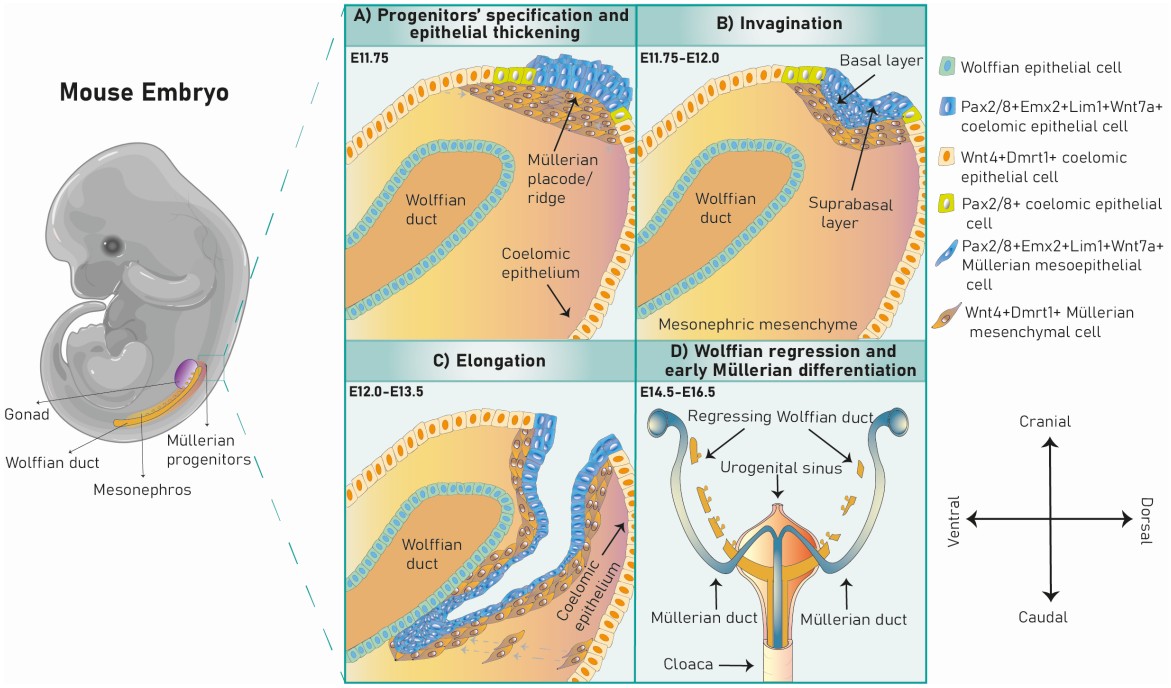


Figure 2. Müllerian duct development and differentiation in mouse
The highly-conserved Müllerian developmental process is depicted in the mouse. (A) This organogenesis starts with the specification of coelomic epithelial progenitors which proliferate creating the Müllerian ridge. (B) The specified progenitors then invaginate into the mesonephros until they establish physical contact with the Wolffian duct. (C) From this point onwards, the tubes undertake a caudal elongation step to expand their length. (D) In female embryos, the permanence and differentiation of the Müllerian ducts is promoted, whereas the male Wolffian ducts fully regress.

At the time of Müllerian initiation, the mesonephros already hosts the Wolffian ducts, which have been shown to play a paramount role during all stages of the female duct growth, except for this early progenitors' differentiation. During this early stage, mesonephric coelomic epithelial stem cells commit to a Müllerian fate and proliferate into the placode driven by bone morphogenetic protein (BMP) signalling acting through SMAD proteins and an unknown BMP effector. This was proven by studies using the soluble protein noggin, a natural BMP inhibitor (Atsuta and Takahashi 2016), and bulk RNA sequencing (Roly et al. 2020a) in the coelomic epithelium of chicken embryos at the time of the Müllerian progenitors' specification. The use of the BMP inhibitor blocked the placode formation, whereas the analysis of the RNA sequencing data showed the expression of many genes belonging to the BMP pathway. Among the mesonephric coelomic Müllerian progenitors, a subset is characterised by the expression of conserved urogenital progenitor markers such as *Pax2*, *Emx2*, *Lim1*, *Pax8*, *Pbx1*, *Hnf1b*, *Dach1*, *Dach2* and *Wnt7a*, while an alternative subset expresses *Wnt4* and *Dmrt1* (Table 1, pg 24) (Kobayashi et al. 2004, Guioli et al. 2007, Orvis and Behringer 2007, Ayers et al. 2015, Jacob et al. 1999, Atsuta and Takahashi 2016, Prunskaite-Hyyryläinen et al. 2016, Vainio et al. 1999, Schnabel et al. 2003, Lokmane et al. 2010, Davis et al. 2008).

Gene Family	Gene name	Symbol	Protein type
Paired box (Pax)	Paired box gene 2	Pax2	Transcription factor (TF)
	Paired box gene 8	Pax8	TF
WNT	Wnt family member 4	Wnt4	Secreted signalling protein
	Wnt family member 7a	Wnt7a	Secreted signalling protein
Dach/dachshund	Dachshund family transcription factor 1 and 2	Dach1 and Dach2	Transcriptional co-factors
(Doublesex/MAB-3) Domain	Doublesex and mab-3 related transcription factor 1	Dmrt1	TF
Homeobox	LIM homeobox 1	Lim1/Lhx1	TF
	Empty spiracles homeobox 2	Emx2	TF
	Pre-B-Cell leukemia transcription factor 1	Pbx1	TF
	Hepatocyte nuclear factor-1 beta	Hnf1b	TF

Table 1. Embryonic molecular markers of the Müllerian ducts.

1.2 Progenitors' invagination

Both classes of Müllerian progenitors invaginate to create the Müllerian tube, with the first class (expressing urogenital progenitor markers) giving rise to the epithelial lineage forming the tubal lumen, and the second one dedicated to the mesenchymal lineage forming the tubal stroma. Invagination takes place until the tip of the new growing tube reaches the Wolffian ducts and Müllerian cells establish physical interactions with Wolffian cells (Gruenwald 1941, Guioli et al. 2007, Orvis and Behringer 2007, Hashimoto 2003, Kobayashi et al. 2004, Jacob et al. 1999, Dohr and Tarmann 1984). The described cellular process by which these Müllerian progenitors are able to detach from the coelomic epithelium and invaginate into the underlying mesonephric space is the epithelial-mesenchymal transition (EMT). One of the essential steps for the appropriate execution of this EMT is the breakage of the coelomic epithelial basal membrane (BM), an event which has been demonstrated to happen before the invagination of Müllerian progenitor cells and to be controlled by the FGF (fibroblast growth factor)/ERK (extracellular signal-related kinase)/MAPK (mitogen-activated protein kinase) pathway. This was shown in chicken embryos where downregulation of the *Fgfr2* receptor and its downstream effector protein *Rac1*, prevented BM disruption and invagination of Müllerian progenitors (Atsuta and Takahashi 2016). However, this study was not able to identify the ligand triggering the pathway activation. Furthermore, the expression of an array of additional FGF-regulated genes was also detected in the same model during the invagination stage (Roly et al. 2020a) and, interestingly, the FGF pathway seemed to interact and cross-talk with the BMP pathway at these early stages too (Atsuta and Takahashi 2016). Moreover, BMP signalling controls the expression of the transcription factor *Pax2* in Müllerian epithelial progenitors and Pax2 itself, as well as the FGF pathway, has been shown to control the expression of *Lim1* in

the same stem cell subset (Atsuta and Takahashi 2016). Expression of the transcription factor Lim1 is essential for the invagination step as the loss of this gene in mouse or its deficiency in chicken leads to the same invagination-defective phenotype (Kobayashi et al. 2004, Atsuta and Takahashi 2016, Orvis and Behringer 2007). Nevertheless, the extent of Pax2 control over the expression of *Lim1* has been questioned given the divergent phenotypes of different lack-of-function models. In chicken embryos, *Pax2* downregulation highly reduces *Lim1* expression and blocks Müllerian duct invagination (Atsuta and Takahashi 2016). However, in transgenic mice lacking a functional Pax2, *Lim1* expression is unaffected and invagination is still carried out by Müllerian progenitors although lack of Wolffian ducts in this model eventually impedes Müllerian elongation (Kobayashi et al 2004, Torres et al. 1995). This dichotomy could be explained by the presence of *Pax8*, a gene missing in the chicken genome but present in the murine one, which could act redundantly and compensate for the *Pax2* loss in mouse. Functional redundancy among these transcription factor family members has previously been demonstrated in *Xenopus* and mouse in the embryonic kidney, an organ that is developmentally very similar to the reproductive ones (Bouchard et al. 2002, Carroll and Vize 1999). Loss of other early Müllerian stem cell markers such as *Hnf1b* and *Emx2* presents a similar mutant phenotype of Müllerian agenesis, however, further molecular mechanisms behind this phenotype are not reported and would warrant additional investigation (Lokmane et al. 2010, Miyamoto et al. 1997). Mouse lines with individual homozygous and double heterozygous mutations in the Müllerian embryonic markers Dach1/2 present normal Müllerian ducts, however, in Dach2$^{-/-}$;Dach1$^{+/-}$ and Dach2$^{+/-}$;Dach1$^{-/-}$ animals the expression of *Lim1* and *Wnt7a* is severely reduced and Müllerian hypoplasia is observed, which suggests that these cofactors regulate these genes during Müllerian development (Davis et al. 2008). Moreover, *Wnt4*, a member of the WNT family expressed by mesonephric and Müllerian mesenchymal cells, is also required for Müllerian invagination as both knock-out and knock-down models show important defects during Müllerian invagination but not progenitors' specification (Vainio et al. 1999, Prunskaite-Hyyryläinen et al. 2016, Guioli et al. 2007). Accordingly, another Müllerian mesenchymal-specific mouse model bearing *Wnt4* and *Wnt5a* knock-outs has been reported to display Müllerian abnormalities (St-Jean et al. 2019).

1.3 Müllerian caudal elongation

The last step of Müllerian development is the caudal elongation of the tube along the cranial-caudal axis of the mesonephros. During this growth, Müllerian cells proliferate and maintain a close interaction with the Wolffian ducts, which serve as guides and signalling sources but do not contribute with cells to the female tube (Guioli et al. 2007, Orvis and Behringer 2007, Hashimoto 2003, Gruenwald 1941, Dohr and Tarmann 1984, Kobayashi et al. 2005, Carroll et al. 2005). Wolffian-expressed genes such as *Wnt9b* and *Gata3* have been found to be essential for Müllerian elongation. Gata3 is a transcription factor whose loss in mouse halts Müllerian elongation (Grote et al. 2008). Wnt9b is a glycoprotein secreted by the male duct, which influences Müllerian growth through the activation of the canonical WNT pathway (Carroll et al. 2005). In a mouse model lacking this gene, Müllerian development is successful until the elongation phase, where it is blocked (Carroll et al. 2005). Additionally,

it has been found that mice knockout for *Lim1* in the Wolffian ducts exhibit an important decrease in *Wnt9b* expression and the same elongation-arrest phenotype, which points at this transcription factor as a possible Wnt9b regulator (Pedersen et al. 2005). Other Wolffian-independent factors controlling the elongation stage are the Müllerian-expressed G protein receptor Gpr56 (Roly et al. 2020b), Wnt4, coelomic epithelium cues (Prunskaite-Hyyryläinen et al. 2016) and Dmrt1 (Ayers et al. 2015). Moreover, mice with double homozygous mutations for the retinoic acid (RA) receptors *Rara* and *Rarβ2* suffer from elongation arrest, even in the presence of Wolffian ducts (Mendelsohn et al. 1994) and double mutants for the receptors *Rara*/*Rarγ* and *Rxra*/*Rara* also lack Müllerian ducts (Mendelsohn et al. 1994, Kastner et al. 1997). Moreover, deletion of the *Lats1* and *Lats2* kinases genes also leads to developmental aberrations in the Müllerian ducts such as atypical coiling and to cysts and dilatations in the adult oviducts (St-Jean et al. 2019).

Once the Müllerian elongation is completed, the fully formed Müllerian ducts fuse with each other and with the urogenital sinus. This process is dependent on the membrane protein Lhfpl2 (LHFPL Tetraspan Subfamily Member 2), whose inactivation leads to Müllerian ducts unfused with the urogenital sinus (Zhao et al. 2016). In humans, Müllerian ducts not only fuse with each other and the urogenital sinus but they also undertake two additional steps: (1) midline fusion and (2) central septal resorption to create the human uterine cavity, in contraposition to the bicornuate uterus in mouse (Figure 3, pg 26). This developmental divergence is what causes the pronounced anatomical gross difference between the murine and human FRT. Furthermore, the murine oviducts present a coiled structure, whereas the human fallopian tubes present a curved tubal shape. The oviduct and FRT morphology also differs in other species due to developmental reasons, but the organs are otherwise extremely similar.

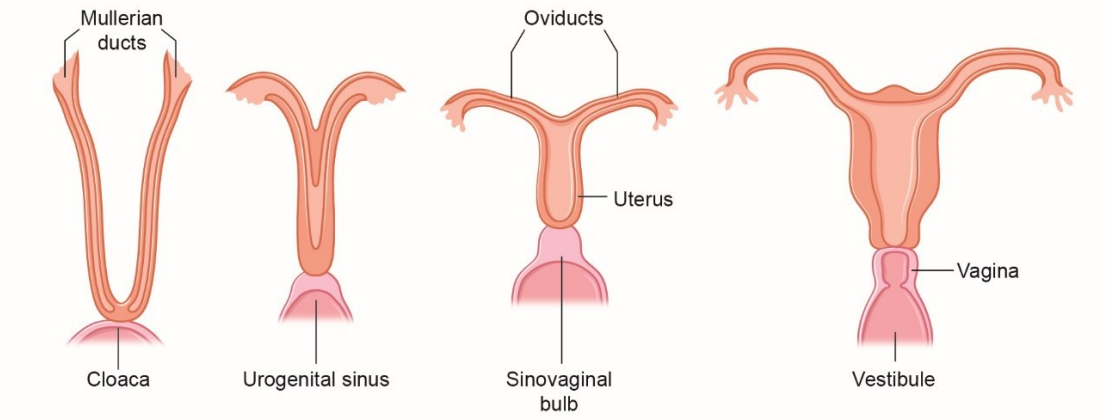


Figure 3. Müllerian duct development and FTR formation in humans
The human-specific development and differentiation of the Müllerian ducts is represented. First, the Müllerian ducts develop as individual tubes that caudally fuse with each other and the urogenital sinus. Second, they fuse with each other at the midline and the formed central septum undertakes resorption to allow the formation of the uterine cavity.

2 Müllerian duct differentiation

Following the development of the Müllerian ducts, the male Wolffian ducts disappear in genetically female embryos and further development and differentiation of the female ducts is promoted. These events are known as sex determination. The Wolffian regressive process has been shown to depend on the transcription factor COUP-TFII (also known as Nr2f2), expressed by Wolffian mesenchymal cells (Zhao et al. 2017). Subsequently, the immature but fully developed Müllerian ducts switch on several transcriptional programmes to differentiate into the mature female reproductive organs. This priming can be first observed with the segment-specific activation of genes belonging to the Hox transcriptional cluster and the RA signalling pathway in the Müllerian mesenchyme (Taylor et al. 1997, Ma et al. 1998, Nakajima et al. 2016, 2019). The Hox genes *Hoxa9*, *10*, *11* and *13* become expressed around E14.5-15.5 in mouse (Taylor et al. 1997, Ma et al. 1998, Nakajima et al. 2016) and are spatially restricted to the cranial (*Hoxa9*), middle (*Hoxa10/11*) and caudal (*Hoxa13*) areas. Their paralogues *Hoxd9*, *10*, *11* and *13* also follow this peculiar cranial-caudal expression pattern (Dollé et al. 1991). Each marked segment will respectively differentiate into the adult oviducts, uterine horns and cervix/upper vagina. Curiously, this expression pattern is conserved during adulthood (Taylor et al. 1997, Ma et al. 1998) and across species, including humans. Depletion of the *Hoxa10* and *Hoxa11* function in independent homozygous knock-out mouse models shows an anteriorisation of the cranial first quarter of the uterine horns into oviducts (Satokata et al. 1995, Benson et al. 1996, Gendron et al. 1997), revealing a putative patterning role for these transcriptional factors. Intriguingly, under *Hoxa9*-null conditions no mutant reproductive phenotype has been reported in mouse (Raines et al. 2013, Fromental-Ramain et al. 1996).

Regarding the RA pathway, RA signalling displays a decreasing gradient at this time, being strongly activated in the cranial part, which develops into the oviducts, and inhibited in the lowest sections, the prospective upper vagina (Nakajima et al. 2016). The detected expression of the enzyme *Cyp26a1* (Cytochrome P450 Family 26 Subfamily A Member 1) in the caudal Müllerian end has been proposed to be involved in this inhibition as it acts as an RA level regulator by metabolising RA in many tissues (Nakajima et al. 2019). In the upper Müllerian section, the RA-activation time window lasts between E14.5-P0 in mouse (Nakajima et al. 2016). At E14.5, it seems that a partial fate decision has already been reached as cultures of proximal and mid-duct portions differentiate into epithelial cells resembling the oviduct and the uterine epithelium respectively (Nakajima et al. 2016). However, treatment of cranial Müllerian portions with an RA-signalling inhibitor induced differentiation into $Trp63^{+}$ vagina-like epithelia, a phenotype that could be rescued by RA addition (Nakajima et al. 2016). The importance of this developmental stage is further highlighted by the gradual acquisition of the expression of the epithelial differentiation marker E-cadherin (*Cdh1*) and apicobasal polarity in the Müllerian epithelium (Orvis and Behringer 2007, Stewart et al. 2013, Ghosh et al. 2017, Prunskaite-Hyyryläinen et al. 2016).

Moreover, several molecular drivers of Müllerian differentiation have been identified through lack-of-function research strategies. In this context, absence of the proteins Wnt4, Dach1/2 and Wnt7a were found to induce mutant reproductive phe-

notypes such as hypoplastic oviducts for the first two and uncoiled short oviducts displaying posteriorisation for the latter (Vainio et al. 1999, Miller et al. 1998, Prunskaite-Hyyryläinen et al. 2016, Parr and Mcmahon 1998, Davis et al. 2008). One proposed function of Wnt7a in the oviducts that could potentially explain this phenotype is related to the maintenance of the epithelial morphology: *Wnt7a* expression seems to be influenced by VANGL Planar Cell Polarity Protein 2 (Vangl2), a member of the planar cell polarity (PCP) pathway which has been shown to control the epithelial morphology and polarisation in the oviducts postnatally (explained in the following paragraph) (Vandenberg and Sassoon 2009). The *Wnt7a*-null phenotype is replicated when *Ctnnb1* is knocked out in Müllerian mesenchymal cells, whereas stabilisation of the Ctnnb1 protein impairs E-cadherin expression in the epithelium, also yielding aberrant oviducts (Deutscher and Humphrey 2007, Stewart et al. 2013).

Taken together,this evidence indicates that tight regulation of the canonical Wnt pathway seems key for Müllerian differentiation and that Wnt7a could act as a signalling molecule in this context. Furthermore, the loss of the TGFβ receptor 1 (*Tgfbr1*) in Müllerian mesenchymal cells triggers smooth muscle (SM)-related oviduct aberrations such as cysts (Li et al. 2011). Similarly, mice lacking the RNase Dicer1 in Müllerian mesenchymal cells present short oviducts with cysts and lack of epithelial folds, which impair the reproductive function (Hong et al. 2008, Gonzalez and Behringer 2009).

3 Oviduct formation

The presented differentiation events are executed by the Müllerian ducts prenatally, however, it is important to highlight that a second round of substantial differentiation also occurs during the first two weeks of the postnatal life and it is at this vital point when the upper Müllerian segment noticeably differentiates into the oviducts. This is achieved through the activation of transcriptional programmes that (1) distinguish each section-specific cell types with a unique signature and (2) orchestrate the important morphological remodelling occurring at this time.

In this time window the oviducts acquire their characteristic mature protruded luminal morphology, their pseudostratified epithelium and the specialised surrounding histological layers. As previously mentioned, the Wnt noncanonical PCP pathway has been shown to be involved in these processes through the maintenance of the organ morphology and the establishment of the oviduct epithelial polarity from P2 to adulthood (Shi et al. 2014). In this study, the molecule Cadherin EGF LAG Seven-Pass G-Type Receptor 1 (*Celsr1*) was identified as a crucial molecular driver underpinning epithelial polarity in the oviduct. However, despite the high degree of differentiation established at this time, it is not until week 3 or 4 post-birth that the oviduct epithelium has been demonstrated to fully commit to a fate using electron microscopy and heterotypic recombination experiments (Yamanouchi et al. 2010, Komatsu and Fujita 1978). These experiments also demonstrated a prominent role of the underlying oviduct mesenchyme in specifying the oviduct epithelial identity and the same assays, carried out in different organs of the young FRT, have highlighted this as a shared circumstance in the entire reproductive system (Cunha 1976, Kurita 2001).

The acquisition of the oviduct epithelial identity and the main differentiation processes determining this are also established postnatally. Immediately after birth, all oviduct cells are primed with the expression of the embryonic marker *Pax8* and display a homogeneous epithelial columnar morphology ((Figure 4, pg 29), Ghosh et al. 2017, 2020). These cells can act as unipotent stem cells giving rise to daughter cells with either the same stem fate through self-renewal or to a more committed epithelial progeny such as a *Pax8*$^{+}$ lineage, known as secretory cells, and a *Pax8*$^{-}$ lineage known as ciliated cells by P3-P4 (Ghosh et al. 2017, Shi et al. 2014).

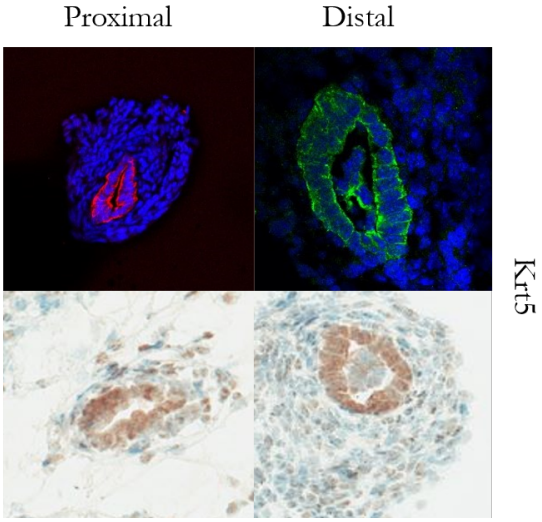


Figure 4. Histological characterisation of the newborn murine oviducts
Top row, the cranial (left) and caudal (right) oviducts of mouse newborns are stained with a keratin marker (Krt5) and immunofluorescence images are presented (nuclei stained in blue by DAPI and Krt5 in red or green). Bottom row, proximal and distal sections are stained with the same marker by immunohistochemistry.

The epidermal growth factor (EGF) and Notch molecular pathways have been reported to regulate these processes, with the first activating the second. Subsequently, the activated Notch pathway would regulate the maintenance of the secretory subtype in the oviduct epithelium. Evidence for this is provided by the increased ciliogenesis upon Notch signalling inhibition, leading to an abnormal number of ciliated cells (Zhu et al. 2019). Furthermore, interruption of the Notch pathway in another study reduced stem cell proliferation and the expression of their marker genes while promoting ciliogenesis (Kessler et al. 2015). Tight control of these differentiation events is essential for the maintenance of the right epithelial cell identity and frequency at a given time, as secretory and ciliated cells display different functions in the oviduct.

Ciliated cells are cuboid cells with hair-like motile structures called cilia located at their apical side and they aid in establishing the oviduct fluid flow (Komatsu and Fujita 1978). The rhythmic movement of the epithelial cilia is known as the oviduct ciliary beat frequency (CBF), and the fluid flow generated within the oviducts can be measured by its particle transport speed (PTS). Secretory cells do not present motile cilia but microvilli, and contain granules in their cytoplasm, which are released into the extracellular lumen in the oviducts contributing to the

mucous fluid (Komatsu and Fujita 1978). This oviduct fluid is of dual origin given that it is composed by factors secreted by secretory epithelial cells and also blood plasma transudate. This fluid is complex and it contains carbohydrates, ions, lipids, proteins, immune cells and extracellular vesicles, which serve as vehicles for intercellular communication by enclosing molecules such as proteins, mRNAs, miRNAs and lipids (Leese et al. 2008, Buhi et al. 2000). Some of these molecules are also metabolic factors such as lactate, pyruvate, glucose and aminoacids (Leese et al. 2008, Leese 1988, Hugentobler et al. 2007). Interestingly, only a small percentage of the fluid proteins seems to be secreted by the oviduct in pigs, which suggests that the oviducts themselves do not synthetise the majority of the proteins in the fluid and that these could come from the plasma instead (Acuna et al. 2017).

Although it is known that the function of the epithelial subtypes is not limited to the establishment of the oviduct fluid and its flow, investigation of further specific roles for the epithelial subtypes is currently lacking. Nevertheless, research has been able to identify distinct molecular features among these subtypes, such as the differential expression of the cytoskeletal protein tubulin beta-4 chain (*Tubb4a* and *Tubb4b*) and the transcription factor Forkhead Box J1 (*Foxj1*) (Blatt et al. 1999, Chen et al. 1998) by ciliated cells and the expression of the embryonic transcription factor *Pax8* (Perets et al. 2013) and the oviductal glycoprotein 1 (*Ovgp1*) by secretory cells. Although the oviduct secretory type is the only in the body expressing *Ovgp1*, this gene seems dispensable for fertilisation, as mice with a null *Ovgp1* mutation are fertile (Araki et al. 2003). However, a role in protein phosphorylation cascades necessary for sperm capacitation and polyspermy control through regulation of the zona pellucida has been suggested for this protein (Saccary et al. 2013, Zhao et al. 2013, Coy et al. 2008, Algarra et al. 2016). Recently, more advanced research by Hu et al. 2020 has been able to refine the current epithelial subdivision through the identification of four secretory and one intermediate secretory-ciliated subtypes in adult human fallopian tubes. Another study identified three secretory, four ciliated and three epithelial undetermined clusters (Dinh et al. 2021) and, despite the lack of consensus on the number of subtypes reported by these studies, they both highlighted the existence of heterogeneity within the two epithelial subtypes.

Histologically, the adult oviducts are composed of a mucosa, a muscularis and a serosa layer, also known as endosalpinx, myosalpinx and mesosalpinx respectively. The mucosa is organised into folds known as "rugae" or "plicae" and it contains a pseudostratified epithelium lining a basement membrane and a subjacent lamina propria of connective tissue. The muscularis layer is comprised of an inner circular layer and an external longitudinal layer. The serosa layer is made of an external mesothelium and an inner sheet of interstitial connective tissue. The adult luminal epithelium of the mucosa carries out the organ reproductive function and it is referred as the oviduct parenchyma, whereas the other supportive structures (lamina propria, blood vessels, etc) are named as oviduct stroma. During postnatal differentiation, the oviduct becomes histologically divided into four highly-conserved sections across species: the isthmus, ampulla, infundibulum and fimbriae (Figure 5, pg 31). The fimbriae are the most cranial and the isthmus is the most caudal section. The latter connects with the uterine horns through the uterotubal junction and it is the only oviduct portion that possesses an inner longitudinal muscle layer.

It is followed by the ampulla and the infundibulum, which is the final part of the oviducts. The infundibulum is open to the peritoneum via the ostium tubae and ends with the outer-protruded fimbriae that are in close contact with the ovary.

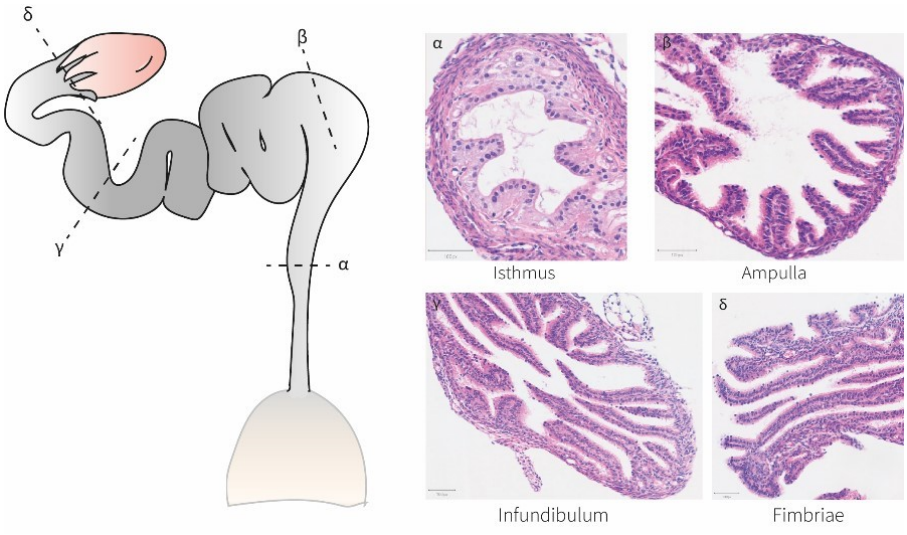


Figure 5. Histoanatomical characterisation of murine oviducts
Left, schematic representation of the murine oviduct. Right, representative histological images of the four oviduct sections (H&E staining). Figure taken from Santana Gonzalez et al. 2021 and authored by Laura Santana Gonzalez.

The mucosa/muscularis layer proportion, the associated folding degree (measured as the frequency and length of folds) and the luminal diameter decreases along the craniocaudal axis, from the fimbriae to the isthmus. The presence of ciliated cells also decreases along this axis, being very abundant in the terminal parts such as fimbria, infundibulum and the ampulla, while the isthmus epithelium is almost exclusively composed of secretory cells. The epithelial layer is situated on top of a stromal layer with a basement membrane acting as a bridge between the two domains. The oviduct stroma contains mesenchymal cells, immune cells and cells composing the blood and lymphatic vasculature. The oviduct blood supply is supported by the ovarian and uterine arteries and veins that branch towards the oviducts and infiltrate them. The way the vasculature is organised is individual-dependant but, commonly, a branch of the ovarian artery supplies blood to the fimbriae, infundibulum, ampulla and distal isthmus, while the proximal isthmus and uterotubal junction are supplied by a branch coming off the uterine artery. Both arterial systems are in contact, therefore, it is possible that the ovary and uterus can transfer molecules to the whole oviduct via the circulatory system, establishing a paracrine communication (Eddy and Pauerstein 1980, Leese 1988). Similarly, the lymphatic system, which is in charge of protein removal and interstitial fluid draining, also supports the oviducts. Three individual lymphatic meshes remove excess liquid from the mucosa, muscularis and serosa in an independent fashion (Eddy and Pauerstein 1980, Leese 1988). Interestingly, the lymphatic vessels become expanded at oestrus, when the oviduct fluid volume increases and there is need of liquid removal and possible waste products.

Historically, the oviducts were thought to be a sterile, immune inert or mildly immune-responsive organ with no pathogenic threats. However, around the

late 20th century, the identification of subepithelial or epithelial-intercalated immune cells such as lymphocytes, neutrophiles, macrophages and granulocytes, as well as early cytokine studies on oviduct response to pathogenic insults, introduced the idea of a complete immune function in the oviducts (reviewed in Cardenas et al. 1998). Nowadays, there is compelling evidence that the oviducts exert a tight regulation of the immune system function in order to fight pathological insults such as bacteria and viruses, while also allowing for the presence of the allogeneic sperm cells and the partially allogeneic blastocyst (Wira et al. 2005). Furthermore, it is now known that the oviducts are not sterile but count on a facultative anaerobic and aerobic bacterial microbiome, which is substantially less rich than the one found in the lower genital tract (Chen et al. 2017).

The structure, cellular inhabitants and morphological organisation of the oviducts are intimately linked to their biological functions, which are mainly mediated by the epithelium. However, this has been a traditional and rather simplistic view, given that the functional and cyclic complexity of the oviduct requires a high degree of biological regulation based on the precise and integrative orchestration of many cell types and dynamic transcriptional programmes. Hence, it is important to keep in mind that other cellular subtypes also play an important role not only in maintaining the oviduct reproductive functionality, but also supporting its integrity and homeostasis as well as protecting the organ from tissue damage during the continuous biological changes of the surrounding environment.

4 The oviduct and the reproductive function

The main functions of the oviducts are the transportation and nutrition of the female and male gametes, sperm selection and the provision of an adequate environment for successful fertilisation and early development of the pre-implantation embryo.

4.1 General transport mechanisms in the oviduct

Generally, the transport of sex gametes and pre-implantation embryo is achieved by two mechanisms: (1) the ciliary movement of ciliated cells and (2) the muscular contractions mediated by SM cells. These contractile movements have been reported to be controlled by prostaglandin E2 (PGE2) and F2α (PGF2α), endothelin-1 (ET-1) and angiotensin II (Ang II) (Wijayagunawardane et al. 2001, Spilman and Harper 1975). Additionally, the ciliary movement is important for mucus clearance, although the extent of its contribution to a successful fertilisation remains unclear. For example, women with the Kartagener syndrome, characterised by a ciliary dyskinesia that triggers abnormal ciliary function or structure, can present both fertile or infertile phenotypes, although researchers suggest that cases in which fertilisation was achieved could have been due to an atypical ciliary movement (McComb et al. 1986).

4.2 Ovulation, egg pickup and oviduct-egg communication

The oviducts lie in close proximity to the ovaries, which contain the female gametes. In the ovaries, immature oocytes are found in a state of arrest at prophase I of meiosis I. Approximately every 28 days in human and 5 days in mouse and following the increase in the levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) during the ovulatory cycle, follicular growth is promoted and meiotic oocyte progression from prophase I to a new blockage at metaphase II is undertaken. This meiosis reactivation is known as maturation and also equips the oocyte with developmental competence to become a viable blastocyst post-fertilisation. This has been demonstrated by the fact that bovine oocytes matured and developed *in vitro* displayed a 26.4% success rate in blastocyst formation vs the 49.3% of those that were matured *in vivo* and developed *in vitro* (van de Leemput et al. 1999). In another study, this blastocyst formation rate was 39.2% vs 58.2% following *in vitro* and *in vivo* maturation, respectively (Lonergan et al. 2001). Moreover, *in vivo* maturation also yielded a higher number of viable blastocysts (Dieleman et al. 2002). In line with these findings, oocytes from early stages of the preovulatory period had an *in vitro* blastocyst-forming rate of 39.3-41.5% while oocytes from late stages and more advanced in maturation reached a 57.6% (Humblot et al. 2005). Although the exact mechanisms underlying oocyte maturation and controlling their developmental potential are unknown, leptin has been proposed as a hormone positively influencing these processes in a cumulus cell-dependent and independent manner (Boelhauve et al. 2005, Paula-Lopes et al. 2007). Cumulus cells are a type of granulosa cells, somatic cells surrounding the oocyte during folliculogenesis. Together they form the cumulus-oocyte complex (COC) where cumulus cells support the oocyte during maturation and ovulation. This function is further endorsed by gene expression and proteomic studies in bovine oocytes which delineated their role controlling the oocyte developmental competence and survival, as well as in progesterone secretion. These studies highlighted the cumulus cells expression of the enzyme Glutathione S-transferase A1 (*GSTA1*), the FSH receptor (*FSHR*), aromatase (*CYP19A1*), steroid-metabolism enzymes such as steroidogenic acute regulatory protein (*STARD1*), cholesterol side-chain cleavage enzyme P450scc (*CYP11A1*) and 3-beta-hydroxysteroid dehydrogenase (*HSD3B1* and *HSD3B2* human genes) and the TGFβ and MAPK signalling pathways (Salhab et al. 2011, Salhab et al. 2013). A recent proteomic study also highlighted that cumulus cells of *in vivo*-matured COCs differentially expressed proteins related to the complement and coagulation cascades (CCC), extracellular matrix (ECM)-receptor interaction, required for adhesion to the oviduct epithelium, and steroid biosynthesis (Walter et al. 2020).

After oocyte maturation, ovulation occurs, with the consequent release of a mature COC, which resolves and completes the second meiosis arrest once it is fertilised by a spermatozoon. Although human and mouse have different cycles, the hormonal fluctuations are similar in both species, therefore, the mouse oestrus cycle is used as a representative explanatory example. In mouse, four consecutive stages occur: proestrus (rising oestrogen concentration), oestrus (oestrogen peak and consequent ovulation), metestrus (increasing progesterone concentration and decrease in oestrogen) and dioestrus, which is the largest stage and lasts around 2 days (peak of progesterone concentration and minimal levels of oestrogen) (Figure

6, pg 34). Following ovulation, the COC is captured by the projected fimbriae, enters the oviduct ostium and then migrates towards the ampulla, where it becomes stationary (Kölle et al. 2009, Lam et al. 2000). These processes are possible thanks to the regulated adhesion of the cumulus cell matrix, rich in hyaluronan, to the ciliated epithelial surface (Lam et al. 2000) and a high CBF environment produced by cilia beating following the fimbriae-isthmus direction (Kölle et al. 2009, Norwood et al. 1978). Due to the importance of ciliated cells for these events, they are strategically enriched in the cranial parts of the oviducts (Figure 7, pg 34).

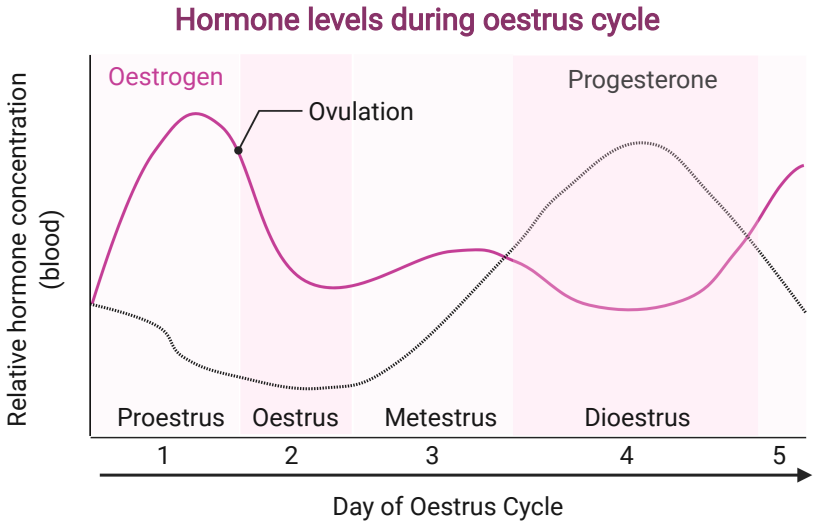


Figure 6. The hormone concentration during the oestrus cycle of mice

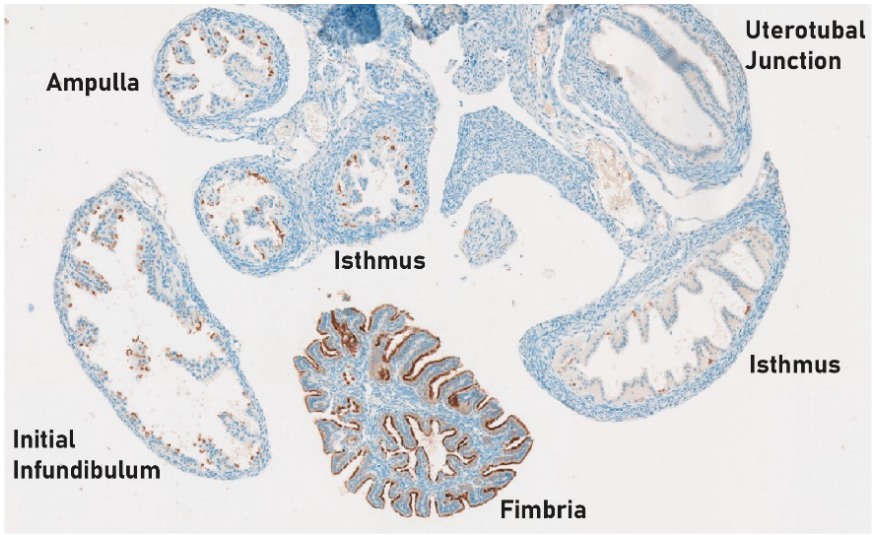


Figure 7. Distribution of ciliated epithelial cells in the oviducts
The different sections of the oviducts are portrayed and stained with IHC using the ciliated marker Tubb4.

The cumulus cell matrix is elastic and soft, key physicomechanical properties for oviduct pickup and bonding to the epithelium (Curtis 2016). Their importance can be especially observed when the COC is transported across the ostium. Because the COC's size is too big to fit in the ostium opening, the COC decreases its diameter, which is achieved by reducing the elasticity of the cumulus cell matrix (Talbot et al. 1999). Interestingly, factors such as the change in the cumulus cell matrix properties during *in vitro* treatments and chemicals contained in cigarette smoke seem to negatively affect COC pickup and posterior epithelial adhesion (Talbot et al. 1999), being the latter a plausible explanation for the negative impact of smoking on female fertility (Stillman et al. 1986).

After entering the ostium, the COC is transported to the ampulla, the section where fertilisation occurs. While the oocyte awaits until the encounter with spermatozoa, interactions between the underlying oviduct epithelium and the COC occur. Effective interactions are suggested to happen independently from oocyte maturation but dependently on oocyte viability. This was concluded after attachment between a bovine immature oocyte and the ampullar epithelium was observed (Kölle et al. 2009) and after attachment to a degenerated oocyte (defined as counting on few layers of cumulus cells and compromised cytoplasm) was unsuccessful. Furthermore, attachment to the canine oviduct epithelium has been shown to improve the quality of COCs and subsequent embryos *in vitro*, which could suggest the same beneficial effect *in vivo* (Lee et al. 2019). Exosomes secreted by canine oviduct epithelial cells were identified as the vectors mediating these interactions. These exosomes induced the activation of the EGFR/MAPK signalling pathway in cumulus cells of the attached COC, increasing their viability and proliferation and decreasing the levels of reactive oxygen species (ROS) and the apoptosis rate (Lee et al. 2020). This interaction network has been further characterised by gene expression analyses, which found that cultured bovine oviduct epithelial cells change their transcriptome in reponse to COC binding, primarily upregulating proteins such as the metalloproteases MMP-1/2, and the TGFβ signalling molecule TGFβ3 (Gabler et al. 2008). Similarly, a transcriptomic study in cow oviducts also showed that the epithelial transcriptome undertakes dynamic changes in response to the COC (Bauersachs et al. 2003). Additionally, the oviduct epithelium is known to interact and respond through transcriptional changes not only to COC's contact, but also to follicular fluid (Hasan et al. 2020), the early embryo (Lee et al. 2002, Schmaltz-Panneau et al. 2014) and sperm cells (Fazeli et al. 2004, Alminana et al. 2014). Curiously, in the case of bovine sperm cells, they were able to trigger changes in the oviduct epithelium gene expression in a contact-free paracrine fashion *in vitro* (Reshi et al. 2020).

4.3 The sperm journey and interactions with the female reproductive tract

Spermatozoa begin their journey into the female reproductive system and towards the COC in the vagina. Thereafter, they are able to cross the uterine horns and reach the oviducts. Many sperm cells are released during ejaculation but only a small proportion of them will make it to the oviduct and, eventually, only one of them will be able to fertilise the egg. In humans, it has been estimated that around 300

million sperm cells are delivered in the vagina but only 300 make it to the fallopian tube. This highly selective process is due to the fact that during the spermatozoa journey to reach the oocyte, they need to overcome a series of obstacles in the female reproductive tract. Only the most capable cells are able to overcome the acidic pH in the vagina, the opposite mucus current, the triggering of the immune reaction in the vagina and the uterine horns and the physical convolutions of the oviduct lumen. The sperm population is no longer classified according to the classical division into abnormal, immotile and motile cells since new categories related to sperm motility (Martínez-Pastor et al. 2011), DNA methylation patterns (Jenkins et al. 2015) and ssDNA and dsDNA fragmentation (Ribas-Maynou et al. 2012) have been proposed and linked to their quality and fertility capacity. Recently, the idea of a dedicated sperm-cell selection by the FRT has gained interest due to the evidence arising from sperm heterogeneity and detected levels of discrimination against sperm cells bearing fragmented DNA in the mouse (Hourcade et al. 2010).

When spermatozoa enter the oviducts, a proportion of them are able to bind to the isthmic epithelium, which provides a specialised reservoir enclave in an oestrus cycle-independent manner. Here, the fertility capabilities of the sperm cells are maintained until encounter with the COC, although avoidance of polyspermy is also a possible function of this reservoir. The isthmic reservoir has been observed in many species such as mouse (Suarez 1987), cow (Kölle et al. 2009, Camara Pirez et al. 2020), hamster (Smith and Yanagimachi 1991), horse (Lefebvre et al. 1995) and pig/boar (Suarez et al. 1991), both *in vivo* and *in vitro*, although in humans this has only been observed *in vitro* (Pacey et al. 1995, Baillie et al. 1997). Furthermore, this sperm-epithelial interaction has been reported to decrease in fallopian tubes from post-menopausal women and also patients with endometriosis (Baillie et al. 1997, Reeve et al. 2005).

The sperm cells attached to the oviduct epithelium can be motile or stationary (Baillie et al. 1997) and they preferably bind to the cilia of ciliated epithelial cells on the acrosome zone of the membrane (Suarez et al. 1991, Kölle et al. 2009, Camara Pirez et al. 2020). Nevertheless, interactions between sperm heads and secretory cells have also been observed (Fléchon and Hunter 1981, Hunter et al. 1991). Furthermore, sperm cells have been shown to bind to the oviduct ampullar epithelium, although these events are uncommon under physiological conditions (Camara Pirez et al. 2020, Baillie et al. 1997). The biochemical process underlying these physical interactions is mediated by a lectin-like molecule binding to a carbohydrate. The sperm's head expresses the lectin-like factors and recognises carbohydrates expressed by oviduct epithelial cells. This union is Ca^{2+}-dependent and the carbohydrates involved in this interaction seem to be highly species-specific. For example, asialofetuin and its terminal sugar galactose is the equine effector molecule (Lefebvre et al. 1995), fetuin and the sialic acid residue are the ones in hamsters (DeMott et al. 1995) and rats (Cortés et al. 2004), and mannose is the key carbohydrate in pigs (Wagner et al. 2002). In humans, this situation in less understood but it has been demonstrated that human sperm can bind the fallopian tube epithelium through the molecule fucosyltransferase-5, which binds carbohydrates (Huang et al. 2015). The same study proposed asialofetuin as the possible carbohydrate sugar, although further research would be required to conclusively determine the sugar identity in humans. Sperm cells and oviduct epithelium have been shown to

not only interact and communicate in a cell-to-cell contact basis, but also through paracrine signalling using oviductal extracellular vesicles secreted by the epithelium, also known as oviductosomes, which have been shown to regulate fertility functions (reviewed in Al-Dossary and Martin-Deleon 2016). Oviductosomes have been primarily studied in bovine and murine models, however, it has been reported that fallopian tube epithelial cells release extracellular vesicles similar to the murine ones (Bathala et al. 2018).

In the vagina and uterus lumen, the presence of the foreign sperm cells triggers an immune response, activating the production of pro-inflammatory cytokines, which causes a leukocytes influx (Rodriguez-Martinez et al. 1990, Song et al. 2016). Paradoxically, this situation does not happen in the oviducts and they act as reservoirs, sheltering and nourishing sperm cells, when they would be expected to trigger a hostile immune reaction. This immunosuppression has been reported to be mediated by the anti-inflammatory cytokines TGFβ1 and IL10, which are strongly expressed in the bovine oviduct and whose expression in epithelial cells and neutrophils is increased after their contact with sperm cells *in vitro* (Yousef et al. 2016). This contact also promoted an increase of PGE2 production by both cell types *in vitro* and when exogenous PGE2 was added to an oviduct epithelium culture, these cells downregulated the pro-inflammatory cytokines TNF and IL1B (Yousef et al. 2016). Previously, a PGE2 increase by oviduct epithelial cells upon sperm attachment had been identified, although it had exclusively been linked to the aid in transport by increasing muscular contraction (Kodithuwakku et al. 2007). Regarding a possible immunological role, PGE2 had also been reported to control the suppression of sperm phagocytosis by neutrophils in the oviducts (Marey et al. 2013). Later research found that this was sperm-specific, since it was not observed for bacteria, and aided by epithelial cells, demonstrating yet another protective role of the oviduct epithelium towards spermatozoa (Marey et al. 2020). Interestingly, incubation of oviduct neutrophils with media derived from oviduct epithelial culture also induced the expression of TGFβ1 and IL10 (Yousef et al. 2016). Taken together, these evidences suggest that under physiological conditions the oviducts finely tune a selective immunosuppressive state that allows them to welcome sperm cells, while still retaining their protective function against pathogenic threats.

Sperm cells stay at the isthmic portion until ovulation happens. Then, to carry out fertilisation, they detach from the oviduct epithelium and travel through the oviducts to find the ovulated COC in the ampulla. This release from the tubal reservoir seems to be controlled by capacitation, a process essential for fertilisation. Capacitation is a maturation step where molecular changes remodel the sperm membrane in preparation for sperm hyperactivation, the acrosome reaction (AR) and, consequently, the COC fertilisation. Hyperactivation is a change in the flagellum movement, which becomes ample and asymmetrical, aiding the sperm in detaching from the isthmus and progressing towards the oviduct (Demott and Suarez 1992). The acrosomal reaction is an exocytosis process undertaken by the sperm when it has reached the COC and where enzymes are released to enable the fusion with the COC membrane and subsequent penetrance. The exact molecular mechanisms undertaken during capacitation and how they trigger hyperactivation and AR are poorly understood, however, it seems that cholesterol excretion, changes in the phosphorylation pattern of proteins and membrane hyperpolarisation, content dis-

tribution and changes in ionic permeability, are directly involved (reviewed in Puga Molina et al. 2018, Stival et al. 2016 and De Jonge 2005). In bovine spermatozoa, the decrease of a sperm-head protein known as PDC-109 during capacitation has been linked to a sperm-epithelium affinity reduction (Ignotz et al. 2001), whereas the implication of hormones in mediating the sperm detachment has been discarded (Suarez et al. 1991, Baillie et al. 1997). In summary, a series of molecular alterations seem to reduce the sperm affinity for the membrane carbohydrates of the isthmic epithelium, which coupled to hyperactivation, results in sperm liberation.

4.4 Fertilisation and early development of the pre-implantation embryo

Ahead of the detached sperm cells there is an arduous path filled with obstacles such as mucus, entwined epithelial protrusions and an opposite flow current. Furthermore, upon contact with the COC, the sperm cells need to pierce its two protective layers: (1) the corona radiata made of granulosa cells and (2) the zona pellucida, which is a glycoprotein membrane surrounding the oocyte's plasma membrane. The ciliary beating rate is also very high at the time of spermatozoa detachment from the oviduct epithelium and it is their hyperactivation state, along with the oviduct smooth muscle contraction, which allows them to move against the current and travel towards the COC (Kölle et al. 2009). Hyperactivation is controlled by the Catsper Ca^{2+} protein channel located on the main part of the sperm tail. Two independent mouse models bearing null mutations for *CatSper1* and *CatSper2*, are unable to undertake hyperactivation, which leads to reduced sperm motility, prevents appropriate travelling through the luminal mucus and penetration of the zona pellucida, and ultimately results in male sterility (Ren et al. 2001, Quill et al. 2003, Carlson et al. 2003).

To date, sperm cells have been reported to detect the right direction for fertilisation following three mechanisms, thermotaxis, chemotaxis and rheotaxis, which guide sperm movement towards higher temperature, chemoattractants and against mucus flow, respectively (fully reviewed in Perez-Cerezales et al. 2015). Thermotaxis and rheotaxis are long-range mechanisms as they drive the sperm through long distances in the oviducts, whereas chemotaxis is a short-range process effective in the vicinity of the COC. In rabbits (Bahat et al. 2003) and pigs (Hunter and Nichol 1986), a temperature gradient is established in the oviducts between the isthmus and the ampulla shortly after ovulation. The lowest temperature of the gradient is found in the isthmus and the highest in the ampulla, with a 2°C difference, and only a subset of sperm cells that have undergone capacitation follow this increasing temperature gradient towards the ampulla (Bahat et al. 2003). Furthermore, during chemotaxis two different sources of signalling molecules have been proposed. First, chemotaxis towards follicular fluid has been described in a small fraction of human and mouse spermatozoa (those capacitated) (reviewed in Eisenbach 1999). Second, the human oocyte and the surrounding cumulus cells have been found to secrete factors attracting sperm cells *in vitro* (Sun et al. 2005) with progesterone as a potential candidate (Teves et al. 2006). Further evidence comes from the discovery of a G protein-coupled olfactory receptor (hOR17-4) in human spermatozoa, which mediates chemotaxis responsiveness to signalling molecules released by the egg (Spehr et al. 2003). Regarding rheotaxis, the movement against a current flow, it has been

reported to happen in mouse and human in a capacitation-independent fashion and with CatSper channels involved (Miki and Clapham 2013). In this study, after sexual intercourse in mice, epithelial cells increase their secretion to the luminal space and the fluid flow from the oviducts to the uterine horns increases. This process is dependent on the hormone prolactin and the established flow lowers viscosity, clears debris from the oviduct lumen and creates a stream axis that guides the sperm to the COC. Interestingly, this is in consonance with what is described by Camara Pirez et al. 2020, where ciliated cells respond to the sperm entrance and, after the creation of the reservoir, the CBF and epithelial endoplasmic reticulum activity is increased. Taken together, these evidences suggest that the sperm cells that arrive to the COC have been filtered based on their capacity to respond to these stimuli.

A full revision of the processes occurring after the sperm has arrived to the COC's proximities can be found in Yeste et al. 2017. Briefly, a capacitated spermatozoon undertakes an acrosomal reaction and binds to and pierces the zona pellucida of the COC. Binding of the sperm cell to the zona pellucida is mediated, as with oviduct epithelial cells, by carbohydrate recognition. Furthermore, its membrane fuses with the oocyte membrane and the oocyte undertakes activation, which leads to the cortical reaction (hardening of the zona pellucida to prevent polyspermy), triggers the exit from the meiosis II arrest and starts embryonic development. Some molecules have been found to play a role in the oviduct functions during these events. For example, the oestrogen receptor α (*ERα*) in oviduct cells aids fertilisation from different angles given that following its disruption, sperm migration is greatly reduced and the immunosuppression in the oviduct is disturbed, leading to the expression of proteases and downregulation of their inhibitors, which causes degradation of the zona pellucida and the embryo (Winuthayanon et al. 2015). Also, the transcription factor Wt1 has been found to contribute to fertility by repressing the expression of the serine protease Prss29, which causes infertility and impairs appropriate embryonic development (Nathan et al. 2017).

Following successful fertilisation, a zygote is created and detaches from the ampullar epithelium continuing its journey towards the uterine horns for implantation. Transportation to the uterus seems to be tightly regulated and involves embryo fate decision, as it was found in rats that 1-cell embryos transferred to oviducts would progress towards early development in the oviduct, whereas when 4-cell stage embryos were transferred, most of them were subsequently transported to the uterine horns (Ortiz et al. 1989). Furthermore, the highest number of genes expressed by the oviduct epithelium following the initial contact with the zygote was found at the 4-cell stage (Chang et al. 2000). It is in the oviducts, where the first maternal-embryonic communications are established and, while the zygote is in the oviducts, they respond by decreasing the ciliary beating rate and increasing the number of secretory cells until it enters the uterine environment. Local changes also occur around the embryo to provide vascular supply, leading to morphology changes of the tubae uterinae arteria, which increases the proximity between its surface and the oviductal lumen (Kölle et al. 2009). Another elegant example of this cross-talk comes from the *in vitro* evidence that the oviduct epithelium triggers the production of interferon-tau in embryos, which induces an anti-inflammatory phenotype in immune cells (Talukder et al. 2018). This has been further observed in bovine oviducts where the presence of the embryo triggers the downregulation of immune

genes related to inflammation (Maillo et al. 2015). Also in equines an effect on the expression of immune genes in the oviducts following sperm contact was also found (Smits et al. 2016).

5 The oviduct as the origin and site of diseases

In line with the paramount importance of the oviducts in female health and reproduction, alterations of the oviduct biology can lead to serious pathologies occurring at different stages in life. During embryonic development, malformations of the Müllerian ducts can lead to a varied range of congenital disease phenotypes. In this instance, genomic aberrations and specific gene mutations have been identified as the major cause and can partially explain the phenotypic complexity found in patients. Furthermore, at later adult stages, diseases such as ectopic pregnancy, malignancy and severe infections, can also occur in the oviducts. These diseases are individually presented here in order to understand the magnitude of their impact on the quality of life and fertility of patients and how some of them represent life-threatening conditions.

Following a chronological order, the first range of pathologies affecting the FRT are those arising during embryonic development and known as Müllerian duct anomalies (MDAs), reviewed in Santana Gonzalez et al. 2021b. Briefly, MDAs display a huge range of phenotypic combinations depending on the embryonic time at which the aberration happened and how many organs of the FRT were affected. In addition to the complex clinical pictures that these disorders present, they also display variable penetrance, where even the same genomic abnormalities can lead to complete different outcomes, suggesting further levels of non-genomic regulation of the phenotypic expression. Müllerian conditions vary from complete agenesis, the most serious, to dysplasia, the mildest, and these malformations can be found on an individual or several FRT organs. Furthermore, these aberrations can exclusively affect the reproductive system or be found in association with abnormalities in other organ systems. In the general healthy female population, the incidence of MDAs is 6.7%, although in women with a record of infertility and in those with recurrent miscarriages the incidence is higher, with a 7.3% and a 16.7%, respectively (Saravelos et al. 2008). The most common form of MDAs is the Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome (OMIM 277000), which is generally characterised by agenesis of the vagina, cervix and uterus or, alternatively, the same features with fusion or differentiation defects in the uterus instead of aplasia (Morcel et al. 2007, Guerrier et al. 2006). This syndrome affects 1 in 4500-5000 newborn female babies. During fertility-related events, the consequences of MDAs can be found at two main levels: (1) before and during pregnancy and (2) during labour. In the first one, MDAs increase the risk of impaired reproduction, recurrent miscarriage during the first trimester of pregnancy, foetal malpresentations and retarded growth, whereas in the second instance MDAs are linked to premature birth, caesarean section and placental retention (Rock and Schlaff 1985, Acién 1993, Maneschi et al. 1995, Green and Harris 1976, Lin et al. 2002). Regarding the genetic entities behind these abnormalities, the chromosomes X, 1, 3, 9, 16, 17 and 22, have been identified as hotspots for MDAs-related genomic rearrangements. Moreover, mutations and genomic imbalances affecting Müllerian master embryonic genes, such as *LHX1*,

HNF1B, *WNT4*, *EMX2* and the HOX cluster, have also been reported as causative events.

The aforementioned developmental diseases occur during the early oviduct lifespan, however, there are also a number of pathologies affecting this organ during adulthood. Among these pathologies, ovarian cancer is the first presented due to its extreme severity and terrible impact on women's life. Ovarian cancer has been identified as the most lethal of the gynaecological cancers given its poor prognosis, partly related to its late-stage detection. Only in the United Kingdom, it causes more than 4000 deaths annually, whereas this number raises to 150.000 worldwide (Cancer Research UK, 2021). Ovarian cancer is a blend of complex and heterogenous diseases originated by the uncontrolled division of cells *in situ* and the later spread to other body locations, at more advanced stages of the disease, in a process known as metastasis. It is currently thought that the mechanisms driving these tumorigenic processes are the gradual accumulation of mutations in normal epithelial cells in charge of the homeostasis and correct function of the organ, which eventually culminate in their transformation into cancer cells. The most aggressive and common version of ovarian cancer is high-grade serous ovarian cancer (HGSOC), a subtype diagnosed in 70-80% of the cases. Contrary to the past belief in the scientific community and what its name suggests, recent studies have reported the fallopian tubes/oviduct secretory epithelium as its possible originating source, in contraposition/addition to the ovarian epithelium (Ng and Barker 2015, Perets et al. 2013). Following this theory, pre-cancerous secretory cells of the fallopian tubes would acquire a first round of molecular alterations, evolving into early lesions termed serous tubal intraepithelial carcinomas (STICs), which share markers with HGSOC such as p53, Ki-67, Rsf-1, cyclin E and fatty acid synthase (Ahmed et al. 2010, Sehdev et al. 2010, Kuhn et al. 2012). The cells comprising these early lesions would be then shed on to the ovarian epithelium where they would progress into an HGSOC phenotype.

Additionally, ectopic pregnancy and severe infections can occur in the oviducts as well. Ectopic pregnancy is an implantation that takes place outside the uterine endometrium and its incidence in the UK is 11 per 1000 pregnancies with 10.000 cases reported per annum and a mortality rate of 0.35 per 1000 ectopic pregnancies (Elson et al. 2016 on behalf of the Royal College of Obstetricians and Gynaecologists, Rana et al. 2013). Diagnostic strategies for this condition are transvaginal ultrasonography, measurement of human chorionic gonadotropin (CG) and progesterone, computed tomography, measurement of vascular endothelial growth factor, laparoscopy, etc. (Farquhar 2005, Rana et al. 2013). This disease is a medical and life-threatening emergency as more than 75% of maternal deaths during the first trimester and 9-13% of the total of deaths caused by pregnancy factors are related to ectopic pregnancy (Farquhar 2005). In 90-95% of the ectopic pregnancy cases, retention of the pre-implantation embryo occurs in the fallopian tubes, preventing its transport to the uterus and the commencement of pregnancy. The vast majority of abnormal implantations take place in the ampulla, the natural site for fertilisation, and the identified risk factors hitherto are of maternal origin (Rana et al. 2013). In tubal ectopic pregnancies, dysfunctional G-protein-coupled cannabinoid receptor 1 (CNR1) and β2-adrenergic receptor (ADRB2) were found to impede embryo transportation in the oviducts as they control smooth muscle

movement (Wang et al. 2004).

Furthermore, infections and inflammation of the oviduct (salpingitis) are pathological conditions that can also occur and are triggered by pathogens such as *Chlamydia trachomatis*, responsible for the sexually-transmitted disease chlamydia. Although salpingitis is generally a consequence of sexual activity, rare cases induced by other non-sexually transmitted bacteria such as *Streptococcus pneumoniae* and *Escherichia coli* have been reported (van der Putten 2008, Arda et al. 2004). Tubal infection induces an inflammatory response that has been associated with an interruption of the tubal functionality and a reduction of sperm cross-talk with the oviduct epithelium resulting into subfertility and infertility in cows (Owhor et al. 2019). Interestingly, the uterine endometrium was not exempted from this infection and endometritis was always found along with salpingitis. Furthermore, in mouse models, interruption of the ovulated egg transport into the oviduct and consequent infertility was also reported in salpingitis conditions (Tuffrey et al. 1990). Taken together, these serious conditions can have very negative effects on the female quality of life, posing a high risk to fertility and pregnancy development, and in the case of malignancies and ectopic pregnancies, a direct threat to patients' life.

6 Prospects in the field and the role of this project

Despite years of research that have improved our knowledge about the conduits of life, their underlying biology and deregulation under certain pathological conditions, there is still a large amount of poorly understood aspects and the progress in the field is still slow, especially, when compared with the rate of discovery for other organs. This situation can be partially explained by the fact that the tubes are historically understudied, even within the FRT system, having attracted less attention than other organs due to their less "catchy" role in human biology and pathology. Although there has been increasing research efforts and resources for the study of this organ in the last years, paramount insights regarding stem cell biology and differentiation, the identity of the organ cell constituents, their specific functions and their cross-talks as well as the roles and molecular processes underlying the physiology and pathology, have been long lacking for the oviducts.

An expanding and powerful research strategy holding great potential to boost our knowledge in the field is based on functional genomics. Functional genomics focus on the study of the genomic contribution to biological processes and phenotypes. This research approach utilises a genome-wide strategy and comprises the diverse, but complementary, divisions of genomics, epigenomics, transcriptomics, proteomics and metabolomics. Among these disciplines, transcriptomics provide a unique dynamic vision of the gene expression signatures that characterise different tested conditions. In other words, transcriptomics obtain a gene expression-based phenotypic snapshot of a biological system at a given time, under a given condition. The humble, yet extremely powerful, beginning of the transcriptomic history took off with the invention of the quantitative reverse transcription polymerase chain reaction (RT-qPCR), which allowed the identification and quantification of gene transcripts expressed by a biological system at a certain time. However, this technique presents limitations for large scale investigations due to its low-throughput nature. Therefore, the development of high-throughput breakthroughs such as microarrays

and, concretely, next-generation sequencing (NGS), represented an incomparable omics revolution and a shift in the transcriptomic research paradigm. The genesis of microarrays in the 90s allowed a large-scale simultaneous interrogation of the expression of thousands of genes for the first time and prompted their peculiar description by the journal Science in 2001 as "*the hottest thing in biology and medicine since the advent of the polymerase chain reaction a decade ago*" in A Heaping Serving of Gene Chips, Please. Science, 19 Oct 2001: Vol. 294, Issue 5542, pp. 483. Microarrays are a genetic tool based on reference DNA probes attached to a substrate and a sample containing query nucleic acid sequences incubated with these references. Only complementary and expressed strands in the query hybridise and their average expression can be quantified, whereas non-complementary ones will be removed and not accounted for. Despite the progress they promoted, microarrays are an experimental technique in which previous knowledge of at least partial sequences of the query RNAs is required and they are now replaced by emerging NGS, which possess a higher throughput power and a less biased approach, relegating microarrays to an honorary grandparent's role to the current transcriptomics generation.

Within the RNA NGS, two modalities can be distinguished: (1) bulk RNA-sequencing (bulk RNA-Seq) and (2) single-cell RNA sequencing (scRNA-seq). Bulk RNA-Seq became popular in the late 2000s and, as microarray-based techniques, measures the average expression of each gene for all the cells contained within a sample. However, unlike its predecessor, it offers higher reproducibility with less technical variability and additional features such as higher sensitivity in detecting gene expression, and the ability to identify alternative splice variants and novel mRNAs (Marioni et al. 2008). Bulk RNA-Seq has driven much of the biomedical progress in the century. For instance, its use allowed the identification of a gene-expression signature to diagnose high-risk acute-leukaemia patients who do not respond to standard clinical treatment and require an alternative management (Ng et al. 2016). However, this method does not allow to characterise the heterogeneity within complex cell populations, hindering the possibility to study specific subpopulations and identify the correlation between these and specific biological processes. This limitation is particularly relevant when a biological process is mediated by a rare cell population such as stem/niche cells. Moreover, variations in specific cell-type proportions captured in the bulk RNA-Seq sample might introduce biased/inconsistent results (e.g. the presence of a dominant but "irrelevant" cell population can completely mask the expression of relevant genes by a rare cell subgroup). The emergence and development of gene expression tools applicable at the single-cell resolution level provided a solution to this scenario by allowing for the unbiased characterisation of individual cell identities. This, applied to the study of multiple organs, allowed a better understanding of their underlying biology, the discovery of rare cell populations previously unidentified and the hierarchical relationship between diverse cellular lineages. The first publication in the field was that of Tang et al. 2009 and since then, scRNA-seq has led many of the most impactful discoveries and breakthroughs in a vast range of fields, species and applications (Hwang et al. 2018, Tang et al. 2019, Kulkarni et al. 2019, Aldridge and Teichmann 2020). Furthermore, another powerful and complementary approach that has gained increasing attention over the recent years, is the study of systems biology. In this research approach, the key principle is that an organ biology is based on the syn-

chronised action of complex and heterogenous cellular systems. This is in contrast to the more traditional synecdochic approach of representing the whole organ by the study of a representative part of it, commonly, the cell population more directly involved in the organ function.

As exciting as these state-of-the-art areas stand, there should also be caution as they constitute high-throughput experimental techniques, which are based on the production and delivery of big amounts of data. This poses two associated challenges that relate to: (1) big data obtention and production from a biological system and (2) extracting meaningful biological information from the data mass. The researcher has to face and deal with them by mastering the experimental technique used to generate the data and, also, through adopting a robust analytical pipeline introducing the least possible biases in both steps.

Aware of the challenges and the knowledge gaps characterising the gynaecological studies of the oviducts (or fallopian tubes), the present project aims to contribute to the field by utilising state-of-the-art transcriptomic approaches in order to advance our knowledge about the oviducts underlying biology, non-genetic heterogeneity, functionality and deregulation.

II Aims and objectives

- To design and optimise an adequate single-cell dissociation protocol for the mouse oviducts **(Chapter 1)**
- To decipher the cell systems comprising the mouse oviduct landscape and to provide a functional insight into their orchestration of the oviduct biology (**Chapter 2**)
- To draw an inter-species comparison of oviduct cell populations in human and mouse and characterise their cross-talk and interactions (**Chapter 3**)
- To discover potential molecular interactions and pathways within the oviduct that are conserved across species and between the homeostatic and diseased states (**Chapter 4**)

III Methods

1 Technical optimisation of oviduct single-cell isolation

1.1 Mouse work

All female mice were purchased from the company Charles River. In order to reduce possible variability linked to age, all used adult mice were 12-week-old. Furthermore, to capture potential dynamic changes in cell composition and gene expression as a result of the oestrus cycle, representative samples were taken from mice across all oestrus stages. When oviducts coming from different animals were isolated, they were pooled together before mechanical and enzymatic dissociation in order to correct for any individual mouse effect.

1.2 Sample processing and tissue dissociation

For each experiment, the mouse oviducts were identified, harvested and rapidly cut and processed to preserve the single-cell transcriptomes as much as possible. In this regard, it is also important to highlight that the organs and the single cells were at all times kept at 4°C, except for the incubation with digestive enzymes. In this project, the first cocktail used for single-cell dissociation consisted of: 2 mg/ml Trypsin (Sigma), 0.5 mg/ml DNase I (Sigma) and 100 U/ml Collagenase Type I (Invitrogen) in 1X Dulbecco's phosphate-buffered saline (DPBS, Sigma). The second, and finally selected cocktail, contained: 1.25 mg/ml Liberase TM Research Grade (Sigma, medium Thermolysin concentration) and 0.5 mg/ml DNase I (Sigma) in 1 ml of medium (HEPES [Gibco, 1:100], GlutaMax [Gibco, 1:100] in DMEM/F12 (Gibco)). Then, this cocktail was added to finely minced oviducts in a 1.5 ml DNA LoBind Eppendorf tube. DNA low-binding tubes were used for all experiments to prevent the cells and the genetic material from binding to the plastic walls of the Eppendorf tubes. The tube containing the sample was placed in a 5382 ThermoMixer C v.3.0.2 at 37°C and 800 rpm vibration speed for the sequential times of 8, 15, 30 and 45 minutes. After that time, the suspension was filtered through a 30 μm filter (Sysmex) attached to a 15 ml tube, previously washed with 1X DPBS 1% BSA. Then, the filtered cell suspension was centrifuged for 5 minutes at 1600 rpm at 4°C, the supernatant was discarded and the remaining cells were prepared for further analysis.

1.3 Fuorescence-activated cell sorting (FACS)

The cell pellet produced at the end of the tissue processing step was resuspended in a Blocking Solution of 90 μl FACS Buffer (1X DPBS, 1% bovine serum albumin [BSA, Sigma-Aldrich], 2 mM EDTA pH 8.0 [Invitrogen]) and 10 μl of Mouse FcR Blocking Reagent [Miltenyi, 130-092-575], and incubated for 5 minutes. This Blocking Reagent is used to increase the specificity of the individual antibodies used for subsequent labelling to their target by blocking their non-targeted binding to immune cells expressing Fc receptors, such as B cells and macrophages. Subsequently, the cells are stained with anti-mouse EpCAM-APC antibody [Miltenyi Biotec, 130-102-234, 1:10

concentration] and anti-mouse CD45-FITC antibody [Miltenyi Biotec, 130-116-500, 1:50 concentration] for 10 minutes at 4°C in the dark. In the experiment detecting the MHC Type I, the antibody used was an anti-mouse H-2-FITC antibody produced in rat [Biolegend, clone M1/42, 125508, 1:10 concentration]. Then 10 μg/ml DAPI was added to the samples and they were further resuspended in FACS buffer to be sorted with a BD FACSAria™ II sorter. Alive EpCAM^{+}CD45^{-}DAPI^{-} oviduct cells were enriched to represent 70% of the final sample, while EpCAM^{-}CD45^{+}DAPI^{-} and EpCAM^{-}CD45^{-}DAPI^{-} oviduct cells constituted 15% each. The exact number of cells (10.000 cells approx.) were directly sorted into the appropriate volume of FACS buffer to create the 10X libraries in the least amount of time. The FlowJo v.10.7.2. programme was used to analyse and represent the FACS data, generating the plots showing the sorting strategy.

1.4 RNA extraction and quantitative reverse transcription polymerase chain reaction (RT-qPCR) for the quantification of immediate early genes (IEGs) expression

The RNA extraction from mouse oviducts was carried out with the RNeasy® Plus Micro kit (Qiagen, 74034). From the extracted RNA in each sample, 1 μg was reversed transcribed into cDNA with the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher, 4368814). The thermal cycler programme used can be found in Table 2, pg 47. The qPCR was performed with the kit SYBR™ Green PCR Master Mix (ThermoFisher, 4309155). The primers used for this analysis were designed to amplify the mouse genes *Fos*, *Junb*, *Actb* and *Gapdh* (see sequences below), and the thermal cycler programme used can be found in Table 3, pg 47. Quantification of *Fos* and *Junb* gene expression in test samples was performed using the $\Delta\Delta$Ct method normalising them against the expression of a control sample (oviducts where RNA was extracted immediately after harvesting and without any tissue processing) and internally against the level of the two housekeeping genes *Actb* and *Gapdh*.

List of primers used:

Gapdh:

Forward: 5'- GGTGAAGGTCGGTGTGAACG -3'

Reverse: 5'- ACCATGTAGTTGAGGTCAATGAAGG -3'

Actb:

Forward: 5'- GTTCCGATGCCCTGAGGCTC -3'

Reverse: 5'- CGGATGTCAACGTCACACTTCAT -3'

Fos:

Forward: 5'- TGTCTACCCCTGGACCCCTT -3'

Reverse: 5'- TGCGCAAAAGTCCTGTGTGTT -3'

Junb:

Forward: 5'- GAAGACCGTGTCAGGAGGCA -3'

Reverse: 5'- CTCCTCCGATTCCGGCACTT -3'

	Step 1	Step 2	Step 3	Step 4
Temperature (°C)	25	37	85	4
Time (min)	10	120	5	∞

Table 2. The thermal cycler programme used for the cDNA reverse transcription

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Temperature (°C)	94	94	56	72	72	4
Time (min)	2	30 s	30 s	1	5	∞
		25 cycles				

Table 3. The thermal cycler programme used for the qPCR of immediate early genes (IEGs)

2 Single-cell transcriptomic atlas of the mouse oviducts and data validation

2.1 Nuclei extraction from mouse oviducts for single-nucleus RNA-Seq (snRNA-seq)

The extraction of individual nuclei for the creation of scRNA-seq libraries was performed with the Nuclei Isolation Kit: Nuclei EZ Prep (Sigma, NUC101). The protocol for these experiments started with the snap freezing of fresh oviducts in liquid nitrogen and their storage at -80°C until use. All the procedures were carried out on ice and at 4°C unless otherwise stated. At the time of the experiment, the oviducts were minced in small pieces, 2 ml of ice-cold nuclei PURE lysis buffer were added to the sample in a glass homogenizer (KIMBLE Dounce tissue grinder set, Sigma) and these were grinded 25 times with pestle A and 25 times with pestle B. After that, the sample was incubated with 2 more ml of ice-cold PURE lysis buffer for 5 minutes. Then the sample was centrifuged at 500g for 5 minutes at 4°C. The supernatant was discarded and 2 ml of Nuclei PURE lysis buffer was added to the pellet followed by a 5 minutes incubation at 4°C. A second centrifugation with the same conditions was carried out and the pellet was resuspended in 2 ml of ice-cold nuclei suspension buffer (NSB). 500 μl of this nuclei suspension was thoroughly mixed with 900 μl of sucrose cushion buffer I. Then, 500 μl of the sucrose cushion buffer I was added to a 2 ml Eppendorf tube and the nuclei suspension was carefully laid on top of it without mixing. This layered solution was centrifuged at 13000g for 45 minutes at 4°C and the supernatant was removed, leaving 100 μl. 1 ml of NSB was added to this volume of pellet and mixed until the nuclei were completely resuspended. The mixture was then passed through a 10 μm strainer to remove cell debris and the concentration was adjusted for a requirement of approx. 10000 input

nuclei for a Chromium Single-Cell 3' library preparation in the volume specified by the 10X protocol.

2.2 10X Genomics

The 10X Genomics Chromium Single-Cell 3' Gene Expression Kits, Chips and Beads for version v2 and v3 were used in this project. Each library was created from 10000 oviduct sorted cells (composition specifications in the FACS cell sorting section) following the 10X Genomics protocol. In short, cells mixed in an enzymatic cocktail, barcoded beads and oil were loaded into the 10X chips to form single cell emulsion droplets and to perform reverse transcription within the drop. During this reverse transcription each cell transcriptome was labelled with a unique barcode and each single cDNA molecule was labelled with a unique molecular identifier (UMI). Next, all the cDNA molecules from all the original cells in the sample were pooled to create the libraries by converting them to an appropriate size and adding sequencing adapters. The created libraries were sent to the external company Novogene to be sequenced in a NovaSeq sequencer (Illumina).

2.3 Pre-processing and initial filtering of scRNA-seq data

The fastq files obtained after single-cell sequencing were analysed with the 10X Genomics cellranger v5.0.0 algorithm using the command cellranger count with the option of exon and intron alignment. The reference genome used was the mouse mm10-3.0.0. The count matrix output was then loaded in R v4.0.3. and the subsequent analysis was carried out in this system. First, a Seurat object was created with the Seurat package v3.9.9.903 (Stuart et al. 2019) and an initial filtering was achieved by removing low-quality cells that exhibited less than 200 detected genes, a percentage of detected mitochondrial genes greater than 5 and/or a gene-UMI correlation lower than 0.9. Also, genes that were not expressed in at least three cells were filtered out.

2.4 Doublet identification and removal

Further filtering was done through doublet identification and removal using the R package DoubletFinder (McGinnis et al. 2019a). DoubletFinder was applied with the default settings to the Seurat objects created from all the scRNA-seq datasets on an individual basis as per the author's recommendation.

2.5 scRNA-seq integration, clustering and differential-expression (DE) analysis

To perform further analyses, all the datasets were integrated following Stuart et al. 2019. For that, all the datasets were first normalised with the SCTransform method and then, 3000 genes were selected for integration of the data. Then, the PrepSCTIntegration function was used to prepare the datasets for integration and

the functions FindIntegrationAnchors and IntegrateData were used to identify anchors and integrate the data, respectively. The data was then dimensionally reduced by principal component analysis (PCA) with the function RunPCA and the uniform manifold approximation and projection (UMAP) was calculated with the RunUMAP function. To perform clustering on the data, two functions were used: (1) the function FindNeighbors (using 50 dimensions), which computes the k.param nearest neighbours of the dataset; (2) the function FindClusters (resolution = 1.4), which identifies clusters of cells by a shared nearest neighbour (SNN) modularity optimization based clustering algorithm. To perform DE analysis on an integrated object, it is necessary to additionally run the standard Seurat processing in the RNA data slot. That is, log-normalising the data, selecting high-variance genes (2000 genes were specified) and scaling the data (I regressed out the variable "percent of mitochondrial genes" so they would not be taken into account when scaling). After that, the function FindConservedMarkers was used to calculate DE genes in each cluster and an associated p-value was generated with the calculations for each gene. In some instances, for large clusters with highly DE markers, the p-values appeared as 0, however, this is an R artefact meaning that, when the p-value is less than 2.225074^{-308}, R cannot display those numbers and they are treated as 0. If after this complete analysis, clusters of bad-quality cells were found downstream, an additional filtering was carried out, filtering those cells out.

2.6 Visualisation of the gene expression and associated metrics

The UMAP plot representations were based on the functions DimPlot from the Seurat package and plotUMAP from the scater package v1.18.6. Feature plots were represented with the FeaturePlot function from the Seurat package or the plotUMAP from the scater package. Violin plots were plotted with the function plotExpression from scater or VlnPlot from Seurat. Heatmaps were created with the DoHeatmap Seurat function. Dot plots were created with the DotPlot function from Seurat. Pie charts were depicted by the geom_col and coord_polar functions of the ggplot2 package. Bar charts were done with the geom_bar function of ggplot2. Lollipop charts were created combining the geom_point and the geom_segment ggplot2 functions. Ridge plots were created with the RidgePlot Seurat function.

2.7 Study of biological terms enriched in specific clusters

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Dennis et al. 2003) was used to further characterise the epithelial clusters by detecting the main enriched biological processes and gene ontology (GO) terms and discovering functional-related gene groups and molecular pathways. The analysis was run keeping the default settings and only biological terms with a p-value <0.05 were considered. This less conservative method is adopted, at the risk of increasing the number of false positives, in order to enable the discovery of new relevant biological processes, which would not be possible when using conservative methods.

2.8 Generation of 3D mouse oviduct organoids

Single oviduct cells were obtained following the same protocol previously described. It is important that all steps in this protocol are carried out on ice to minimise cell damaging and preserve cell viability, to wet the pipette tip that is being used between steps to avoid cell-wall sticking and to keep Matrigel on ice to avoid solidification of the matrix. When working with Matrigel, it should be thawed at a stable temperature of 4°C before use, avoiding temperature fluctuations. After enzymatic dissociation of the oviducts and filtration to obtain single cells, they were washed in cold 1X DPBS, counted and resuspended at a concentration of $5x10^5$ cells per ml in Matrigel. 50 μl drops of Matrigel ($2.5x10^4$ cells) were plated on 12-well plates pre-warmed at 37°C. After plating the drops, the plates were flipped to set the drops hanging and placed in the tissue culture incubator for 30 minutes to allow Matrigel gelation. After that time, the solid drops were covered with previously warmed organoid medium and placed in a tissue culture incubator to allow organoid growth. The organoid medium composition is made of DMEM F12, HEPES (1:100), GlutaMax (1:100), 1% Penicillin-Streptavidin [ThermoFisher, 15140122], B27 [ThermoFisher, 17504044] (1:50), EGF [Sigma, 11376454001] (0.1 μg/ml), Y-27632 Rock inhibitor [Sigma, Y0503] (10 μM). Medium was replaced with pre-warmed, freshly prepared one every three days.

2.9 Passage of 3D mouse oviduct organoids

First, the organoid medium is removed and the wells washed with cold 1X DPBS. Second, a 1X Trypsin solution is added to every well with enough volume to cover the Matrigel drops. Then the Matrigel and the organoids were dissociated through direct pipetting of the drops in each well at least 30 times. Subsequently, the organoid-trypsin mixture was transferred to a 1.5 ml Eppendorf tube and placed in a 5382 ThermoMixer C v.3.0.2 at 37°C for 20 min at 800 rpm vibration speed. The mixture was thoroughly pipetted 15 times once at minute 6.5 and next at minute 13. The mixture containing trypsin and cells was centrifuged at 400g for 10 minutes at 4°C and the supernatant was then discarded. The formed cell pellet was resuspended in medium containing fetal bovine serum (FBS) to neutralise any trypsin leftovers and cells were centrifuged again with the same specifications as before. Supernatant was discarded and cells washed once more with 1X DPBS. Then, the pellet was resuspended at a concentration of $5x10^5$ cells/ml in Matrigel and 50 μl drops were plated as previously described.

2.10 Immunofluorescence staining

To carry out immunofluorescence on mouse organoids, these were generated directly in 8-well chamber slides (Thistle Scientific). Once the organoids were grown, the medium was removed and the samples were briefly washed in ice-cold 1X PBS. Then, they were fixed in cold 4% paraformaldehyde (PFA) for 10 min. After fixation, the wells were washed three times with ice-cold PBS to remove any PFA leftovers and organoids were permeabilised with a permeabilization buffer (1X PBS, 0.5% Triton X-100) for 10 min at room temperature (RT). Subsequently, they were washed

with wash buffer (1X PBS, 0.2% Triton X-100, 0.05% Tween-20) and incubated in blocking solution (wash buffer, 1% BSA) for 1 hour. The used primary antibodies were the mouse monoclonal anti-β-Tubulin IV (Tubb4) antibody (clone ONS.1A6, reference number T7941, Sigma; 1:200 dilution in blocking buffer) and the rabbit polyclonal anti-Pax8 Antibody (10336-1-AP, Proteintech; 1:50 dilution in blocking buffer). These antibodies were incubated with the organoids at 4°C overnight. Then, the wells were washed three times with washing buffer and subsequently incubated with the secondary antibodies donkey anti-Rabbit IgG (H+L), Alexa Fluor 594 [ThermoFisher, A-21207] and donkey anti-Mouse IgG (H+L), Alexa Fluor 488 [ThermoFisher, A-21202] (both 1:1000 in blocking buffer) and DAPI (1:1000 in blocking buffer) at RT for 1 hour. Finally, the stained organoids were washed three times with washing buffer and mounted in mounting medium without DAPI (Vector Laboratories). They were imaged using an inverted confocal microscope (Zeiss 780). The same staining protocol was used for oviduct tissue sections. First, the oviducts were snap-frozen in liquid nitrogen, they were embedded in the cryo-embedding media OCT and then frozen at -80°C. Sections of 10 μm were cut with a cryostat and kept at -80°C until use. On the day of staining, the slides were brought to room temperature and washed in ice-cold 1X PBS to remove any OCT leftovers. A hydrophobic pen (PAPpen, Abcam) was used to draw circles around the samples on the slide (to minimise working volumes), which was left to dry for 5 min at RT before proceeding with the same staining steps described for the organoids. Slides were ultimately mounted with a coverslip before imaging.

2.11 Immunohystochemistry (IHC)

To carry out IHC, mouse oviducts were fixed in ice-cold 4% PFA for 24 hours at 4°C and then processed and embedded in paraffin. The paraffin blocks were sectioned into 2.5 or 4 μm sections and Heat Induced Epitope Retrieval (HIER) in retrieval buffer (pH 6) was performed. IHC was performed using a Leica Bond Max autostainer (Leica Microsystems) or an Autostainer plus Link 48 (Dako). Primary antibodies were incubated for 1 hr and the following products were used: anti-Pax2 polyclonal antibody produced in rabbit (1:200; Biolegend, 901002), anti-Cd13 polyclonal antibody produced in rabbit (1:500; Proteintech, 14553-1-AP). Counterstaining with haematoxylin to label the cell nuclei was performed after detection of the primary antibody with an HRP-conjugated anti-rabbit secondary antibody and subsequent DAB (3,3-Diaminobenzidine) reaction. Finally, an Aperio slide scanner (Aperio) was used to scan the stained slides at 40X magnification.

3 Cross-species comparison of oviduct populations composition and interactions in mouse and human

3.1 Publicly available normal human fallopian tube scRNA-seq datasets

The SMART-Seq2 produced scRNA-seq dataset from Hu et al. 2020 was retrieved from the Gene Expression Omnibus (GEO), National Center for Biotechnology Information (NCBI), with the GEO Accession number GSE139079. The single-cell data produced by the 10X Genomics protocol by Dinh et al. 2021 was retrieved

from GEO, NCBI, with the GEO Accession Number GSE151214.

3.2 Patient samples

The two samples used in this project come from donors recruited under the Gynecological Oncology Targeted Therapy Study 01 (GO-Target-01, research ethics approval #11-SC-0014), with all patients appropriately informed and consented before surgery.

3.3 Human 10X Genomics and data analysis

The followed protocol was the same described for mouse samples and two scRNA-seq libraries were created from the fallopian tubes of two pre-menopausal patients. The only differences were the use of the human reference genome hg19-3.0.0. when using the cellranger count v5.0.0 algorithm, the number of datasets that were integrated using the Seurat package (the SMART-Seq2, the 10X Genomics dataset by Dinh et al. 2021 and the 10X Genomics dataset created in this project) and the specification of the resolution parameter for FindClusters (resolution = 1.2). These experiments were all done in collaboration with the DPhil student Zhiyuan Hu, except for the FACS cell sorting, which she carried out independently.

3.4 Statistical correlation analysis and generation of the correlation matrix

The statistical correlation using the Pearson and the Spearman method was computed with the cor function from the stats package v4.0.3. The correlation matrix was generated with the corrplot package v088. The correlation matrix was computed using the averaged normalised expression value of highly variable genes. For that, the 2000 most highly variable genes across oviduct cells and across fallopian tube cells were identified with the FindVariableFeatures algorithm from the Seurat package. Only the common genes to both species were kept (271) and, of these genes, the 50 genes with the highest variances across the oviduct and fallopian tube cell clusters were selected.

3.5 Cell-cell ligand-receptor analysis with CellPhoneDB

This algorithm requires two files as input data: (1) the scRNA-seq normalised count data containing gene expression values with rows as genes and columns as cells and (2) an annotation file containing two columns called 'Cell', indicating the name of the cell, and 'cell_type', indicating the name of the cluster that the cell belongs to. Both files were extracted from the Seurat object corresponding to the oviducts and the one corresponding to the fallopian tubes. Then, the analysis involving CellPhoneDB was run on the Computational Biology Research Group (CBRG) supercomputational cluster. For that, the commands "module load python-cbrg" followed by "cellphonedb" were used to load

Python and the CellPhoneDB algorithm. Then the virtual environment was activated with "source cpdb-venv/bin/activate". For the mouse data the analysis command "cellphonedb method statistical-analysis mouse_cellphonedb_meta.txt mouse_cellphonedb_count.txt –counts-data=gene_name" was used. However, for the human data the subsampling feature was used "cellphonedb method statistical-analysis cellphonedb_meta.txt cellphonedb_count.txt –counts-data=gene_name -subsampling –subsampling-log false" given that the number of cells was too big to complete the analysis.

4 RNA-Seq and correlation analyses

4.1 Download of high-grade serous ovarian cancer (HGSOC) transcriptomic data

The AOCS (Australian Ovarian Cancer Study) dataset (Tothill et al. 2008, GEO Accession Number GSE9899) was obtained from the GEO, NCBI. The ovarian cancer TCGA (The Cancer Genome Atlas) dataset (Cancer Genome Atlas Research Network 2011) was downloaded from UCSC Xena Data Hub (https://tcga.xenahubs.net); Broadfirehose (https://gdac.broadinstitute.org).

4.2 Generation of dot plots, violin plots and scatter plots for the correlation analysis

For the dot plots, the function ggplot and geom_point from the package ggplot2 v3.3.3 were used. The violin plots regarding the tumour transcriptomic data were created with the function ggplot, geom_violin, geom_boxplot and geom_jitter from the ggplot2 package. The violin plots depicting gene expression from scRNA-seq datasets were computed with the VlnPlot function from the Seurat package. The scatter plots were created with the function ggscatter from the ggpubr package v0.4.0.

4.3 Statistical correlation analysis and p-value calculation

P-values for the differential expression analysis were calculated with the non-parametric Wilcoxon test. For that, the function wilcox.test from the stats package v4.0.3 was used. The alternative hypothesis was specified as "greater" to test whether the median gene expression of EMT-high tumours is greater than the median gene expression of EMT-low tumours. Correlation analysis were run with the function ggscatter and the spearman method.

IV Results

1 Chapter 1. Technical optimisation of oviduct single-cell isolation

1.1 Introduction

In the last decade the scientific community has witnessed an unprecedent revolution in the development and implementation of state-of-the-art technologies such as NGS and their corresponding single-cell approaches. The progress achieved in the single-cell transcriptomics field has translated into the generation of a range of powerful high-sensitivity scRNA-seq platforms, widely used by interdisciplinary research approaches and responsible for exceptional discoveries and innovation, in both, clinical and biological research areas (Hwang et al. 2018). This technological advancement, supported by the exponential genesis and optimisation of bioinformatics resources, has enabled the success of scRNA-seq tools in providing long-sought insights into the non-genomic cell heterogeneity, the interaction landscape and the associated driving biological/deregulated processes of a vast range of tissues, species and pathologies (Wagner et al. 2016, Jindal et al. 2018, Aldrige et al. 2020, Hie et al. 2020).

In this project, scRNA-seq is utilised to accomplish a non-genomic cell heterogeneity investigation of the mouse oviducts and the human fallopian tubes, and to identify potential cell populations interactions and developmental hierarchies. The general workflow in a scRNA-seq experiment includes the initial stages of tissue/organ dissection as well as careful handling and processing to obtain viable single cells (Figure 8, pg 55). Subsequently, the cells can optionally be enriched for specific subpopulations before being used as input material for RNA extraction and library preparation. Finally, the transcriptomic molecules generated through library preparation undergo sequencing, which generates high-throughput data subject to bioinformatic analysis. Single-cell RNA sequencing is a complex and highly sensitive technology, which requires detailed precision at every step of the experimental execution. Hence, in order to maximise the chances of a successful single-cell RNA library preparation and the consequent high-quality data acquisition, several experimental variables should be carefully accounted for, in any robust research strategy (Nguyen et al. 2018).

First, the organ of interest is identified and carefully harvested. Subsequently, the tissue should be appropriately dissociated into the maximum number of individual cells exhibiting optimal viability (Reichard and Asosingh 2019). To this end, it is important to understand the tissue composition and architecture of the target organ. In terms of composition, the oviducts are solid tubes comprised of an internal epithelial lining attached to an ECM, which contains embedded mesenchymal cells, and is surrounded by a smooth-muscle layer and an external mesothelial layer. The ECM is a non-cellular network, which supplies not only a scaffold and structural foundation to the tissue, but also triggers and transduces biological signalling (Theocharis et al. 2016, Frantz et al. 2010). The primary components of the connective matrix are collagens, which mediate the establishment of cell-matrix connections, although proteoglycans such as decorin and versican, and glycoproteins like fibronectin, elastin and laminin, are also key elements of the ECM (Halper et al.

2014). In the oviducts, the ECM is present in the basement membrane, a specialised layer lying between the connective meshwork and the basal domain of the epithelial sheet. The majority of cells, including the epithelium, are known to synthesise and secrete components of the basement membrane. Besides the ECM and epithelial components, the oviducts are also infiltrated by blood and lymphatic vessels as well as immune cells.

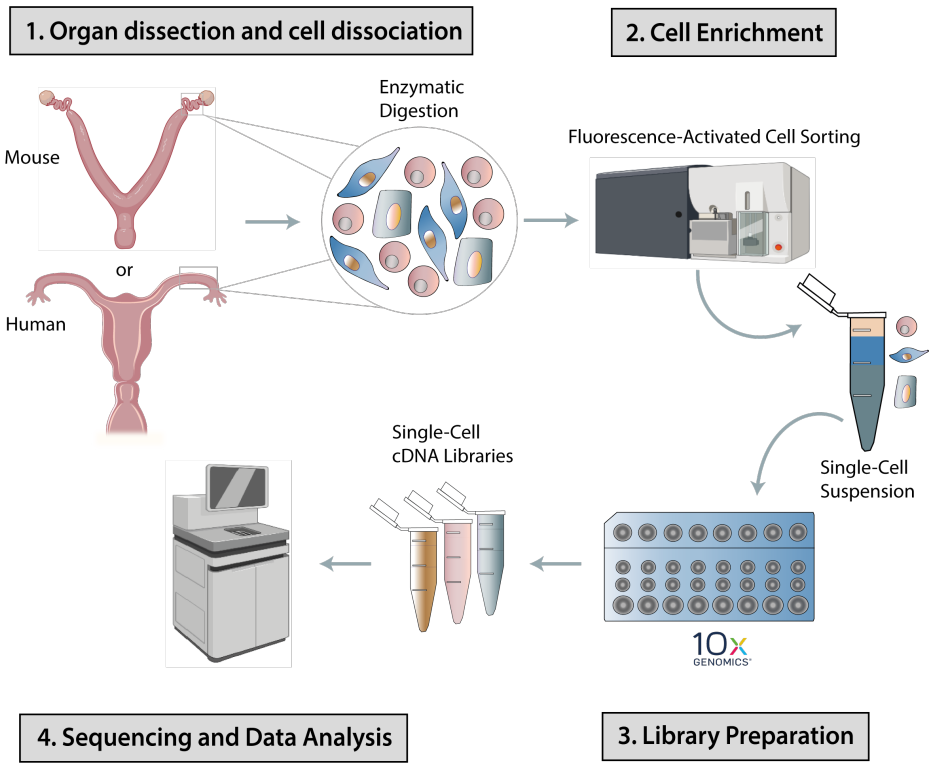


Figure 8. Single-cell RNA sequencing experimental workflow with mouse oviducts or human fallopian tubes
First, the organ of interest is isolated and the cells composing the tissue are enzymatically dissociated and individually separated. Second, a viability control and populational enrichment through FACS is performed and only the cells passing the threshold imposed by the acceptance criteria are sorted. These cells represent the material input for single-cell RNA library preparation, which will be subsequently converted to cDNA libraries subject to sequencing. The output readouts are then analysed following a bioinformatics pipeline

Overall, these diverse cell lineages comprise the oviduct population landscape, however, in order to establish the 3D architecture of the organ, those components organise into tissues through the establishment of a stable and a solid connective net. The connective repertoire of the oviducts is dominated by two main communication axis: cell-matrix adhesions as well as intra and inter-lineage cell-cell interactions (Figure 9, pg 57). Typically, the latter are established by a membrane adhesion receptor binding cytoplasmic plaque proteins, which act as a connecting bridge with the cell cytoskeleton. Cell-ECM or basement membrane adhesions are regulated by adhesion receptors such as integrins, the proteoglycans syndecans, collagens or hemidesmosomes. Hemidesmosomes are basal cell domain-basement membrane unions formed by the adhesion receptors integrin $\alpha 6\beta 4$ binding to laminins from the basement membranes (Walko et al. 2015). This receptor then binds the

cytoplasmic plaque protein plectin, which connects the protein complex to the intermediate filaments of the cytoskeleton, known as keratin or cytokeratin filaments in the epithelium. With respect to cell-cell coupling, they are connections established between adjacent cells and can be classified as tight or occluding junctions, gap junctions, adherens junctions and desmosomes (Gumbiner 1996, Zihni et al. 2016). Tight and adherens junctions generally localise to the intermediate regions between the apical and basolateral sides of the epithelial cells, desmosomes and gap junctions localise on the basolateral membrane. Tight junctions are barriers selectively controlling material transport across neighbouring epithelial cells based on parameters such as size and charge. They form extremely proximal unions of the adjacent plasma membranes and physically delimit the apical and basolateral territories of the membrane of epithelial cells, preventing the merge of different membrane-compartment components. These connections are mediated by a varied array of adhesion proteins which are integral membrane components: claudins, occludins, tricellulin, blood vessel epicardial substance (BVES), junctional adhesion molecules (JAMs), coxsackievirus adenovirus receptor (CAR) and angulins (Zihni et al. 2016). They interact with the cytoplasmic plaque proteins ZO-1, 2 and 3, which attach them to the cytoskeletal actin filaments (Zihni et al. 2016). Adherens junctions are connections composed of cadherins, which are $Ca2^{+}$-dependent adhesion receptors, of which, the most broadly studied in epitheliums is E-cadherin (Gumbiner 1996). To form the adherens junctions, E-cadherin establishes a complex with actin filaments of the cytoskeleton through catenins and cytoplasmic plaque proteins. A more organised and calcium-independent cell-cell junctions are the desmosomes, which are structures formed by the link of cadherins (desmogleins and desmocollins) to cytoplasmic plaque proteins (desmoplakin and plakoglobin) and these to cytokeratins (Gumbiner 1996). Finally, gap junctions are intercellular connections consisting of membrane pores established by the connexin proteins and acting as channels for the transport of solutes and metabolites between adjacent cells (Evans and Martin 2002).

Taken together, solid tissue samples are compact networks comprised of cellular layers and an extracellular matrix, rich in physical adhesions between these two components. An appropriate tissue dissociation requires the disintegration of the cell-cell and cell-matrix attachments, while preserving single-cell integrity. Although this is generally achieved through enzymatic digestion, the details of this step are strongly dependent on sample nature and should be optimised based on organ-specific variables such as type of intercellular interactions and ECM composition.

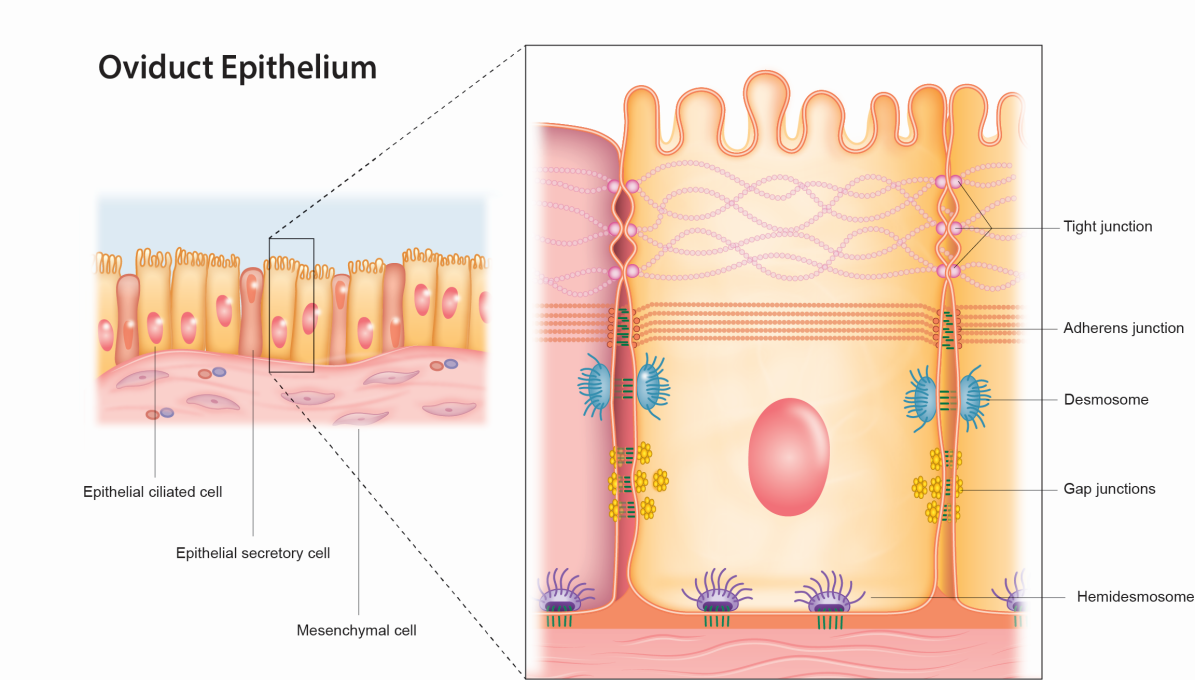


Figure 9. Cell-cell and cell-matrix connections in the oviduct epithelium
The connective landscape of the oviducts is dominated by cell-cell and cell-ECM contacts. The former are established through desmosomes, gap junctions, adherens junctions and tight junctions, whereas the latter are mediated by hemidesmosomes, integrins, syndecans and collagens

Beyond the identification of proper dissociation and processing conditions, this project specifically aimed to catalogue cellular heterogeneity of the oviduct. Consequently, a balanced and comprehensive representation of the diverse cell lineages composing this organ was paramount and, for that, performing cellular enrichment for underrepresented cell lineages was taken into consideration. Specific surface marker-directed cellular enrichment can be accomplished by the use of techniques such as FACS and magnetic bead-based cell selection (MACS). FACS principles are based on flow cytometry, a high-throughput technique, which relies on the light-scattering properties of individual cells in a heterogeneous sample and measures their size and granularity, or complexity (Adan et al. 2017, Vembadi et al. 2019). When coupled with fluorescence, flow cytometry is able to classify and separate different cell phenotypes as cells can be directly marked with a dye, for example, DAPI to account for cell death, or a dye-labelled antibody targeting a marker antigen. Similarly, MACS is based on antibody labelling of cells, but in this case the antibodies are coated over superparamagnetic beads, with marker-positive cells trapped on the beads surface and then separated from the rest of the sample through a magnetic field. Cell enrichment methods, along with the previous single-cell preparation stages, constitute a substantial source of unsought technical artefacts, which can disturb the original cell expression signatures and originate batch effects. Hence, to ensure that the obtained data are valid, confounding factors should be identified and controlled to the maximum extent. Typically, a procedure-induced cellular stress response accounts for one of the most important variables to be minimised when dissociating and sorting primary samples into single cells (Box et al. 2020, van den Brink et al. 2017, Denisenko et al. 2020, O'Flanagan et al. 2019, Machado

et al. 2021, Adam et al. 2017). These changes can be triggered by tissue enzymatic dissociation as well as flow cytometry, and the artificially induced genes can affect several cellular programmes, ultimately altering the native transcriptional profile of the analysed cells.

Therefore, given the nature of this project, the first critical objective to achieve was the optimal preparation of a representative and viable single-cell suspension from the mouse oviducts. For that, a dissociation protocol was designed, tested and extensively optimised to meet the quality requirements for the subsequent RNA library preparation experiments. In this chapter, the focus is strictly limited to the optimisation work carried out prior to the library preparation step. Further optimisations and implementations of technical variability control strategies executed at later stages, such as batch effect control, will be presented in the following chapters. During the pre-library preparation stage, variables such as enzyme type and dissociation time, along with cell-type enrichment technologies, were assessed to find the best compromise between a viable single-cell dissociation, minimisation of confounding factors and an adequate population representation of the mouse oviduct cells, which can enable the generation of high-quality data for an accurate biological interpretation. To this end, the first addressed points were the creation of a protocol to dissociate and obtain viable single cells from the oviducts, the maximisation of cell recovery and the identification and prevention of confounding factors during the technical processing of the oviducts.

1.2 Enzymatic dissociation followed by FACS enrichment performs efficiently for the population-balanced isolation of viable oviduct single cells

An ideal cell suspension for downstream single-cell RNA-Seq experiments should contain highly-viable cells with good cell-surface marker expression and minimal presence of cell death, cell debris, cell aggregates and clumps. The target for these experiments was to obtain at least 10.000 oviduct-derived individual cells for every RNA library. Therefore, in order to establish a protocol for single-cell preparation from the oviducts that yields satisfactory quality and consistent reproducibility, the first addressed steps in this project were the organ dissection, mechanical dissociation and enzymatic digestion. First, the oviducts were identified as the funnels connecting the ovary and the uterine horns in mice and were micro dissected, rinsed, and mechanically cut into the smallest possible fragments to expose the luminal side, increasing the epithelial surface area to the action of the digestive enzymes used. After successful oviduct isolation and mechanical dissociation, a first enzymatic cocktail was tried. Considering the tissue architecture and structure of the oviducts, crucial components of a dissociation mixture should be enzymes able to hydrolyse the ECM and the cell-cell adhesion proteins, as well as, deoxyribonucleases (DNAses) to digest DNA molecules released by apoptotic or disintegrated cells that can cause cellular clumping and aggregation. The first trial was performed with an enzymatic formulation pre-established in the lab. This contained: Collagenase Type I, to break down the ECM by digesting the peptide bonds making up collagen; Trypsin, a member of the serine protease family that cleaves lysine and arginine-mediated cell-cell protein bonds, and DNase I. This blend, allowed the separation of discrete oviduct

cells after 45 minutes of incubation at 37°C. However, lack of reproducibility on the obtained cell numbers and inconsistency between replicate experiments was detected. This might have been due to the use of crude and impure Collagenase I and Trypsin enzymes, as they were derived from bacteria. Advantages on stability and effectiveness during tissue dissociation had been reported when using commercially purified cocktails instead (Cavanagh et al. 1997), which along with the fact that trypsin had also been determined to have a harsh effect on tissue epitopes (Khan et al. 2016, Huang et al. 2010) and to trigger RNA degradation (Vrtačnik et al. 2014), prompted the exploration for an alternative formulation.

Generally, to overcome these undesirable effects and to reduce the composition variability and presence of undefined factors, commercial mixtures are preferred. Commercial preparations are carefully purified and manufactured under controlled conditions so the formulations are endotoxin-free and present less or absent lot-to-lot variability on enzymatic activity. According to the literature and technical brochures, the commercial mixture Liberase^{TM} has routinely been used to successfully dissociate tissues into highly viable cells (Panchision et al. 2007, Dolmans et al. 2006, Vanacker et al. 2011). Therefore, I then tested enzymatic dissociation with Liberase^{TM} and DNAse I, at 37°C, with an incubation time of 45 minutes. This strategy resulted into a balanced single-cell dissociation, which delivered more reproducible cell numbers and viability (range 71-97%).

To further characterise the distinguishing features of the oviduct preparation and the major cell populations, flow cytometry and FACS were carried out on samples labelled with antibodies against the epithelial cell adhesion molecule (EpCAM), as an epithelial lineage marker, and the protein tyrosine phosphatase receptor type C (PTPRC or most commonly known as CD45) for the identification of cells of hematopoietic origin (e.g. blood cells). EpCAM is a ubiquitous marker expressed by epithelial cells, whereas CD45 is expressed by most blood cells (e.g. leukocytes), except for mature erythrocytes and platelets. Additionally, cells were stained with a viability marker, DAPI, to exclude dead and dying cells. This staining works under the principle that alive cells are impermeable to DAPI as they maintain membrane integrity, whereas cells that are dead or in the process of dying, have a compromised membrane, which DAPI is able to diffuse through, accessing the nucleus and binding to the DNA. The gating strategy and the results obtained from FACS are represented in (Figure 10, pg 61).

First, to characterise the cells of interest and distinguish them from the cellular debris, a forward (FSC) versus side scatter (SSC) density plot is used, with the first representing an indirect measure of cell size and the second of cell granularity or complexity. On the FSC/SSC plot, two main populations can be identified, one displays a range of cell sizes and complexities, while the second one sits at the bottom left corner of the plot, exhibiting small sizes and relatively limited granularity. The first corresponds to a real heterogeneous oviduct cell population, whereas the latter can be identified as cell debris. Only the oviduct cell populations (˜70%) are gated and selected for a second analysis step, where a forward scatter height (FSC-H) versus forward scatter area (FSC-A) density plot is used to identify single cells and exclude cell doublets. The majority of these experiments showed a high singleton percentage of 95-98%. Once single cells have been selected, a third analysis step is performed combining FSC-A with DAPI, to generate a density plot

used to classify cells based on their viability. When single cells are displayed in the DAPI/FSC-A density plot, three main populations can be identified. On one hand, there is a size-heterogeneous $DAPI^-$ population and, on the other hand, there are two $DAPI^+$ populations. One of the positive groups shows intermediate fluorescence intensity, whereas the second one displays higher values. The intermediate population is characterised by a relatively small size, while the high-intensity population shows a wider size distribution. The intermediate population can be classified as dying or fragmented cells that have uptaken a certain amount of DAPI, whereas the brighter population represents dead cells from all the oviduct cell lineages. Given the downstream experimental purpose, the dead/alive gating strategy followed in this project was strict (71-97% gated cells depending on experiment), as only good-quality highly-viable cells could be used for quality single-cell RNA library preparation. Alive single oviduct cells were then distinguished by their lineage using a two-parameter density plot based on anti-EpCAM-APC versus anti-CD45-FITC labelling. This characterisation showed a very neat separation of a well-defined immune population of $CD45^+$ leukocytes (8-20%) and a $EpCAM^+$ group (3-15%). Both categories make up a modest percentage of the total population and double positives were extremely rare or non-existent. The double negative cells located at the bottom left corner are $EpCAM^-$ $CD45^-$ cells and represent other oviduct cell lineages such as mesenchymal, endothelial, pericytes and mesothelial types. To thoroughly investigate the $EpCAM^+$ and $CD45^+$ subpopulations distribution further, these were backgated to determine their density on the original FSC vs SSC density plot. This analysis identified two subpopulations within the epithelial lineage, the first one presented a range of sizes and granularity, whereas the second presented a more uniform granular profile and varied in size. The first population could correspond to secretory cells as they are known to exhibit internal granules that are secreted into the oviduct lumen. Depending on the level of secretion, secretory cells have been found in different sizes, from bulging cells containing a high number of granules, to small cells that have released their cytoplasmic content. The second population could correspond to ciliated cells, which would display smaller sizes and less granularity variety. Regarding $CD45^+$ cells, two subpopulations were also identified by back gating to FSC vs SSC. One exhibited a varying granularity and uniform size, whereas the second displayed substantial size and granularity heterogeneity.

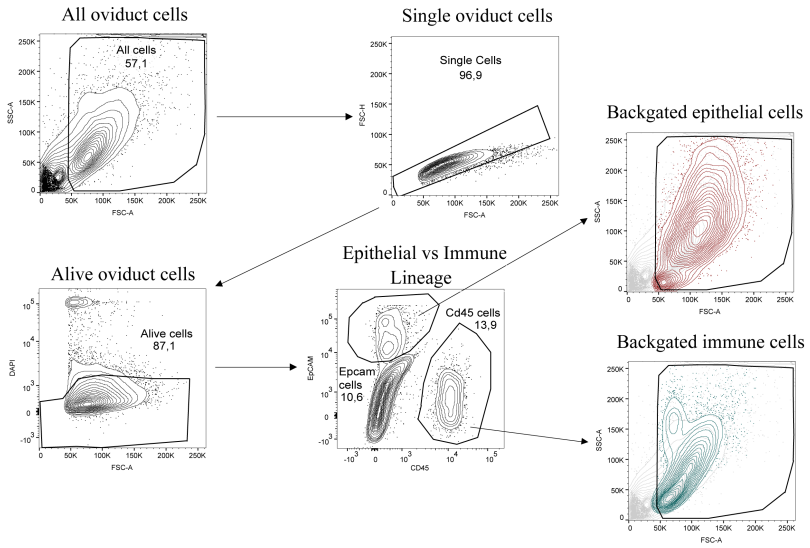


Figure 10. Gating strategy for single-cell FACS sorting of mouse oviducts
The first plot depicts a FSC/SSC gating to distinguish between debris and cells. Cells are then gated and separated into singletons or doublets (FSC-H/FSC-A) and only single cells are subject to a viability analysis through a DAPI/FSC-A analysis. Cells that are alive are negative for the dye DAPI and are selected for further characterisation based on EpCAM/CD45 positivity. $EpCAM^+$ and $CD45^+$ are further backgated to FSC/SSC to observe their distribution within the original sample

Overall, the FACS technique allows for the control of key variables for single-cell RNA profiling experiments such as cell viability, population representation, multiplet exclusion and single-cell separation in the final cell suspension. Furthermore, it permits tight control of cell input concentration, which is essential for the library preparation protocol used in this project. The FACS approach allowed the identification of underrepresented oviduct cell types after enzymatic single-cell dissociation with special attention put on the low yield of epithelial cells (3-15%). In this project, the appropriate study of these cells is paramount as they constitute the organ parenchyma, central for its functionality. Therefore, the low epithelial recovery was a major limiting factor to be tackled. In the native oviducts, approximately half of the cells belong to the epithelial lineage. This fact was validated by analysing oviduct sections stained by hematoxilin and eosin (H&E) or IHC sections with the QuPath algorithm, which identified an average 48.54% of epithelial component across the entire oviduct (Figure 11 A, pg 62). The epithelium/stroma ratio highly depended on the histomorphological sections analysed, with epithelial percentage ranging from as little as 20% to as high as 70% (20,56% [SD (Standard Deviation) 7,14] in isthmus, 48,74% [SD 11,46] in ampulla, 54,93% [SD 5,62] in infundibulum and 69,92% [SD 11,28] in fimbriae) (Figure 11 B, pg 62). Hence, the small percentage of $EpCAM^+$ cells captured by FACS denotes a high loss of epithelial cells, probably due to lineage-specific sensitivity to the dissociation/selection process or tighter intercellular connections, which would lead to cell loss due to increased cell death or poorer dissociation to single cells. To compensate for this, substantial epithelial (70%) and partial immune (15%) FACS enrichment of the final single-cell suspension was adopted in the protocol, with the remaining 15% filled

with EPCAM$^-$ CD45$^-$ cells.

After the initial steps had been set up and a thorough characterisation of the oviduct populations had been undertaken, the dissociation process was further optimised in terms of incubation time with the enzymatic cocktail, an important variable with clear impact on cell viability. To determine the effect of cell dissociation time, the same experiment was carried out in parallel for different samples undergoing enzymatic digestion for 15, 30 and 45 minutes, respectively. Although the general cell viability did not significantly differ among those processing times, it was noticed that epithelial cell recovery decreased with increasing incubation time, likely as a consequence of a viability loss of this cell population due to its higher sensitivity to the dissociation conditions (Figure 11 C, pg 62). Following these results, a time of 30 min was chosen as the best compromise between dissociation efficacy and preservation of the epithelial cells viability. After a 30 min incubation, a number of 1000-3000 alive individual epithelial cells per mouse could be sorted by FACS (500-1500 per oviduct). As I was aiming to have a final epithelial cell content of 70% per sample (7K cells per library prepared), a higher number of mice could have been used to hit the target. However, before this, further optimisation was attempted to maximise the obtained epithelial ratio per mouse and minimise the use of animals.

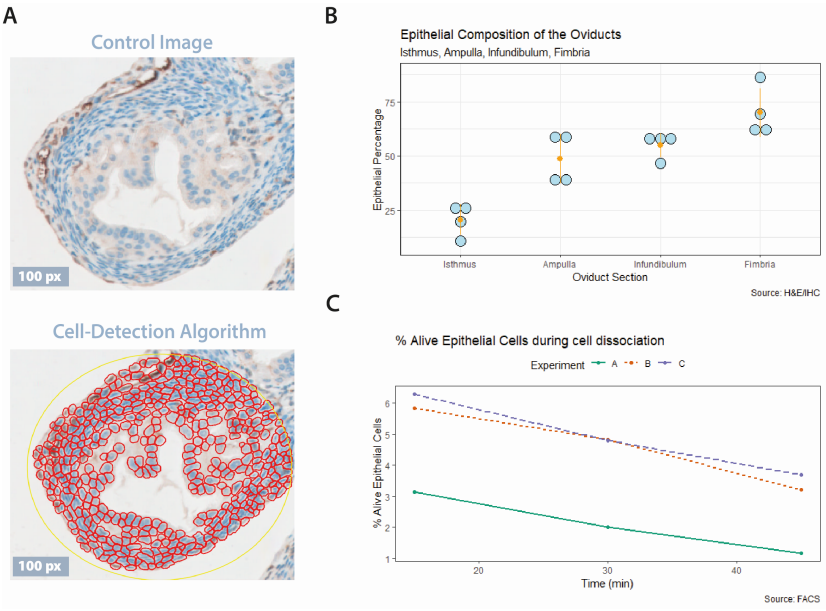


Figure 11. Characterisation of the epithelial loss during oviduct dissociation and FACS sorting

A) The QuPath cell-detection algorithm was used to estimate the epithelial proportion in each of the oviduct portions: isthmus, ampulla, infundibulum and fimbria. A representative isthmic section before and after application of the algorithm mask is shown. B) Epithelial percentage estimates of each oviduct part (blue dots) are shown. Four sections per oviduct portion were analysed and the computed mean and standard deviation are shown as an orange dot and bars, respectively. C) Percentage of viable epithelial cells for different dissociation times. Three replicate experiments (A, B and C) were performed. Px = pixels.

1.3 Multiple consecutive rounds of single-cell dissociation prove to be the best strategy to maximise the number of isolated epithelial cells

In order to optimise the dissociation protocol to enhance epithelial cell recovery, a different multi-step approach was tried (Figure 12 A, pg 64). In this method, the fresh oviduct portions would be incubated with the digestion medium for four shorter sequential periods of time, and at the end of each given period, the detached single cells and debris would be collected and run by FACS, with the remaining tissue chunks being incubated again with fresh enzymatic solution. The same protocol would be repeated for a total of up to 45 min and every round-derived cell sample would be sorted independently and finally pooled together. This strategy improves enzymatic tissue breakdown, and it better preserves viability of early isolated cells, which instead of waiting in stressful conditions for the rest of the tissue dissociation, are separated and rapidly processed. The epithelial numbers during the first two rounds are higher than at later ones, which seems to confirm the previously observed deleterious effect of long dissociation time on epithelial survival (Figure 12 B, pg 64). Furthermore, the proportion of epithelial cells over the total number of cells increases at later stages (Figure 12 B, pg 64). This is due to the fact that most of the debris and easier-to-detach cell types were depleted from the sample at earlier times, leaving tissue chunks enriched for harder-to detach epithelial cells, such as those placed at the base of the luminal protrusions (Figure 12 C, pg 64). This multiple-round approach proved to be much more efficient than the previously tested 30 minutes protocol (Figure 12 D, pg 64), with an almost five-fold increase in epithelial cell recovery. For this reason, this strategy was adopted for subsequent experiments.

Hitherto, an effective method for the generation of a high quality, single-cell mixture from the mouse oviducts was optimised, overcoming the challenges posed by the more fragile nature and hard-to-reach position of epithelial cells. Nevertheless, there was still one important aspect to be evaluated before proceeding with the analysis of the sorted cells: to what extent cell handling and processing affect the transcriptomic profile of individual cells and trigger technical artefacts, which could confound downstream data analysis.

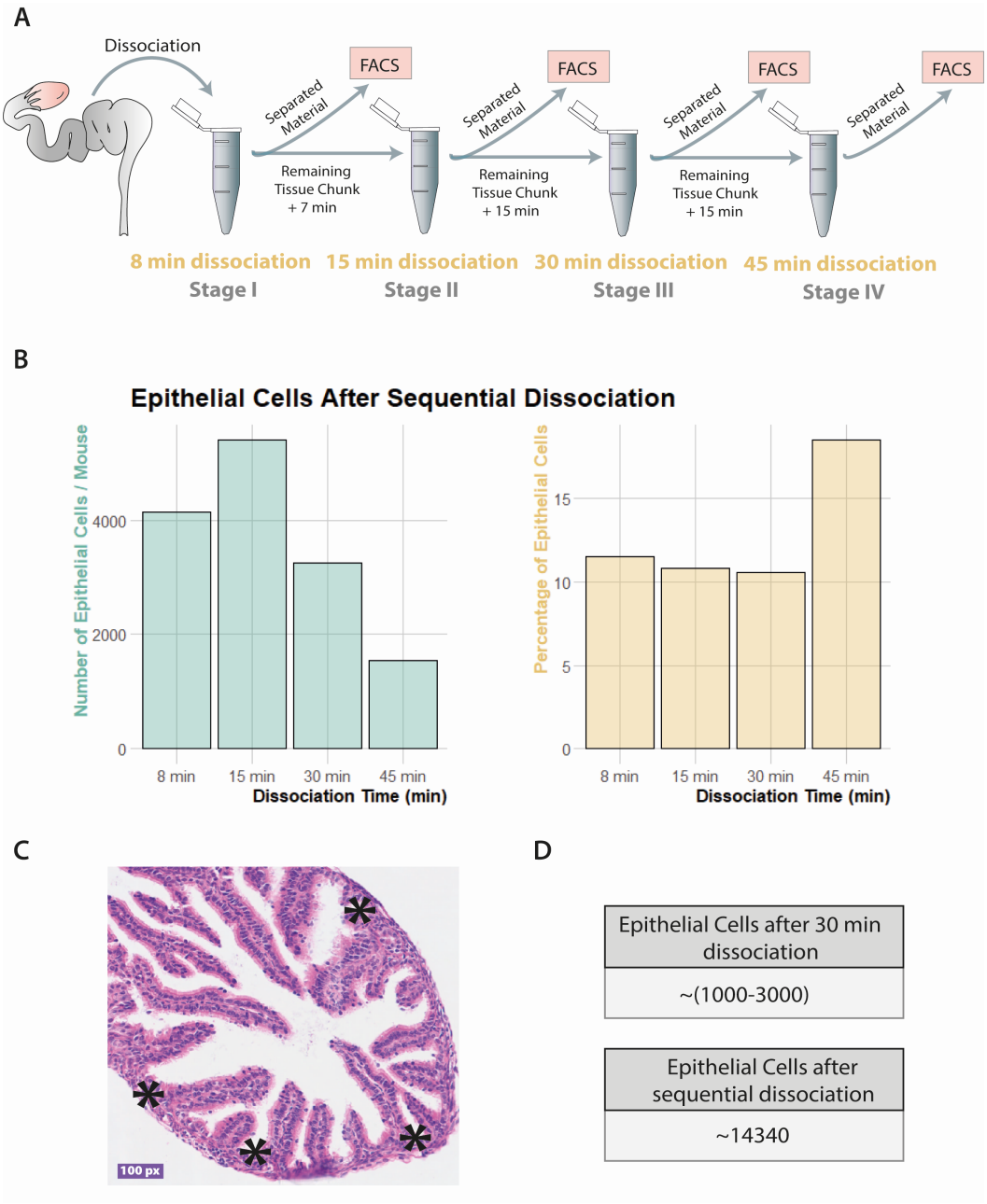


Figure 12. A serial dissociation strategy to enhance epithelial recovery during oviduct single-cell harvesting

A) Schematic representation of the multiple-stage dissociation approach. After every dissociation stage, detached cells are separated from tissue chunks to be FACS sorted, whereas the tissue chunks are placed in new enzymatic medium. B) Representation of the number of epithelial cells sorted after every dissociation stage (green bar graph) and the percentage of epithelial cells out of the alive cells in the sample following every dissociation stage (yellow bar graph). C) Representative ampulla H&E section of the oviducts highlighting zones of difficult reach for the Liberase solution. D) Total number of epithelial cells sorted per mouse after single-step, long incubation protocol (top) and consecutive, multi-round dissociation approach (bottom).

1.4 A confounding stress response dominated by the expression of immediate early genes is detected in oviduct cells after enzymatic single-cell dissociation

Isolation and preparation of primary cells from tissues can be a source of artefacts affecting the transcriptome and the downstream collected data. Whereas FACS sorting has been reported to cause minimal changes in gene expression (Box et al. 2020), the lengthy enzymatic dissociation step at 37°C has often been reported to be a stress source and to disturb the native gene expression profile of single cells from a range of tissues (van den Brink et al. 2017, Adam et al. 2017, Denisenko et al. 2020, O'Flanagan et al. 2019, Machado et al. 2021). This confounding effect is dominated by the strong induction of a set of genes such as the immediate early genes (IEGs) and those coding for heat-shock proteins. In order to understand the effect of the optimised protocol on viable cells and maximise the quality of the acquired data, some libraries were generated and sequenced. A stress signature was then found to be upregulated in all cell types in the oviduct (Figure 13 A, pg 66). To further delineate the cause-effect relationship between the protocol and the detected stress response, a qPCR was performed on oviduct samples following enzymatic digestion for 15, 30 and 45 minutes. Unprocessed oviducts were used as negative control, and the analysis focused on the expression of the IEGs *Fos* and *Jun* (Figure 13 B, pg 66). Their gene expression was not detected in naïve tissue that was not exposed to any processing. However, upregulation of both genes was detected shortly after digestion incubation. The expression of these genes is time-dependant and, following the typical IEG expression pattern, they decreased after 30 minutes of incubation, which might indicate the start of their transcriptional repression.

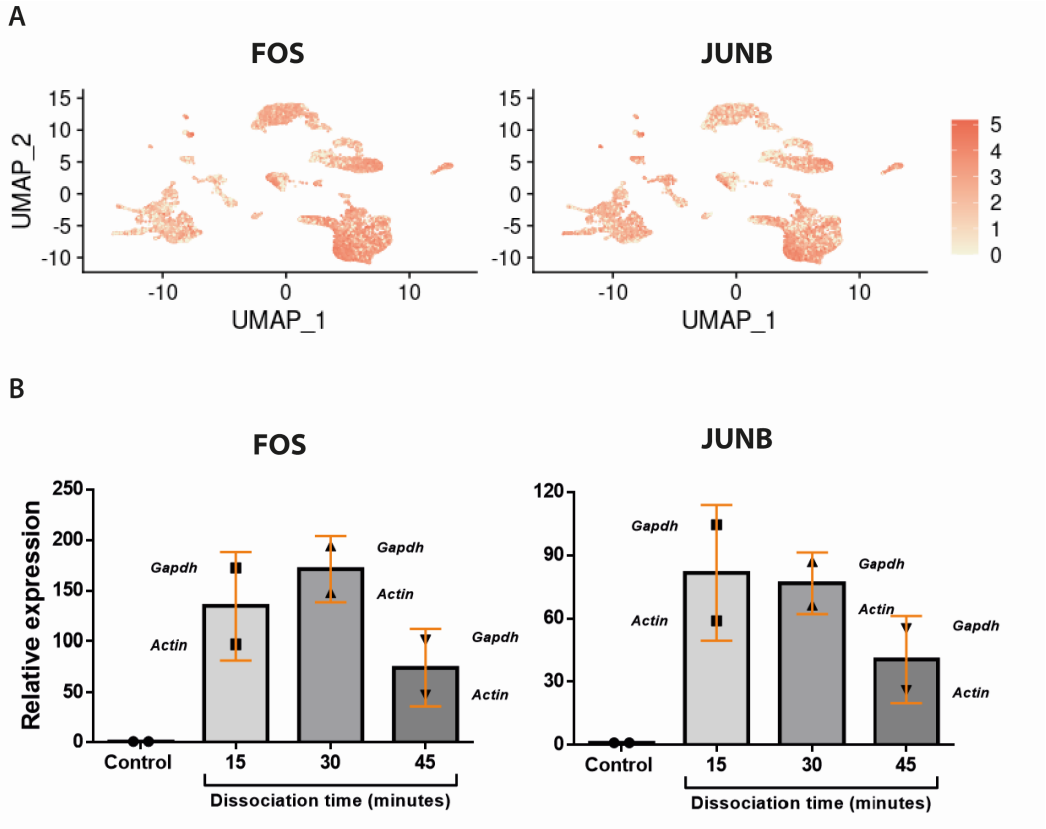


Figure 13. . Detection of a stress response in oviduct cells following Liberase enzymatic dissociation
A) UMAP plot of single-cell sequencing data of mouse oviducts showing the gene expression of the immediate-early genes *Fos* and *Jun*. High expression of both genes is detected in all cells populations. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. B) qPCR analysis of *Fos* and *Jun* gene expression in oviducts upon different dissociation times. Bars indicate the gene expression normalised to control samples (unprocessed oviducts) and for the expression of the two housekeeping genes *Actin* and *Gapdh* (squares/triangles indicate the gene expression normalised for each housekeeping gene as per label). The standard deviation between the two housekeeping gene normalizations is shown in orange.

In order to prevent this technical artefact, an active-cold protease enzyme was identified (Adam et al. 2017) and tried in a follow-up experiment. First, the stress response was measured by qPCR for Liberase^{TM} (37°C) and psychrophilic-protease (6°C) treated oviduct tissue, which showed that the cold protease did not trigger IEGs expression, with levels comparable to the control conditions (Figure 14 A, pg 67). Then, the experiment was repeated and the isolated cells were subjected to FACS, which showed that using the cold protease for cell isolation decreased the cellular yield by more than 50%, as it can be appreciated from the FSC/SSC gating (Figure 14 B, pg 67). Nevertheless, cold-protease dissociated cells exhibited a better cell viability with almost no cell death (DAPI/FSC-A gate). Despite this, the cold protease did not allow the detection of EpCAM^{+} cells, which could be due to its cleaving activity affecting the EpCAM antigen. Therefore, the use of this enzyme as a Liberase^{TM} substitute was discarded for this project due to the disadvantages

related to the decreased cell yield and impossibility of performing epithelial cell enrichment.

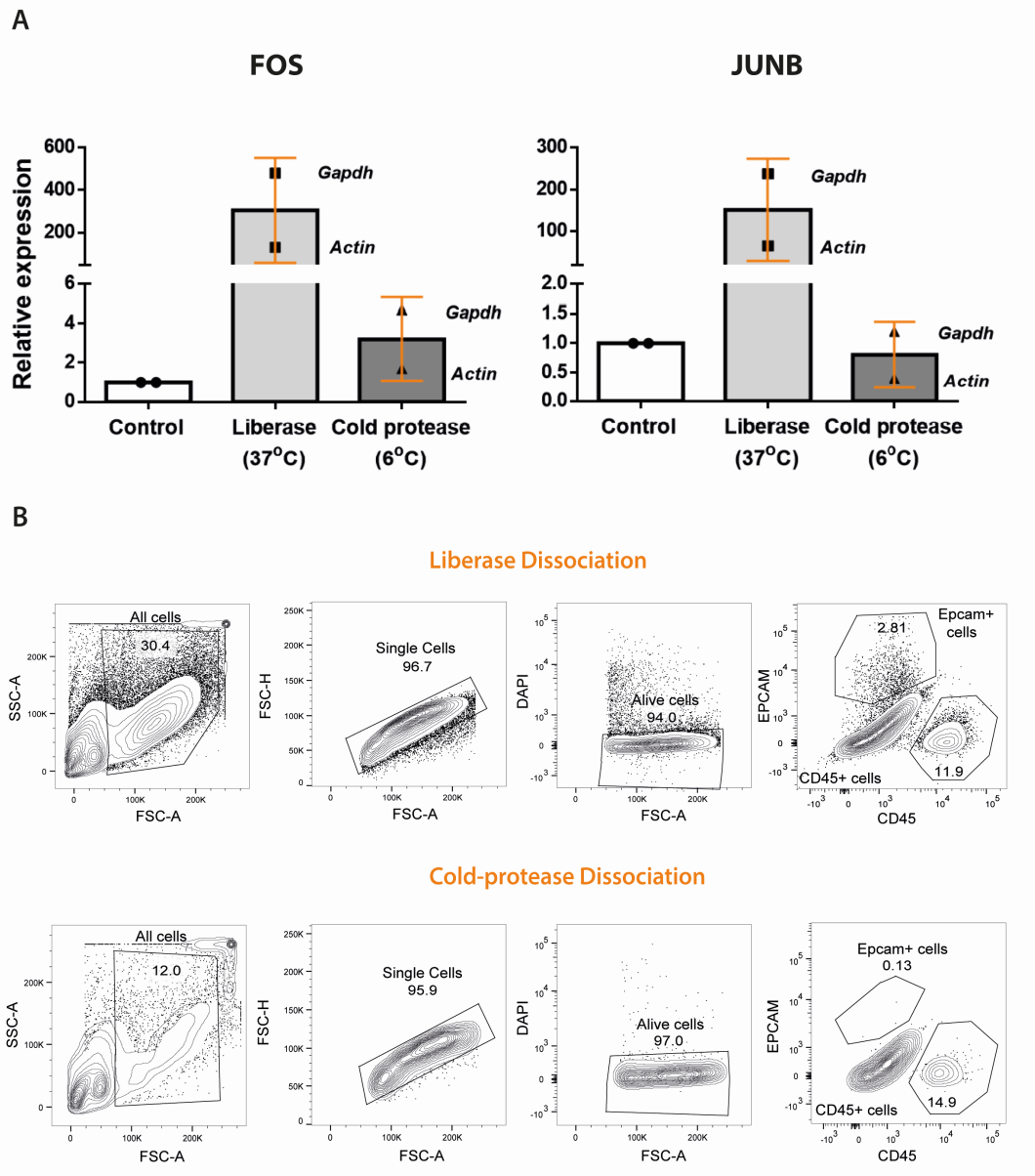


Figure 14. Evaluation of the alternative use of a cold-active protease for enzymatic dissociation of the oviducts
A) Investigation of stress response activation (*Fos* and *Jun* gene expression) in oviduct cells by qPCR of unprocessed oviducts (control) and upon dissociation at 37°C with Liberase™ or at 6°C with a psychrophilic protease (both for 30 minutes). Bars indicate the gene expression normalised to control samples and for the expression of the two housekeeping genes *Actin* and *Gapdh* (squares/triangles indicate the gene expression normalised for each housekeeping gene as per label). The standard deviations between the two housekeeping gene normalisations are shown in orange. B) FACS plots of oviducts dissociated into single cells in the same conditions as A).

1.5 Conclusion

The aim of this chapter was to create a robust and consistent pipeline for single-cell dissociation of the oviducts to use as input for scRNA-seq library preparation. Single-cell sequencing experiments are refined techniques that require extensive optimisation for pre-library preparation stages in order to guarantee high cellular viability, adequate RNA preservation and a well-separated cellular input representative of the original tissue composition. Organ identification, isolation and appropriate dissociation are the first steps in any single-cell analysis protocol. To achieve optimal detachment of individual oviduct cells, a fine balance between robust dissociation, high cell yield, the protection of sensitive cells and the preservation of naïve transcriptomes should be achieved. In the case of the oviducts, the best compromise was accomplished with the use of DNAse I and the commercial enzymatic cocktail Liberase^{TM}, which consists of a blend of precisely purified Collagenases I and II and a non-clostridial neutral protease. Using this formulation, the oviduct lineage landscape was characterised using FACS and main populations and subpopulations of EpCAM^{+}, CD45^{+} and EpCAM^{-} CD45^{-} were identified and sorted. FACS was selected as a suitable technique for this purpose as it allows for the control of key variables such as debris, low-quality cell exclusion and the lineage-specific number of individual cells.

Unexpectedly, epithelial cells, which are an integral part of the oviduct and one of its major components, appeared at a lower than anticipated proportions during the FACS analysis. This underrepresentation can be due to a combination of factors such as epithelial-specific sensitivity to the dissociation process leading to high cell death and the presence of strong intercellular junctions, which limits effective single cell detachment. Epithelial cell loss and depletion following procedures such as enzymatic digestion and cryopreservation have also been documented, not only in human fallopian tubes (Dinh et al. 2021), but also in other systems such as kidney (Denisenko et al. 2020) and lung (Lambrechts et al. 2018). To improve the epithelial recovery, a multi-stage dissociation and a FACS-enrichment approach was adopted. This strategy increased the number of viable epithelial cells sorted, while decreasing the overall number of animals required for the study.

A downside of this implemented dissociation protocol is that the enzymatic activity can only take place at 37°C, and these conditions triggered a conserved stress response in oviductal cells. This gene expression signature was detected at the RNA level by single-cell analysis and it was further confirmed by qPCR. Interestingly, this stress signature is part of a natural mechanism mediating cell adaptation to cell insults and tissue damage (Machado et al. 2021). However, in the context of single-cell studies, it can lead to confounded data analysis by masking the true transcriptional state of cells. Contrary to what other authors such as van den Brink et al. 2017 have described, the stress reaction in the oviducts is cell-type independent, as the majority of cells exhibit it. An alternative protocol involving a psychrophilic protease, functional at 6°C, was evaluated, but this was found to cause a substantial reduction in cell yield and the complete loss of epithelial cells. Therefore, it was concluded that this enzyme is not well-suited for experimental purposes involving epithelial cell enrichment in the oviducts.

Overall, this chapter has aimed at the creation and optimisation of a suit-

able dissociation and enrichment protocol for single-cell RNA sequencing experiments with oviduct cells. Furthermore, it also identified stress-induced artefacts caused by tissue processing, which are considered and addressed in chapter 2.

2 Chapter 2. Single-cell transcriptomic atlas of the mouse oviducts

2.1 Introduction

After having accomplished the optimisation of the single-cell purification step, a single-cell RNA survey was conducted for the mouse oviduct. Single-cell transcriptomics techniques are state-of-the-art tools, which uniquely annexe the fields of cellular, computational and systems biology. They offer a unique approach that can bring comprehensive insights into the cellular composition of the oviduct and the contribution of these cell types to the underlying biological processes of the organ. To date, this technique has been key for exceptional discoveries, which elevated its use among the scientific community (Wagner et al. 2016, Aldridge et al. 2020, Hie et al. 2020). However, this revolution is not exempted of challenges that any new consolidating technology faces (Wagner et al. 2016, Adil et al. 2021). Examples of these are represented by several computational-related variables such as the large scale, the noise and the trend associated with the data. Experimental challenges have also been particularly reported regarding technical robustness and lack of automatization, as the protocols are lengthy, repetitive and mainly rely on manual processes. The scRNA-seq approach not only provides unique insights into the cellular heterogeneity of biological systems by the generation of comprehensive cell atlases, but it is also notable for deciphering their underlying molecular circuitries. Major examples of significant and diverse molecular and cellular knowledge unravelled by scRNA-seq are the characterisation of stem cells and their differentiation processes (Nabhan et al. 2018, Chestnut et al. 2020, Ali et al. 2020), the establishment of cell lineage hierarchies (Mitoyan et al. 2021, Treutlein et al. 2014) and the discovery of novel mechanisms during fertility (Saatcioglu et al. 2019), maternal-foetal interactions in pregnancy (Vento-Tormo et al. 2018), immune response (Shalek et al. 2014) and aging (Ximerakis et al. 2019).

Hitherto, many scRNA-seq projects have been conducted in virtually all known organs and systems at different life stages such as the brain and nervous system (Tasic et al. 2016), heart (Litviňuková et al. 2020), prostate (Crowley et al. 2020), lung (Treutlein et al. 2014, Nabhan et al. 2018), vagina (Ali et al. 2020), uterus (Saatcioglu et al. 2019), etc. Additionally, a broad range of species such as the human (Litviňuková et al. 2020), mouse (Jindal et al. 2018, Tasic et al. 2016, Treutlein et al. 2014), chicken (Li et al. 2020) and zebrafish (Chestnut et al. 2020), among others, have been used for conducting single-cell RNA screening. Furthermore, scRNA-seq has enabled the discovery of previously unknown rare cell identities, which could be particularly advantageous for challenging research such as that limited by the cell sample size, for example, studies scoping the embryonic development, metastatic residual disease (MRD) and stem cell biology (Jindal et al. 2018, Mitoyan et al. 2021). The extent of this technique's power has been further confirmed by multiple validations of the generated data, finding a solid correlation

between the single-cell transcriptomic signatures and real cell populations exhibiting specific properties and functions. The validation step is of paramount importance in bioinformatics-led projects and common strategies used for this are broad, entailing 2D/3D cell culture systems, genetically modified and unmodified animal models and staining techniques such as immunofluorescence (IF) and IHC.

In 2020, the estimated number of published single-cell wide-genome transcriptomics projects was approximately 1000, which along with the rapidly increasing volume of generated data, stimulated the creation of several databases based on collaborative efforts such as the Human Cell Atlas, JingleBells, PanglaoDB, and Tabula Muris (Svensson et al. 2020). Each atlas aims to focus on a specialised domain, counts on different unique features and aids the purpose of establishing the field standards, especially, on data collection, format homogenisation, raw data sharing and accessibility to non-computational biologists. The Human Cell Atlas is a high-scale international repository, which aims to identify all the cell types existing in the human body under physiological conditions (Regev et al. 2017). Additionally, this effort highlights commitment to open source and code sharing with the scientific community, which are essential variables that were not standardised before. The mouse counterpart is represented by Tabula Muris, which comprehensively curates mouse-derived scRNA-seq datasets (Tabula Muris Consortium, 2018). JingleBells is an immune-focused compendium that puts specific emphasis in building a bridge for classic immunologists without a bioinformatics background (Ner-Gaon et al. 2017). Moreover, PanglaoDB is a repository containing processed and analysed data from human and mouse studies (Franzén et al. 2019). Among its useful features are the display of the most commonly differentially-expressed marker signatures in a broad range of cell types and, also, the search of the cell types in which a query gene is most frequently expressed, which proves to be an invaluable reference for cell type identification in novel scRNA-seq datasets. Expanding on model organisms, there is the Single Cell Expression Atlas by EMBL-EBI (https://www.ebi.ac.uk/gxa/sc/home), featuring datasets from multiple additional species such as *Danio rerio* (zebrafish), *Gallus gallus* (chicken), *Drosophila melanogaster* (common fruit fly) and *Schistosoma mansoni* (water parasite). Other useful databases are the one by the Broad Institute, Single-Cell Portal (https://portals.broadinstitute.org/single_cell) and scRNASeqDB (Cao et al. 2017).

Despite the massive effort across disciplines and institutions to undertake cell mapping, classification and characterisation of the complete array of organs comprising a body, the human fallopian tubes and the oviducts continue to be neglected, with little progress in understanding their transcriptome at the single-cell level. Concretely, the understanding of the oviduct cellular heterogeneity under healthy conditions is highly limited and it accounts for one of the few organs that have not been screened and added to Tabula Muris. On the contrary, two publications have addressed these questions in the human fallopian tubes, yet their results scope has been limited to the epithelial component. In the first one by Hu et al. 2020, four secretory, one ciliated and one intermediate secretory-ciliated subtypes were discovered and described for the first time, whereas in Dinh et al. 2021 three secretory, four ciliated and three epithelial undetermined clusters were reported. Although in mouse it is known that secretory and ciliated epithelial subtypes exist and that the former acts as a progenitor for the latter (Ghosh et al. 2017), advanced

knowledge is lacking. This situation has left many outstanding questions in the field, which the current chapter tried to answer by using scRNA-seq. The first aim of this chapter was to decipher the cellular identities comprising the oviduct repertoire and how they establish a cross-talk and cooperation in order to contribute to the maintenance of the organ function and health under physiological conditions. The second aim focused on unravelling the functionality of those identities and the molecular programs underlying their phenotype. To this end, a tailored experimental design was planned and executed. When designing genome-wide transcriptomics projects, it is important to attend to two major endeavours, as previously mentioned. The first is data generation and the second is data analysis. Therefore, the first two challenges faced in this chapter were the selection of the scRNA-seq experimental technique and the design and consolidation of a robust and unbiased data analysis pipeline.

In any scRNA-seq technique, the appropriately dissociated and isolated single cells (chapter 1) undertake a round of lysis and RNA extraction. This RNA is then converted into cDNA, which is subsequently amplified and used for the generation of the sequencing libraries. Commonly, two types of sequencing platforms can be distinguished based on the extent of their read coverage, which can be full-length or partial coverage. The most common examples for each of them are the full-length Smart-Seq2 (Picelli et al. 2014) and the 3' 10X Genomics. Whereas full-length strategies are the most sensitive, capturing the highest number of genes per cell, they are more costly and time-consuming due to the extended manual labour and the low throughput (Ziegenhain et al. 2017). Therefore, they are best suited for projects focusing on specific small cell populations that need to be deeply interrogated. Unlike full-length techniques, non-full length protocols allow for high throughput, achieved by the early addition of cell-specific barcodes to each derived RNA molecule and multiplexing at the cDNA amplification step. Furthermore, these protocols include the addition of UMIs that tag each individual mRNA molecule, enabling the discrimination of PCR-derived cDNA duplication bias, improving gene expression quantification and data noise control. In this context, the 10X Genomics platform is a 3' droplet-based transcriptomics technique, which allows the parallel mRNA library preparation of tens of thousands of cells from multiple samples or conditions.

Due to the different characteristics of the scRNA-seq platforms, there is no best choice in absolute terms, but only an informed one, which depends on the requirements of the specific project. In this case, the aim was to produce a comprehensive characterisation of the mouse oviducts for the first time and, therefore, a platform capable to put up with big numbers was a requirement, reason why the 10X Genomics protocol was chosen. The selection of the 10X Genomics single-cell technology was guided by its scalability but, also, by its reported robustness, price, decreased experimental time and user-friendly pre-analysis bioinformatic tool, which differentiate them from other methods. Moreover, this high-throughput technique has recently been established as the optimal strategy not only for single-cell, but also for single-nucleus RNA sequencing experiments, based on the assessment of several quality and performance metrics (Ding et al. 2020). Although this method presents less resolution than full-length ones, this pitfall can be compensated by increasing the number of screened cells, as it has been shown that the number of

detected populations in scRNA-seq experiments is proportional to the number of cells surveyed (Svensson et al. 2020). The 10X Chromium method is droplet-based, which means that the cells of interest are loaded, along with oil and barcoded beads, into a chip (Figure 15 obtained from the 10X Genomics website, pg 72). In an ideal scenario, one bead and one cell are enclosed in a newly-formed oil droplet inside the chip where the cell will be lysed, freeing the mRNA molecules. Next, an RT-PCR for RNA conversion into cDNA is carried out and all the molecules produced within that droplet are barcoded with the same cell-specific index and a different UMI.

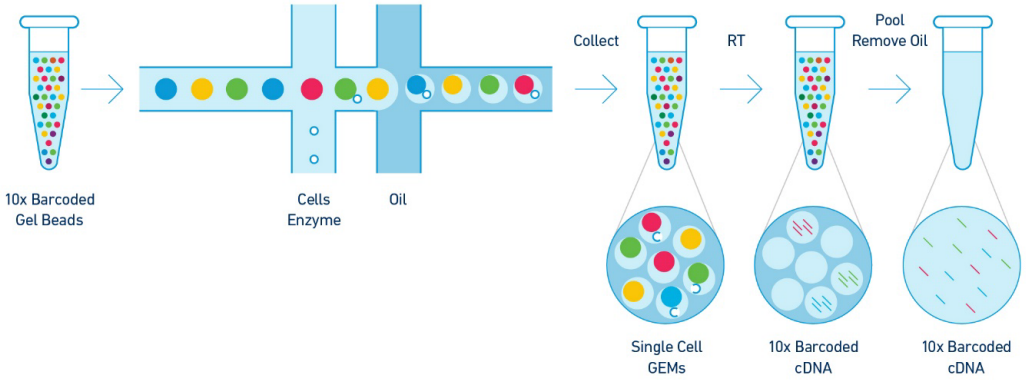


Figure 15. scRNA-seq 10X Chromium protocol

Overall, the purpose of this chapter was to establish an extensive large-scale taxonomy of the cell populations coexisting in the oviducts and, therefore, orchestrating the biological processes underlying the maintenance and the function of the organ. For this, scRNA-seq was used and the obtained results were validated using alternative assays. One of these techniques is another version of RNA sequencing, at the single-nucleus level, whose suitability for the oviducts was assessed and comparisons between the two sequencing methods were drawn.

2.2 Single-cell data analysis setup and workflow

In order to unravel the cellular heterogeneity in the murine oviducts, single-cell RNA sequencing was performed in adult pathogen-free oviducts from C57BL/6 mice (12-week-old) using the 10X Chromium technology of the Single Cell 3' Gene Expression kit for the versions v2 and v3, as specified by their protocol. Subsequently, the raw read-outs were subjected to a sound computational analytical pipeline tailored to minimise bias (Figure 16, pg 74).

In this experiment, seven libraries were created with the 10X Genomics v2' chemistry and five (71.43%) passed the quality controls to be carried forward. Additionally, nine libraries were created with the v3' version and four of them (44.44%) were classified as high-quality. During the first stage of the analysis these libraries were sequenced and the raw data in fastq format served as input for the cellranger count algorithm, which outputs a specific count matrix per cDNA library sample. Cellranger count is a proprietary platform offered by 10X Genomics to perform read

alignment against a reference genome, read filtering and metrics calculations such as cell barcode and UMI counts. The libraries' data were analysed with cellranger count applying the two possible alignment modes: (1) exon alignment and (2) exon plus intron alignment. The second method performed slightly better, improving the number of detected genes and UMIs, therefore, it was kept as the utilised mode. The technical characteristics of the libraries selected for further analysis after application of cellranger count with exon and intron alignment are summarised in Table 4, pg 75.

After running cellranger, every sample-specific count matrix was individually filtered for low-quality cells based on parameters such as mitochondrial genes proportion, library complexity and gene-UMI correlation. Only cells that exhibited a low mitochondrial gene proportion (<0.05), high library complexity (>200 genes) and good gene-UMI correlation (>0.9), were kept in each sample library. Regarding the cut-off values, a mitochondrial genes proportion between 0.1 and 0.05 is accepted as good quality, however, in this project the most limiting value was chosen in order to assure high data quality. Regarding the number of detected genes per cell, different cell types expressed a specific number of genes, therefore, a strict cut-off was kept, while avoiding cell type depletion from the oviduct sample. Similarly, genes from the count matrix were also filtered out if they were not expressed in, at least, three cells. Afterwards, heterotypic doublets from the individual sample datasets were removed from the data using the R package DoubletFinder (McGinnis et al. 2019a), which notably improved the unsupervised clustering results.

These data cleaning steps were paramount to improve the data quality and undertake a meaningful primary and secondary analysis. As shown in Figure 17, the strict quality control step substantially improved the mean number of detected genes (Figure 17 A, pg 75) and UMIs (Figure 17 B, pg 75) per cell. Once every sample-specific count matrix had been appropriately filtered by performance metrics and multiplets' presence, all of them were integrated into a unique dataset using the R package Seurat (Stuart et al. 2019). Subsequently, secondary data analysis involving unsupervised clustering, supervised marker identification and differential expression calculations was carried out using the same package. Finally, further biological investigation was undertaken by performing experimental validation assays and additional bioinformatic analysis.

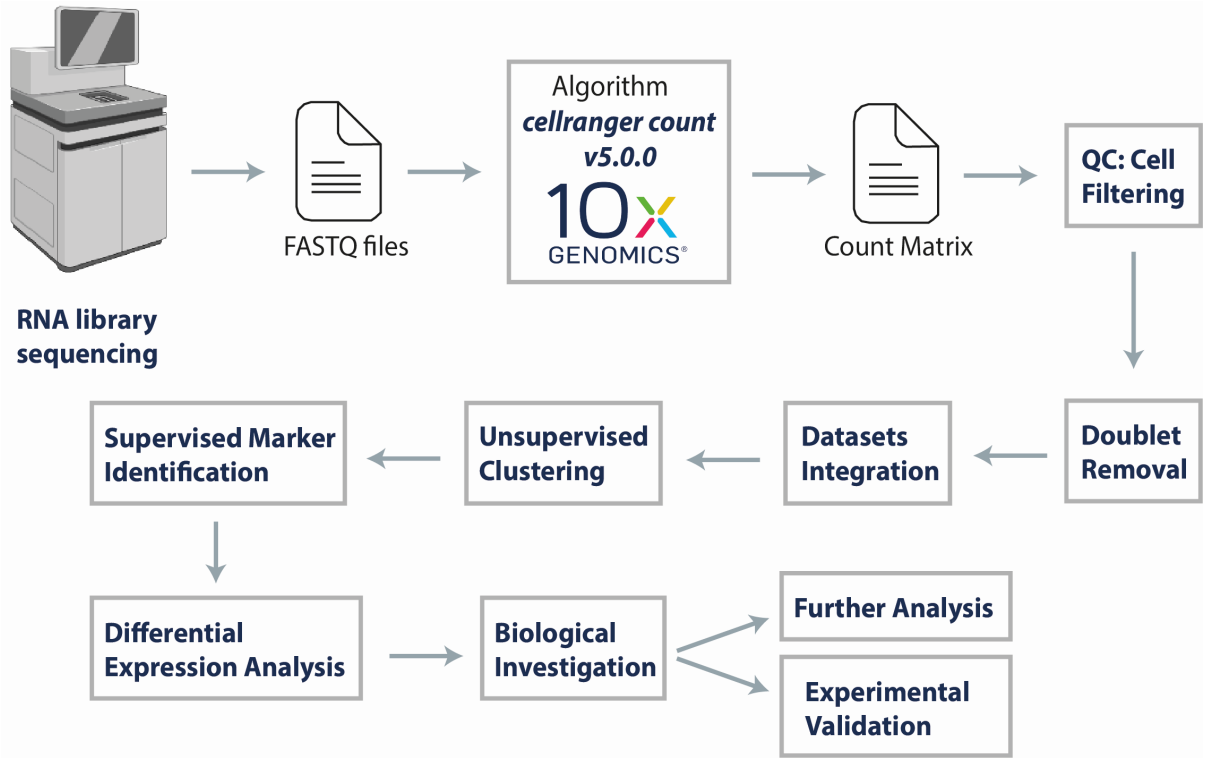


Figure 16. Bioinformatic analytical pipeline for the analysis of scRNA-seq oviduct data

After the single-cell RNA libraries are sequenced at an adequate sequencing depth, the output read-outs are enclosed in a fastq file. The fastq files for each sample library are input to the cellranger count algorithm developed by 10X Genomics and a sample-specific count matrix is output by the platform. This count matrix contains the individual gene counts per cell in the sample and it is ready to be subject to bioinformatic analysis. The first step is to clean the individual matrices from poor-quality cells based on standard metrics and multiplet removal. Then, all the matrices are integrated into a unique object corrected for sample-specific batch effects. Afterwards, unsupervised clustering is applied on the data and the different clusters are identified by a combination of supervised marker identification and a differential expression analysis, which obtains the subtype specific genes expressed/upregulated by a cell population. These gene expression signatures are the foundation for further biological analysis and the design of validation experiments. QC= Quality control.

Sample	10X Genomics chemistry (*)	Estimated Cell Number(*)	Median Genes/Cell (*)	Median UMIs/Cell (*)	Sequencing Saturation (%) (*)	Number of High-quality (HQ) Cells (%)	Median Genes/ HQ Cell	Median UMIs/ HQ Cell
B1	v2	2.833	2.386	5.887	82.9	1333	3181	8506
WGA1	v2	4.635	1.715	3.901	87.9	3000	1972.5	4522
WGA2	v2	1.296	2.156	5.780	81.8	771	2608	7141
E1	v2	4.954	931	1.755	77.1	2653	1231	2666
E2	v2	10.297	1.007	1.836	84.1	6731	1023	1941
S1	v3	8.642	1.492	2.921	36.4	2628	2715	7810
S2	v3	7.520	1.482	2.848	34.1	2376	2679	7548.5
S6	v3	9.559	1.522	2.942	40.0	1847	1837	3894
S8	v3	5.708	2.152	5.394	59.9	2438	2835.5	8480.5

Table 4. Technical metrics per single-cell RNA sample from adult mouse oviducts
Metrics marked with an asterisk (*) were part of the cellranger count output when aligning introns and exons to the mouse reference genome. Metrics without asterisks were calculated after the quality control steps of cell filtering by metrics and multiplet status. The abbreviation HQ stands for High Quality.



Figure 17. Impact of the filtering strategy on data quality
A) Median of the number of genes detected per single cell in each RNA sample is depicted prior and after filtering based on metrics and doublet state. B) Median of the number of transcripts (UMIs) detected per single cell in each RNA sample is depicted prior and after filtering based on metrics and doublet state.

2.3 Single-cell RNA-Seq of the mouse oviduct faithfully captures the main cellular lineages previously reported in this organ

The total number of sequenced cells following the single-cell transcriptomics of adult oviduct tissue specimens was 55.444. Of these, 42.89% (23777 cells) passed the quality control data filtering and were retained for downstream analysis. The minimum representative sample size needed to capture the complete cellular heterogeneity in the oviducts was determined following the negative binomial method developed by the Satija Lab https://satijalab.org/howmanycells. For this model, some input parameters have to be determined by the user: assumed number of cell types, minimum fraction of rarest cell type and minimum desired number of cells per type group. The number of cell types was approximated based on histology experiments and the literature. Five epithelial subtypes (three secretory and two ciliated) in murine oviducts have been reported to date in the preprint Ford et al. 2020 (deposited in Biorxiv in August 2020, currently under review at Cell Reports). Furthermore, although information on cell populations comprising the lamina propria, the muscularis and serosa layers and the immune repertoire in the oviducts was very scarce, at least one mesenchymal, one smooth muscle and one mesothelial cell subtype should be expected in this organ. Regarding the immune populations in the oviducts, one study has reported seven immune subtypes in human fallopian tubes to date, therefore, this number was extrapolated for calculations in mouse (Ardighieri et al. 2014). Following these hypotheses, the total number of expected cell types in the murine oviduct came to fifteen. The frequency of the rarest cell type was based on epithelial cells proportions, as they are the oviductal parenchyma, the cells that carry out the oviductal function, and, therefore, cells of main interest. Epithelial cells were also found to represent a low frequency of the total oviduct population after undertaking FACS (3-15%), contrarily to their prevalence in the naïve organ (around 50%). Selecting the most restrictive value (3%), considering that at least five epithelial subtypes have been documented in mouse and assuming that every subtype contributes equally to the epithelial population, 0.6% is chosen as an input for the frequency of the rarest cell type parameter. Finally, the minimum number of cells per type was adjusted to 100 cells. In summary, assuming 15 cell types in the oviducts with the rarest cell population found at a frequency of 0.006, the probability of seeing 100 cells per type is 99% when a minimum of 22517 cells are profiled. In this project, I met that requirement by sequencing 23777 high-quality cells.

After sequencing and quality control analytics, integration and unsupervised clustering followed using the Seurat package. This revealed the presence of oviduct cell clusters of all the lineages expected from the histological oviductal layers following previous reports (Figure 18A, pg 78). Moreover, this analysis denoted a significant contribution from all the individual libraries to this compendium. Interestingly, minority lineages that were not initially considered, such as pericytes and endothelial cells, were also detected. Clustering also delineated poor-quality cells that had passed the previous quality control filters and they were eliminated from subsequent analysis. Besides, some cells did not present any particular lineage markers and were thereby labelled as NA (not applicable).

The oviduct landscape is dominated by three major populations, immune

(41.09%), stromal mesenchymal (30.58%) and epithelial (19.97%). Despite the enrichment step when FACS sorting the samples, the final dataset composition diverged from the composition of the single-cell suspension loaded onto the 10X chip (discussed in this chapter's conclusion section) (Figure 18B, pg 78). Furthermore, in line with previous observations (Tasic et al. 2016), the diverse cell lineages display different library complexities (Figure 18C, pg 78). Endothelial cells were the ones accounting for the lowest number of expressed genes, which can suggest (1) a more homogeneous or "simpler" function and/or (2) a more efficient adaptation by requiring the expression of less genes to perform their role. The library complexity for the group termed as NA was significantly higher than the other groups, which might correspond to cell doublets that were not recognised by DoubletFinder.

These populations were annotated by the expression of canonical molecular marker signatures such as *Epcam* and the keratins *Krt19* and *Krt18* for epithelial cells (Figure 19, pg 79). For the mesenchymal lineage, matrix-associated proteins such as osteonectin (*Sparc*), the proteoglycan decorin (*Dcn*) and the alpha-2 type I collagen chain (*Col1a2*) were used. These were markers shared with pericytes. The expression of the markers *Cd52* (glycoprotein) and *Ptprc* coding for Cd45 were used to identify the immune lineage. Lastly, the uroplakins *Upk1b* and *Upk3b* were used to identify mesothelial cells and *Egfl7* and *Pecam1* for endothelial cells.

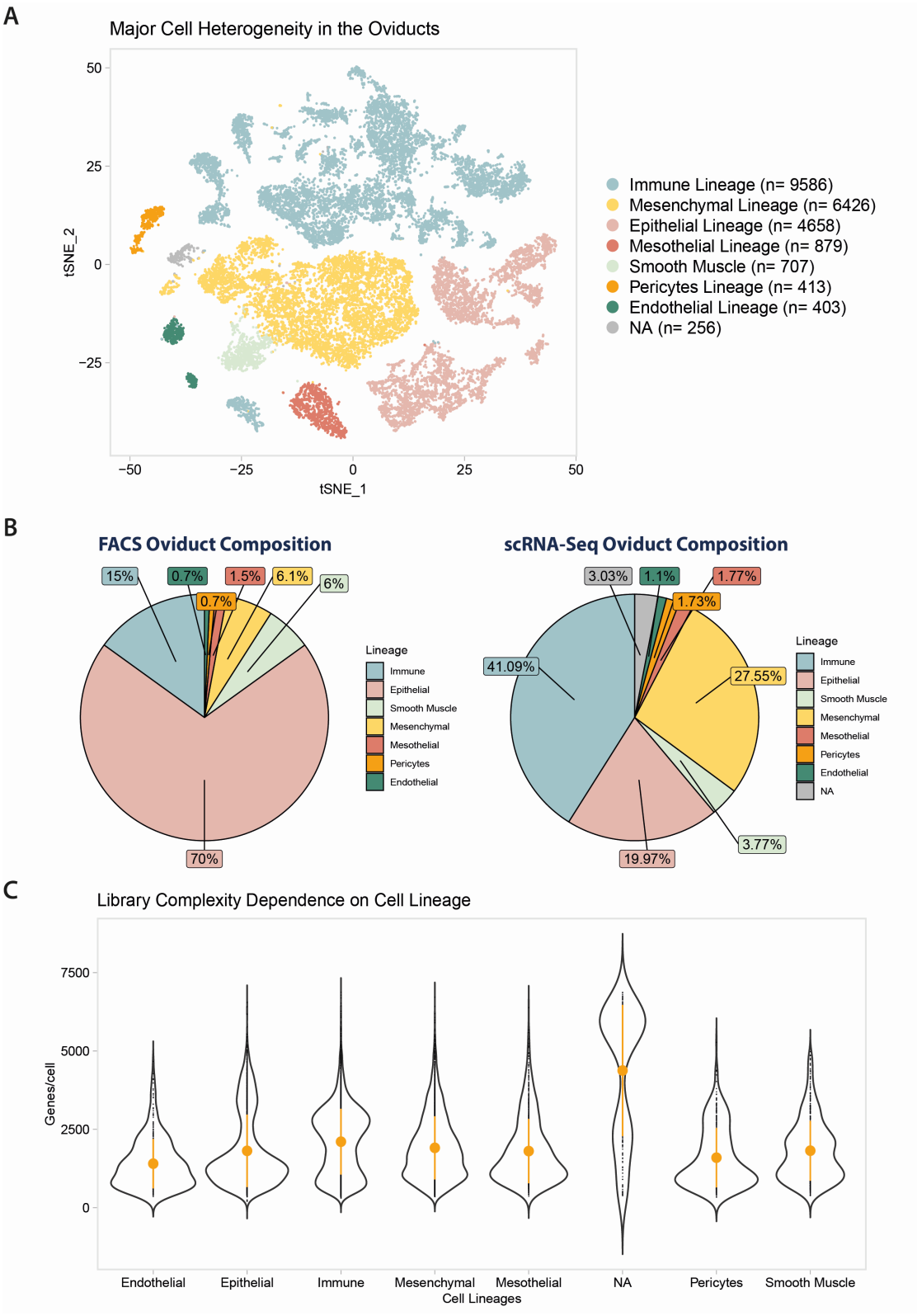


Figure 18. Cell-lineage heterogeneity in the mouse oviduct
A) t-distributed stochastic neighbor embedding (tSNE) plot of the oviduct lineages with the specific number of cells comprising each group. NA stands for Not Applicable. B) Pie chart reflecting the sample cell composition after FACS sorting and scRNA-seq library sequencing and integration. C) Violin plot showing the distribution of the number of detected genes per cell in each oviduct lineage. The mean and standard deviations are presented as orange dots and lines respectively.

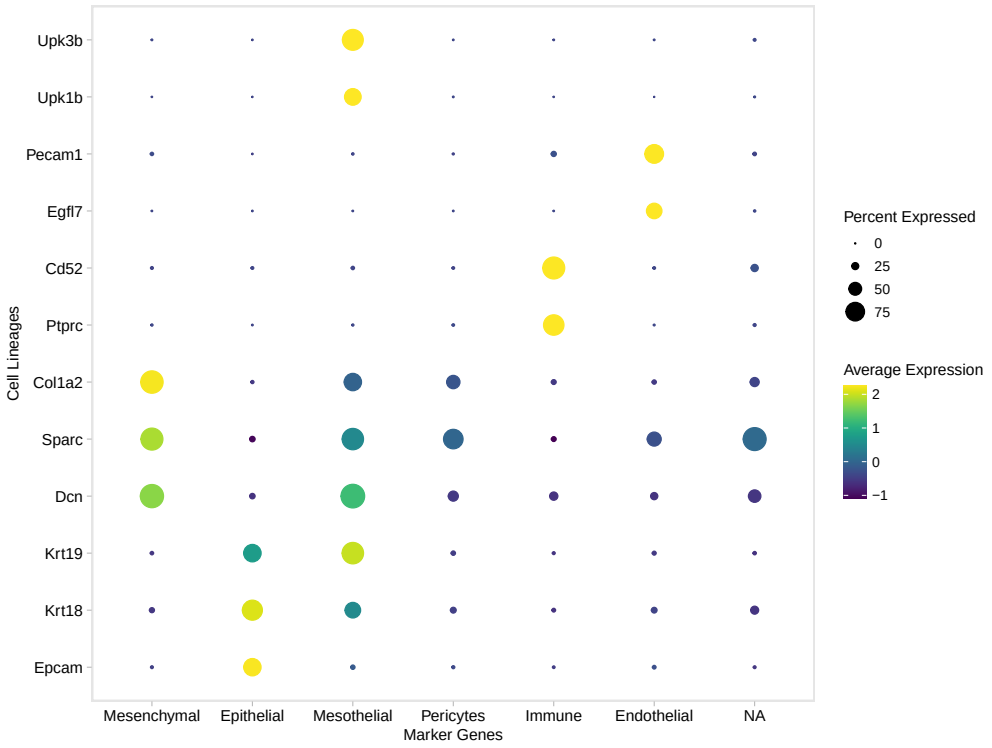


Figure 19. Cell-lineage genetic markers of the oviduct
Dot plot of the molecular markers used to classify oviduct lineages. The size of the dot represents the percentage of cells within the group that expressed the marker, whereas the dot colour represents the marker centered and scaled expression level. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. Centering these values is done by subtracting their average normalised expression and scaling divides the centered gene expression by their standard deviations.

2.4 scRNA-seq analysis of the oviduct epithelium repertoire reveals diverse secretory and ciliated subtypes marked by distinctive biological processes

The oviduct epithelium dataset presented ten discrete clusters in total (Figure 20A, pg 81). These groups could be classified within two general secretory and ciliated subtypes due to their distance in the uniform manifold approximation and projection (UMAP) plot and the expression of the molecular markers *Ovgp1* (secretory) and *Foxj1* (ciliated) (Figure 20B, pg 81). This subdivision was further confirmed by the expression of additional markers such as *Pax8* (secretory) and *Tubb4b* (ciliated). Although *Pax8* expression was predominantly found in secretory cells, it mainly constituted a dropout in the dataset. A dropout is a concept that refers to a gene that is expressed in a naïve cell state, but its transcript is not detected in the scRNA-seq data due to technical limitations. Secretory and ciliated cells were detected in similar ratios as both represented almost half of the dataset (2358 ciliated cells (50.62%) versus 2300 secretory cells (49.38%)). Intriguingly, an intermediate secretory-ciliated phenotype in cells labelled as ciliated, Cil_Mapk10 and Cil_S100g, could be observed as they co-express *Ovgp1* and *Foxj1* (Figure 20B). Nevertheless, these clusters were still considered as ciliated in this study, since the expression of

Foxj1 and *Tubb4b* is indicative of cilia formation. The possibility that these intermediate clusters could be cell doublets was considered unlikely due to a series of evidences: (1) Their library complexities are similar to both ciliated and secretory clusters; (2) a differentiation hierarchy where secretory cells give rise to ciliated cells through an intermediate state has been previously shown using lineage tracing and genetically-modified mouse lines (Ghosh et al. 2017) as well as organoid models (Kessler et al. 2015); (3) no other population in the dataset presents co-expression of two subpopulation-specific markers and, if this was a doublet effect, it would have been commonly observed in non-epithelial clusters of the dataset; (4) the populations express their own characteristic genes; (5) an intermediate ciliated-secretory population has been detected and validated in fallopian tubes (Hu et al. 2020); (6) intermediate populations between functionally different epithelial subtypes in other organs have been reported as well (Crowley et al. 2020).

In order to further investigate the specific biological processes characterising secretory and ciliated populations, the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al. 2003) was used to functionally analyse their differentially expressed genes (DEGs) (Figure 20C and D, pg 81). Ciliated cells further confirmed their identity by displaying enrichment for biological processes related to the cilia function and genes of the smoothened or Hedgehog (Hh) signaling pathway. This pathway is essential for cell-to-cell communication during embryonic development and cancer biology and has extensively been related to cilia as its transduction depends on the primary cilium (Bangs and Anderson 2017, Goetz et al. 2009). Furthermore, ciliated cells were marked by nitric oxide synthase regulator activity, calcium-related metabolism, GTP binding and nucleoside diphosphate kinase activity. Calcium signalling has previously been found to be essential to the cilia function by controlling the Hh signalling (DeCaen et al. 2013, Delling et al. 2013). On the other hand, the secretory epithelial profile revealed a strong presence of endoplasmic reticulum processes, extracellular secretion, ion transport, redox reactions, epithelial cell differentiation and migration and positive regulation of MAPK kinase activity. Furthermore, secretory cells exhibited enrichment in several metabolic pathways, such as lipid and fatty acid metabolism, with the differential expression of fatty acid elongases, cholesterol transporters, acyl-CoA synthetases and phospholipases, among others. Upregulation of the lipid and fatty acid metabolism has previously been reported as a survival mechanism that high-grade serous ovarian cancer cells use to endure chemotherapy (Artibani et al. 2021), which could suggest that the high lipid metabolic profile could be a homeostatic feature of secretory cells that their cancer versions hijack to be able to survive stressful conditions. This subtype was also marked by glutathione transferase activity, which plays a key role in detoxification and controlling oxidative stress, oxidation-reduction processes and biosynthesis of aminoacids. Regarding the glucose metabolism, secretory cells differentially expressed enzymes such as two alcohol and aldehyde dehydrogenases (*Adh1*, avg_log2FC = 0.62, p-value = 2.75^{-41} and *Aldh2*, avg_log2FC = 1.11 , p-value = 7.68^{-83}), aldehyde reductase (*Akr1a1*, avg_log2FC = 1.20, p-value = 1.29^{-110}), aldolase (*Aldoc*, avg_log2FC = 2.15, p-value = 1.77^{-234}) and enolase (*Eno1*, avg_log2FC = 1.02, p-value = 7.12^{-71}). Interestingly, overexpression of aldehyde dehydrogenases has been reported to characterise tumour-initiating and MRD cells in preclinical models of human breast and ovarian cancer (Tachtsidis et al. 2016, Artibani et al. 2021).

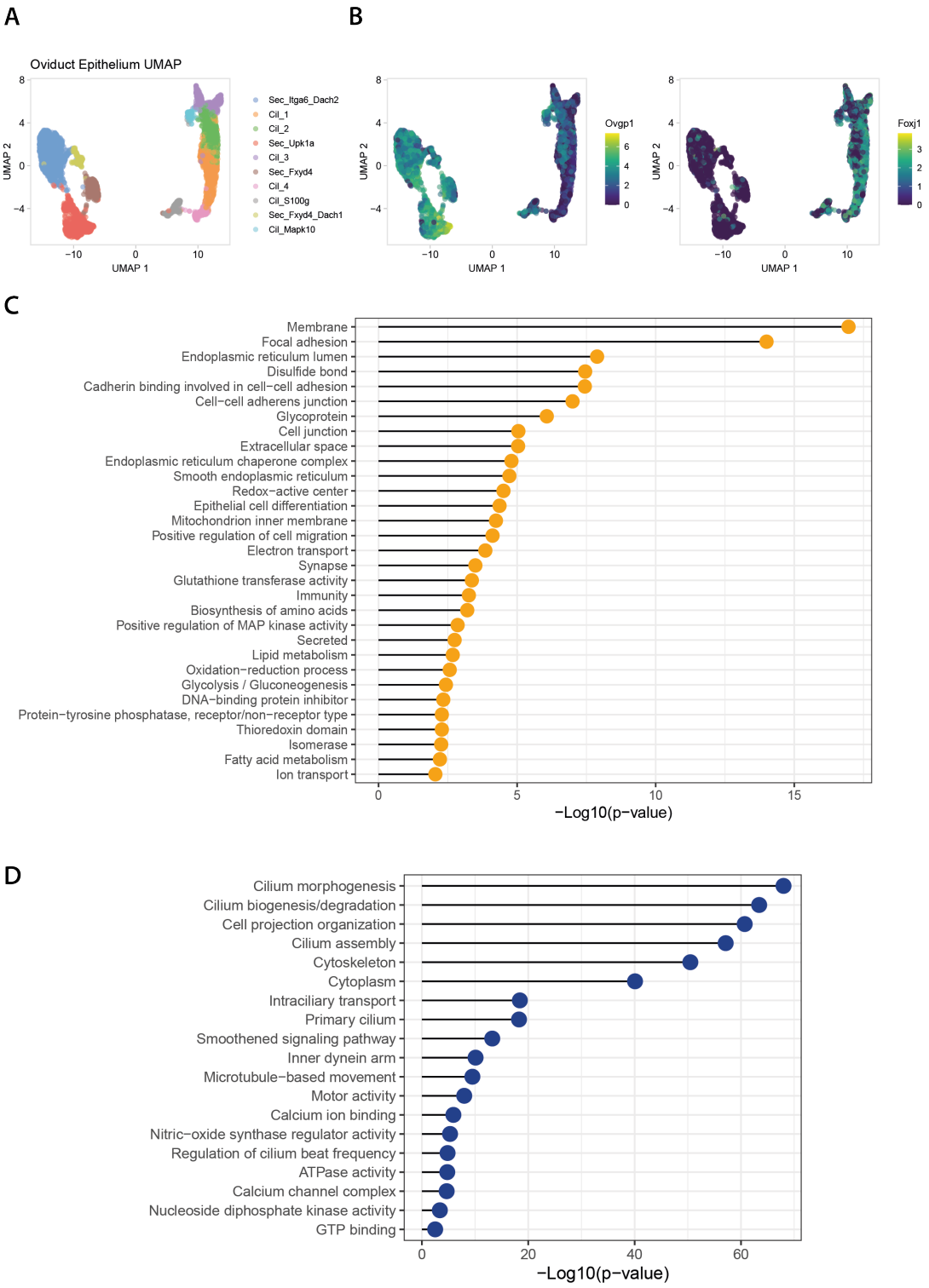


Figure 20. Single-cell characterisation of the oviduct epithelium
A) UMAP embedding of single oviduct epithelial cells (dots) coloured by clusters. B) UMAP plot of single oviduct epithelial cells coloured by the expression of a secretory marker, *Ovgp1*, and a ciliated marker, *Foxj1*. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. C) Dot plot shows the main biological processes enriched in secretory epithelial cells after DE analysis against ciliated epithelial cells. D) Dot plot shows the main biological processes enriched in ciliated epithelial cells after DE analysis against secretory epithelial cells.

Interestingly, secretory cells also showed enrichment in immunity-related terms and expressed multiple genes whose function relate to the immune system, especially the innate arm (Figure 21A, pg 85). Several of those identities were also expressed by ciliated cells, although secretory cells demonstrated a stronger immunity pattern, marked by higher levels of both, gene expression and number of cells expressing a given gene. Within secretory cells, the secretory Itga6_Dach2 cluster exhibited a higher and, occasionally, exclusive expression of some of those immune markers (Figure 21B, pg 85).

Within the antiviral genetic identities, 2'-5'-oligoadenylate synthase-like protein 2 (*Oasl2*), bone marrow stromal cell antigen 2 (*Bst2*, also known as tetherin or Cd317), serine incorporator 5 (*Serinc5*) and interferon-induced transmembrane protein 2 (*Ifitm2*) code for interferon-induced molecules, specifically, an enzyme, in the first case, and membrane proteins in the other cases. The production of interferon cytokines by virus-recognising cells is a common and essential strategy to fight against viral invasion (Platanias 2005) and these four molecules exhibit different action mechanisms to fight viruses upon interferon signalling. Oasl2 is an oligoadenylate synthetase, which activates the RNase L to induce viral RNA breakage and inhibit viral replication (Zhu et al. 2015); Bst2 is a membrane glycoprotein that functions against viruses that bud into a host cell through a tethering mechanism by impairing the virus release (Tiwari et al. 2019); Serinc5 and Ifitm2 are membrane proteins that block the release of host-endocytosed viruses into the cytoplasm. Ifitm2 has been demonstrated to restrict a broad range of viruses such as Hepatitis C (Narayana et al. 2015) and SARS-CoV-2 (Shi et al. 2021). Moreover, Serinc5 was originally discovered as a strong inhibitor of the lentivirus HIV-1 and, during successful infections and development of the acquired immunodeficiency syndrome (AIDS), Serinc5 is successfully counteracted by HIV-1 (Rosa et al. 2015).

Interferons bind to cell receptors in an autocrine or paracrine fashion, triggering the production of antiviral interferon (IFN)-stimulated genes (ISGs) such as the aforementioned. In the oviduct epithelium, a chain of the type I IFN receptor 2 gene for interferon alpha and beta (*Ifnar2*) was found to be expressed by oviduct epithelial cells. This receptor is associated with the tyrosine kinase Jak1, also predominantly expressed in the secretory clusters Itga6_Dach2 and Fxyd4_Dach1 and the intermediate cluster Cil_Mapk10. Jak1 functions through the canonical Janus kinase/signal transducers and activators of transcription (JAK/STAT) signalling pathway (Chmiest et al. 2016). When this pathway is switched on, MAP kinases are in charge of activating cytoplasmic Stat proteins and, interestingly, positive regulation of MAP kinases, is among the biological annotations defining secretory cells (p-value = 0,0014). After activation, Stat1 and Stat2 form a complex with the IFN-regulatory factor 9 (*Irf9*), also found in the dataset, known as ISG factor 3, which is transported to the nucleus and controls the transcription of ISGs (Michalska et al. 2018).

Furthermore, the transcription factor *Irf1* was found to be transcribed primarily from Itga6_Dach2 secretory cells (Figure 21B). This regulator controls the transcription of IFN and IFN-inducible genes and not only responds to viral, but also, to bacterial infections (Feng et al. 2021). Its basal expression in mouse epithelial cells, such as hepatocytes, has previously been documented and determined to drive a broad-range of first line defence mechanisms against multiple viruses (Ya-

mane et al. 2019). Additionally, *Oas2*, a similar gene to *Oasl2*, and *Bst2* have been found to be among the target genes regulated by *Irf1* in bronchial epithelial cells (Panda et al. 2019). The protein tyrosine phosphatase non-receptor type 22 (*Ptpn22*) is also expressed at a higher level in secretory cells. In mice, an allele of this protein has been reported to increase T cell activation in the context of viral infection, favouring its resolution (Orozco et al. 2020). Furthermore, this phosphatase has been reported as a negative regulator of T-cell receptor (TCR) signaling and as a positive regulator of the phosphorylation level of the Src tyrosine kinase Lyn, also found in the immune transcriptomic signature of epithelial cells (Vermeren et al. 2016). Lyn has been identified as a negative modulator of immune receptors in immune cells (Mkaddem et al. 2017) and recent evidence has pointed at a similar role in epithelial cells, which could be also conserved in the oviduct epithelium (Sinha and Lowell, 2016).

Secretory cells also exhibited a strong expression of antibacterial genes. Among them, there is cluster of differentiation 14 (*Cd14*), a cell membrane receptor classically expressed by macrophages. This receptor is part of the innate immune ammunition called pattern-recognition receptors (PRRs), due to their role in recognising conserved pathogen-associated molecular patterns (PAMPs) and orchestrating a transcriptional response leading to the initiation of inflammation and the immune fight (Zanoni and Granucci, 2013). The Cd14 antigen is specialised in recognising lipopolysaccharides (LPS), found on the outer membrane of Gram-negative bacteria. This receptor has been found to be, either integrated on the cellular membrane, or as a secreted molecule. Previously, Cd14 was reported to be expressed and immunologically active in epithelial types such as mouse and human intestinal epithelial cells, where the secreted version is dominant (Basic et al. 2018, Funda et al. 2001), and human corneal epithelium, where Cd14 was located on the cell surface (Song et al. 2001). Furthermore, this protein has also been extensively characterised as a co-receptor of the Toll Like Receptor 4 (*Tlr4*). This receptor belongs to the TLR family of PRRs, essential in the immune response to viruses and bacteria, and whose expression was primarily found in the secretory cluster Itga6_Dach2. In the oviduct context, *Cd14* and this co-receptor's expressions were found to be upregulated following exposure to LPS in bovine oviducts, which reportedly caused reproductive dysfunctions (Ibrahim et al. 2015, O'Doherty et al. 2016). However, another article suggested that these two genes are not involved in the resolution of *Chlamydia muridarum* infection in the mouse oviducts (Imtiaz et al. 2006). Although this effect could be due to the reported low immune reaction triggered by this pathogen's LPS, further investigations aiming to confirm the roles of these PRRs and the existing Cd14 version in the oviduct epithelium, should be conducted.

Other antibacterial genes expressed by secretory cells are the Peptidoglycan Recognition Protein 1 (*Pglyrp1*), which targets the murein peptidoglycans of Gram-positive bacteria, the protein kinase C epsilon (*Prkce*), that participates in the LPS-induced Tlr4 signalling (Aksoy et al. 2004) and the Surfactant Protein D (*Sftpd*), which recognises pathogenic patterns and promotes bacterial elimination through the formation of aggregates and the interaction with macrophages. This molecule can also regulate the adaptive immune system by interacting with antigen-presenting and T cells (Pastva et al. 2007). Furthermore, secretory cells express

lipocalin 2 (*Lcn2*) and lactoferrin (*Ltf*), the latter commonly found in secretions such as breast milk. These proteins capture iron ions that bacteria would use to grow and, in the case of Ltf, it can also release LPS from the bacterial outer membrane (Velliyagounder et al. 2003, Ellison et al. 1988). *Lcn2* is exclusively expressed by the secretory epithelial Itga6_Dach2 and Upk1a clusters, in a decreasing order.

In their fight against bacteria, secretory cells also express proteins of the complement system, which is an essential bactericidal part of the innate immunity that functions as a molecular cascade. During homeostasis, the complement system is tightly kept in a low activity status to allow for pathogen surveillance and to enable the phagocytosis of apoptotic cells without mounting a complete immune response (Merle et al. 2015a, Merle et al. 2015b). C3 is the central molecule of the complement system, and is secreted and mostly circulates in a non-immunologically reactive form, until contact with a pathogen that triggers its activation by cleavage, producing C3a and C3b. C3a (anaphylatoxin) orchestrates a local inflammatory response and C3b (opsonin) binds to foreign molecules, tagging them for subsequent phagocytosis. In the oviducts, expression of *C3* mainly happens in secretory epithelial cells and suggests that C3 is, as in other systems, secreted into the oviduct lumen, where it mainly patrols in its inactive version until pathogenic contact. The Complement Factor I (*Cfi*) is an enzyme also secreted exclusively by oviduct secretory cells and that regulates the constitutive activation of the complement. Cfi, along with a cofactor, ensures that this basal activation does not turn into an immunological response by favouring the cleaving of C3b into an inactive form (Dreismann et al. 2021).

Furthermore, the secretory cluster Itga6_Dach2 expressed the Interleukin-4 Receptor (*Il4ra*), a type-I cytokine receptor for interleukin 4 and 13, which has been suggested not to have an immune role in intestinal epithelial cells (Hoving et al. 2020), whereas a role in proliferation in a colon cancer line was proved (Koller et al. 2010). Following these evidences, the function played in the secretory epithelium of the oviduct requires further investigation. Finally, the pro-inflammatory cytokine macrophage migration inhibitory factor (*Mif*) was found to be strongly expressed by both secretory and ciliated cells. This factor works as an important player in the innate immune system, controlling, among other functions, the upregulation of *Tlr4* (Calandra and Roger 2003).

Taken altogether, this evidence is suggestive of an intrinsic role of secretory cells and, concretely, the Itga6_Dach2 cluster, as the epithelial guardians and sentinels of the immune system in the oviducts, in charge of detecting pathogenic insults and eliciting an appropriate immune response. However, further mechanistic validation is necessary to verify this theory. Interestingly, another observation related to the immune system, was the fact that the major histocompatibility complex Type I (MHCI) seemed to be expressed at very low levels or absent in some oviduct epithelial cells, especially, in ciliated cells. This observation was then verified through the following presented investigations.

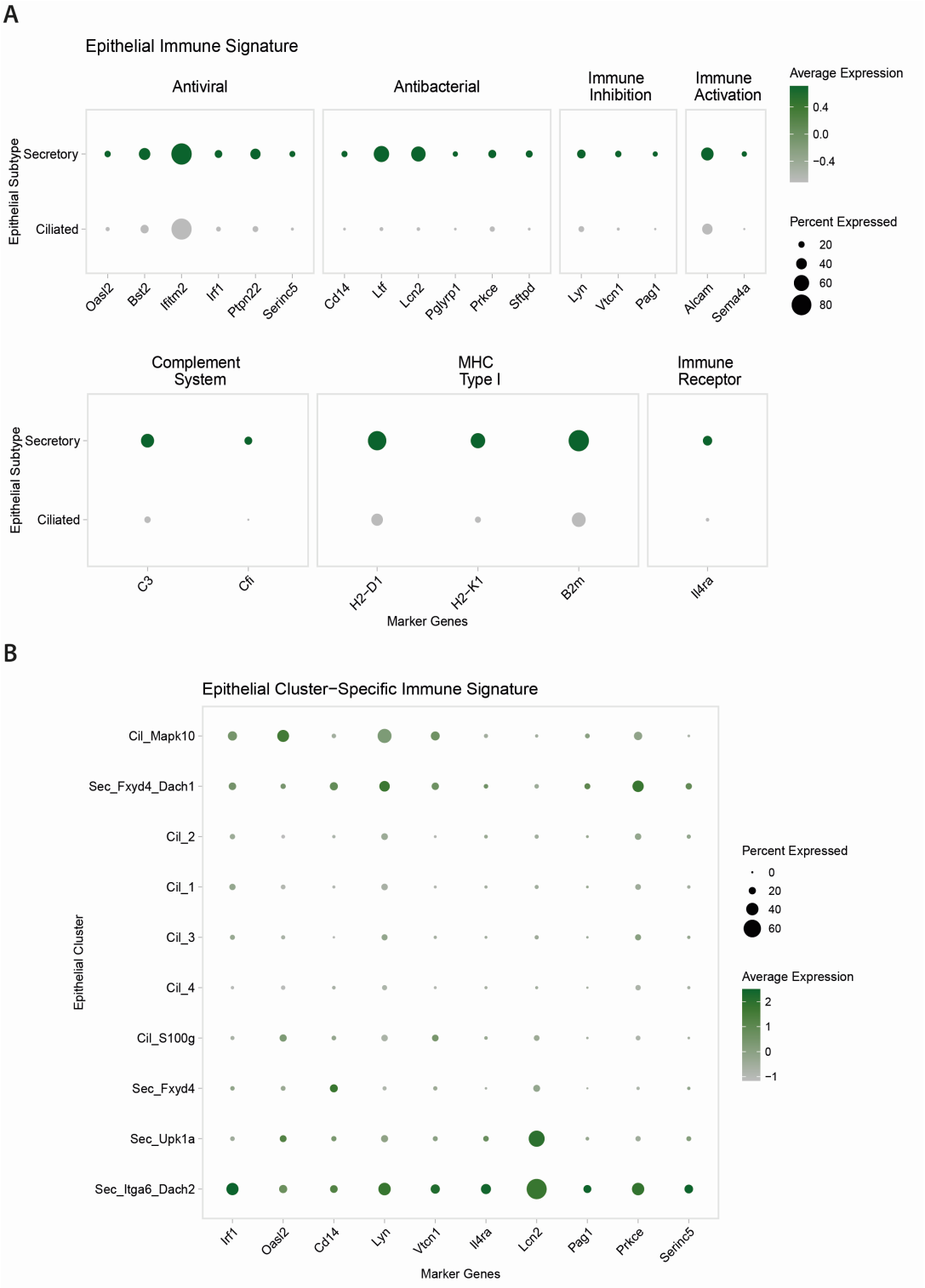


Figure 21. The immune transcriptomic signature of the oviduct epithelium A) Dot plot of the immune-related genes expressed in the ciliated and secretory epithelial types. The size of the dot represents the percentage of cells within the group that expressed the marker, whereas the dot colour represents the marker centered and scaled expression level. B) Dot plot of the immune-related genes expressed in the ciliated and secretory epithelial discrete clusters. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. Centering these values is done by subtracting their average normalised expression and scaling divides the centered gene expression by their standard deviations.

2.5 Low or absent gene and protein expression of the major histocompatibility complex (MHC) Type I is identified in oviduct epithelial cells

The MHC Type I, also known as H-2 in mouse, is an essential protein complex traditionally involved in the adaptive immune response. To that end, it binds endogenous peptides obtained from the hydrolysis of intracellular self or non-self-proteins in the endoplasmic reticulum, an organelle that displays a number of related biological annotations differentially found in secretory cells when compared to ciliated cells (Neefjes et al. 2011) (Figure 20C, pg 81). Specially relevant are the biological annotations "endoplasmic reticulum chaperone complex" and "isomerase activity", as both feature the protein disulfide isomerase gene *Pdia3* (Wright et al. 2004), which helps in the MHC loading, and it is differentially expressed between secretory and ciliated subtypes (avg_log2FC = 1.15, p-value = 4.64^{-76}). After this process, the whole complex travels to the cell membrane, where the MHC class I presents the bound antigen to cytotoxic $Cd8^+$ T cells.

The protein complex is comprised of a polymorphic heavy chain, which can be coded by classical or non-classical molecules, β2-microglobulin (B2m), which is a light chain and the peptide to be presented, fitted into the MHCI binding groove. The classical and non-classical genes encoding the MHCI heavy-chain proteins are located on chromosome 17. Classical MHCI or MHC-Ia genes are found on the D region (*H2-D1*), K region (*H2-K1*) and L region (*H2-L*). Non-classical MHCI or MHC- Ib genes are located on the Q region (*H2-Q1*, *H2-Q2*, *H2-Q4*, *H2-Q5*, *H2-Q6*, *H2-Q7*, *H2-Q10*), M region (*H2-M2*, *H2-M3*) and T region (*H2-T3*, *H2-T5*, *H2-T7*, *H2-T9*, *H2-T10*, *H2-T11*, *H2-13*, *H2-15*, *H2-T22*, *H2-T23*, *H2-T24*). These genes are highly polymorphic, a characteristic that has been proposed to evolve following mutations driven by the pathogenic evasion of the immune mechanisms controlled by this complex (Radwan et al. 2020). Genes encoding the MHCI have largely been reported to be ubiquitously expressed by all healthy nucleated cells except in rare cases such as germinal cells (Garcia-Lora et al. 2003). In other instances, such as cells that have been infected by a virus or tumour cells, the expression of these genes can be downregulated as part of an immune-evasion mechanism (Garcia-Lora et al. 2003, Leone et al. 2013, Ishido et al. 2000, Gabor et al. 2020).

Surprisingly, the expression of tissue-wide classical genes coding for the MHC Type I heavy chain and β2-microglobulin in the oviducts was found to be generally low and even absent in many ciliated cells, a cluster that also expressed these genes in a lower proportion of cells when compared not only to secretory, but all oviduct cells (Figure 22, pg 87). On the other hand, the immune lineage cells presented the highest expression of these genes among all oviduct cell populations. Expression of the classical gene *H2-L* was not found in the oviducts, which is in accordance with the haplotype of the used mouse strain, C57BL/6. The exploration of some non-classical genes expression, which are less polymorphic and their expression is more restricted to specific tissues than classical genes, revealed that the vast majority of them was not expressed in any oviduct cell type except for *H2-T23*, which had previously been reported as a tissue-wide expressed gene in the same mouse strain (Ohtsuka et al. 2008).

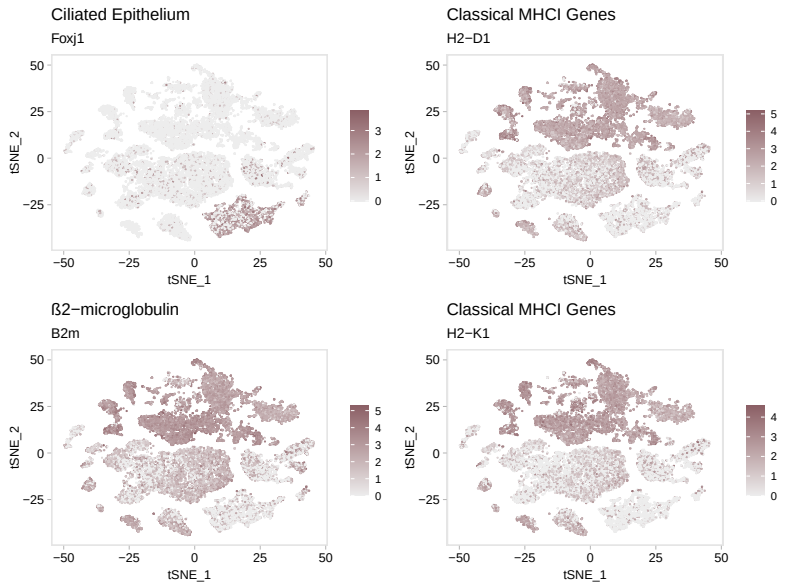


Figure 22. The oviduct gene expression of the MHC Type I components
tSNE plot showing the expression of *Foxj1* to mark the location of ciliated epithelial clusters, the classical MHCI genes *H2-K1* and *H2-D1* and β2-microglobulin (*B2m*). Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

To further investigate whether this lack of expression of the MHCI could be a technical artefact, in which these ciliated-specific mRNA molecules were not captured or amplified during library preparation or not detected during the sequencing step, a validation experiment using an alternative technique was carried out. To this end, FACS was used to profile epithelial and MHC Type I positive cells in the oviducts using antibodies against EpCAM and the MHC Type I. Using the same gating strategy as previously shown in chapter 1, the presence of epithelial alive cells (11.3% out of the total alive cell number) presenting MHC Type I expression (8.17% out of the total alive cell number) was detected, as expected (Figure 23, pg 88). However, around 18% of the epithelial cells (2.03% out of the total alive cell number) proved negative for the expression of MHCI. These were not the only cells with such characteristics, as there was another EpCAM$^-$ population that was also negative for MHCI expression.

To exclude the possibility that the epithelial and non-epithelial cells showing negativity for this complex represented contamination with mature red blood cells (RBCs), the experiment was repeated again depleting the RBCs population from the single-cell sample, which yielded very similar results with a percentage of alive epithelial cells of 15.4 and a percentage of 2.71 and 10.7 for MHCI negative and positive, respectively. Although it could not be determined if the MHCI$^-$ epithelial cells belonged to a particular epithelial subtype due to the current lack of a flow-sorting isolation protocol discriminating both populations, it would be interesting to determine whether the majority of the negative cells are of ciliated nature, as the scRNA-seq data suggest. Furthermore, using the data generated in this project, it would be feasible to select the highest DE genes between the two subtypes and scrutinise them to identify suitable specific membrane antigen panels that can help

the development of an epithelial subtype isolation protocol.

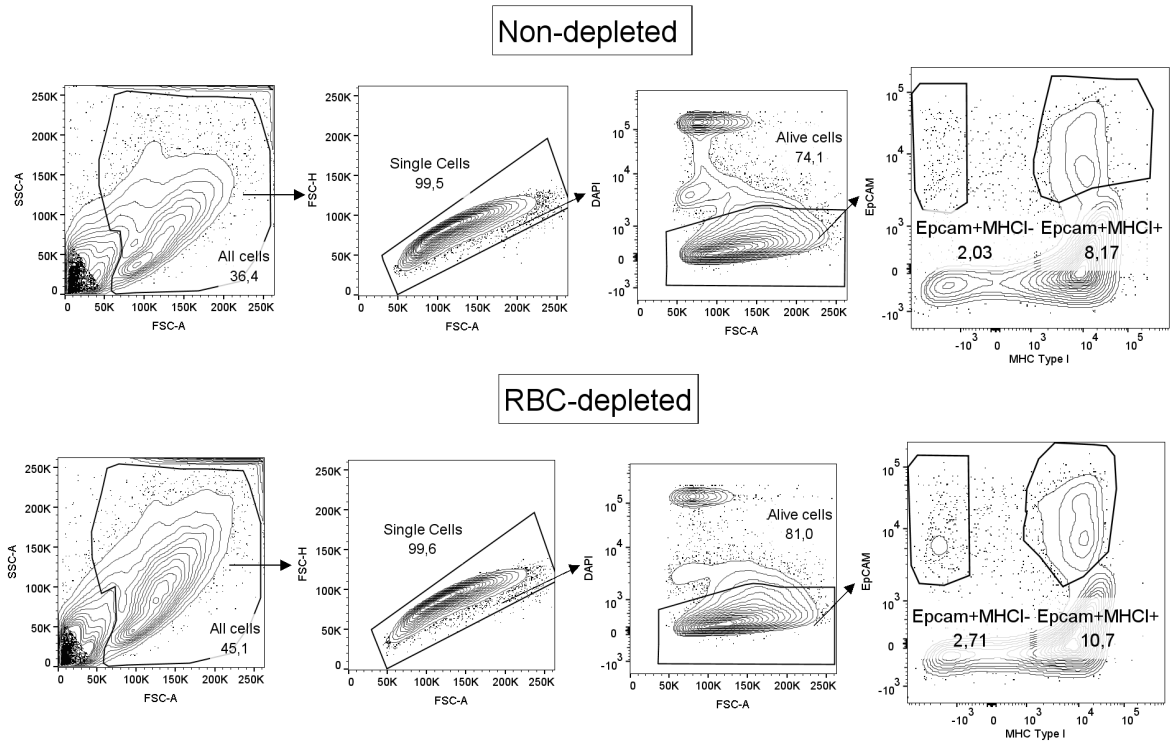


Figure 23. Gating strategy for the flow-cytometry analysis of oviduct epithelial cells expressing the MHC Type I complex
A collation of FACS plots depicting a SCC/FSC, a FSC-H/FSC-A and a viability DAPI/FSC-A selection. Alive cells ($DAPI^-$) are further selected to be classified into the epithelial and non-epithelial cells that show positivity or negativity for MHC Type I (EpCAM/MHC Type I plot). On the first row is shown the first experiment without depletion of red-blood cells (RBC) and, on the second row, is shown the same experiment after RBC depletion.

Despite the unexpected results and the historical ubiquitous-expression report of the MHCI, low/lack of this complex expression has been found before, for example, in some gastric epithelial cells (Ferron et al. 1989, Bockerstett et al. 2018, Daar et al. 1984, Neumann et al. 1997), corneal endothelium (Daar et al. 1984) and skeletal muscle (Daar et al. 1984, Fernández et al. 1991). The last example was also partially confirmed in related cell types in this project, such as smooth muscle cells and pericytes, where the MHCI expression was very low.

In summary, these and the previous analyses in this project, have focused on underlining differential characteristics between the secretory and ciliated epithelial subtypes, demonstrating a possible functional divergence such as that suggested regarding their potential immune role within the oviducts. Next, to further expand on this knowledge and to obtain a deeper insight into their differential biology and identities, their specific subdivisions were separately investigated.

2.6 In-depth characterisation of the four distinct secretory subpopulations, the two ciliated subpopulations and the two intermediate secretory-ciliated clusters

In the oviduct, all the epithelial subclusters can be finely distinguished by well-delimited phenotypes marked by the differential expression of several genes (Figure 24A, pg 91). The secretory epithelial subtype presented four previously unreported clusters Itga6_Dach2 (52.65%), Fxyd4_Dach1 (5.52%), Fxyd4 (11.39%) and Upk1a (30.44%), which displayed a variety of molecular markers (Figure 24B, pg 91).

The first one was enriched for the expression of the Integrin Subunit Alpha 6 (*Itga6*), a voltage-dependent calcium channel protein (*Cacnb2*), *Dach2* (data not shown) and a small rho-related GTP-binding protein, the Ras Homolog Family Member J (*Rhoj*). Furthermore, this cluster displayed enrichment in the expression of *Cd44*. Previously, epithelial cells displaying double positivity for *Itga6* and *Cd44* ($EPCAM^{+}CD44^{+}ITGA6^{hi}Lin^{-}$) were identified in the distal human fallopian tubes (Paik et al. 2012). These cells demonstrated an enhanced capacity for self-renewal and clonal growth, reminiscent of stem-like capacities. These cells constituted 16.6% of the epithelial cells in the fallopian tubes, although this number displayed a high inter-patient variability. The incidence of this cluster in mouse is more than 50% of the secretory cells, though.

The second cluster, Fxyd4_Dach1, was also moderately enriched for *Itga6* expression, highly enriched for the FXYD-domain containing ion transport regulator (*Fxyd4*) and, also, the transcription factors *Pax2* and *Dach1*. This co-expression of *Pax2* and *Dach1* was very interesting as they belong to the group of embryonic markers expressed by Müllerian ducts progenitors. Similarly to the secretory lineage-conserved *Pax8* expression, these markers could have been kept after embryonic differentiation, although in this particular case, in only one secretory subtype representing a clear minority, instead of the entire secretory spectrum. Co-expression of *Pax2* and *Dach1* in this subcluster, their suggested co-expression in Müllerian ducts progenitors (both have been found in Müllerian epithelial progenitors, although no conclusive imaging technique has shown their simultaneity to the best of my knowledge) as well as their co-expression in other embryonic systems, such as the otic vesicle (Heanue et al. 2002), suggest, either a conserved regulatory dependency, or functional co-operation. In chicken embryos, the existence of a regulatory network in which *Pax2* controls *Dach1* expression was discarded (Heanue et al. 2002). However, further investigation would be required to decipher a regulatory mechanism, which could rely on the *Dach1* regulation of *Pax2* expression or/and their protein-protein interaction. Interestingly, *Dach1* has been reported as a tumour-suppressor gene in cancers such as melanomas and bladder, endometrial and prostate cancers (Riggs et al. 2020). In breast cancer it has been reported as a negative regulator of *Cd44* and stem cells (Xu et al. 2017) and, at least in this cluster, co-expression of *Cd44* and *Dach1* is not found. Regarding the other marker, *Fxyd4*, its expression was found to be oestrus cycle-dependant in porcine endometrium, oviduct and ovary (Kim et al. 2018). Its expression is upregulated in these systems after the transition into metestrus and lasts until early proestrus, peaking in dioestrus, which seems to follow the rise and fall in progesterone levels. Furthermore, it has been noticed that this gene increases its expression as early pregnancy advances (increasing pro-

gesterone levels) and that a gene that *Fxyd4* modulates, *Atp1b1*, is regulated by progesterone too (Deng et al. 2013).

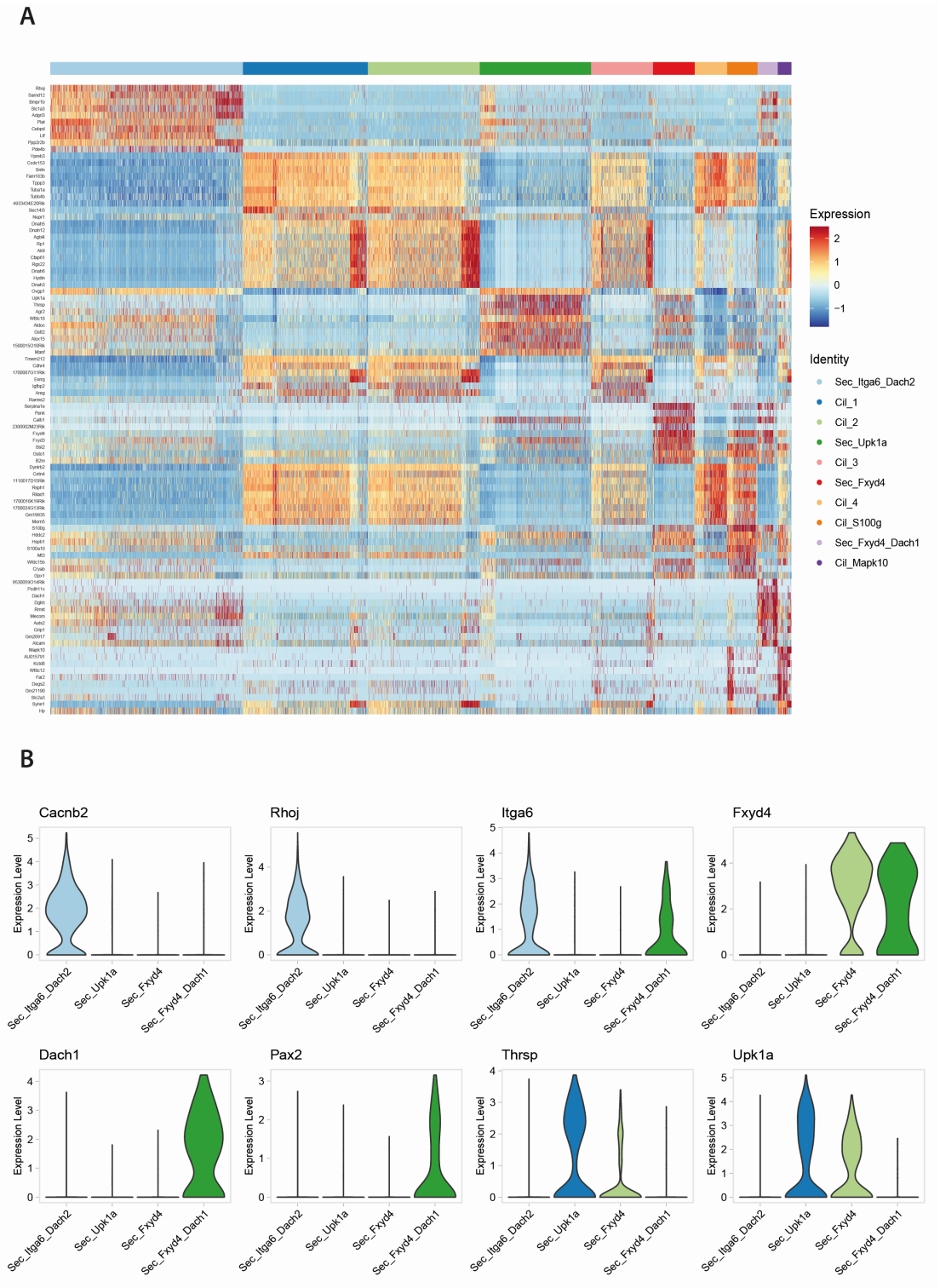


Figure 24. Differential gene expression profiles of the discrete oviduct epithelial clusters

A) Heatmap showing the centered and scaled expression of markers of the discrete oviduct epithelial clusters. B) Violin plots showing the expression of specific secretory epithelial markers. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. Centering these values is done by subtracting their average normalised expression and scaling divides the centered gene expression by their standard deviations.

Furthermore, the third secretory cluster, Sec_Fxyd4, showed enriched and moderate expression of *Fxyd4* and the uroplakin 1a (*Upk1a*), respectively. The last cluster, Sec_Upk1a, was highly enriched for the expression of *Upk1a* and also the thyroid hormone responsive protein *Thrsp*, a gene involved in lipid metabolism. Downregulation of *Upk1a* in the mouse oviduct epithelium following knockout of the ERα suggests that this is a gene controlled by this transcription factor (Winuthayanon et al. 2015, Cerny et al. 2016).

Regarding ciliated cells, they were characterised by the differential expression of cilium and cytoskeleton-related genes, such as cilia and flagella associated proteins (*Cfap20*, *Cfap126*), dynein (*Dynll1*, *Dnah5*, *Dynlrb2*), tubulin polymerization-promoting protein family member 3 (*Tppp3*). Furthermore, there was a strong presence of genes regulating the calcium metabolism, such as *Efcab10*, *Calml4*, *Capsl*, Centrins (*Cetn2* and *Cetn4*) and S100 proteins (*S100a1*, *S100a6*, *S100a11*). No major differences were found among the clusters Cil_1, 2, and 3, which showed a fully differentiated ciliated phenotype. On the other hand, Cil_4 displayed low expression of certain genes, in comparison to the other ciliated clusters. Investigation of the biological processes involving these genes using DAVID revealed annotations such as cell projection (p-value = 1.1^{-20}), cilium assembly (p-value = 4.3^{-10}), microtubule-based movement (p-value = 5.0^{-8}), microtubule motor activity (p-value = 4.2^{-10}), ATPase activity (p-value = 1.4^{-5}), epithelial cilium movement involved in determination of left/right asymmetry (p-value = 7.3^{-4}), JAK-STAT cascade involved in growth hormone signalling pathway (p-value = 8.8^{-3}). This biological picture suggests that this cluster expresses genes related to cilia formation, growth and movement at a lower level, which could indicate a less differentiated cluster or even resting, inactive or immotile ciliated cells. This would be in line with what it has been previously reported in immature ciliated cells, which present primary cilia that are nonmotile and these are substituted by their secondary motile version during maturation (Mao et al. 2018). Alternatively, the fact that pathways related to growth and metabolism are limited can be reminiscent of cellular senescence or exhaustion in this cluster of ciliated cells, fact that if true, would signify that around 9.8% of the total ciliated cells are worn-out ciliated cells going through the homeostatic process of senescence.

Regarding the intermediate clusters, both displayed a high expression of *Ovgp1* (even *Pax8* in the case of Cil_Mapk10), which are known secretory markers, in addition to ciliated markers, such as *Foxj1* (Figure 25, pg 93).

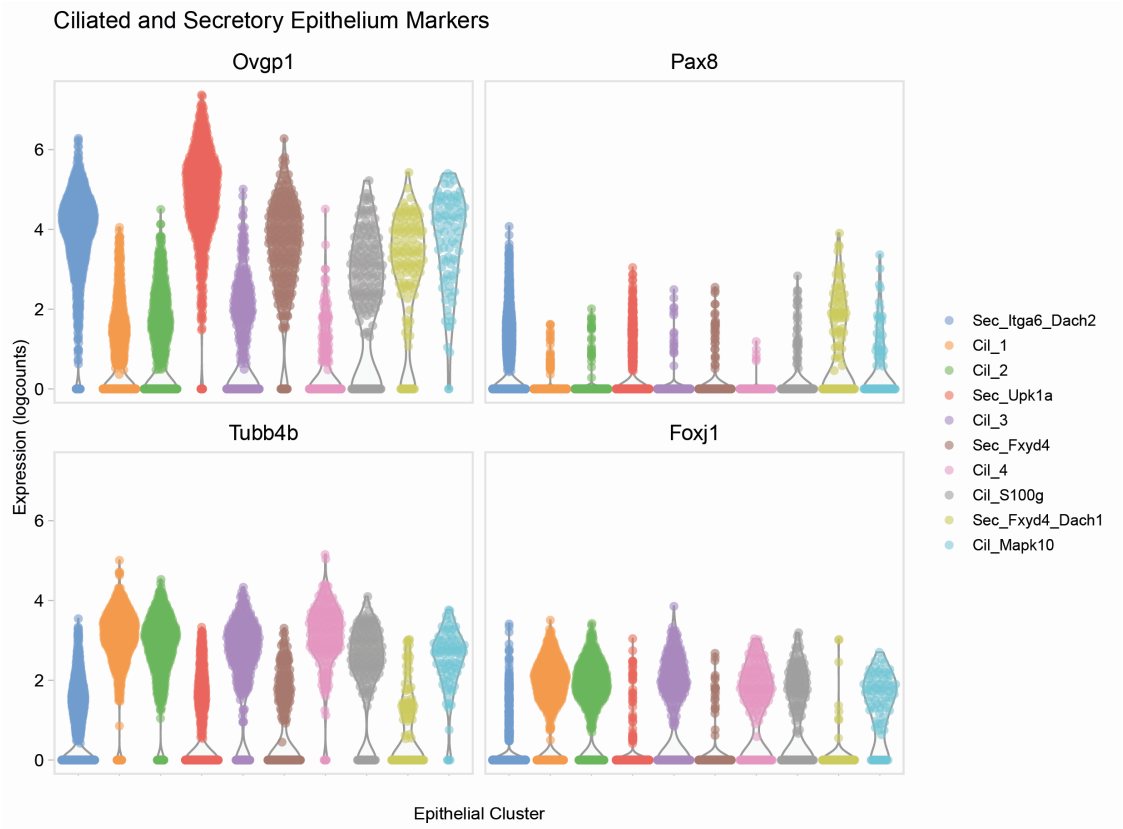


Figure 25. Expression of ciliated and secretory genetic markers in the oviduct epithelial clusters
Violin plots showing the expression of secretory and ciliated epithelial markers per scRNA-seq epithelial cluster. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

The dual nature of Cil_S100g is further demonstrated following gene ontology (GO) analyses. In a first analysis, the DEGs of Cil_S100g versus ciliated cells were used as input in the DAVID portal and the results showed secretory-specific annotated terms. When the reverse was done (DEGs of Cil_S100g versus secretory cells) processes only shared with the ciliated clusters appeared as results. Exclusive biological functions of this cluster that appeared after the GO analysis were peptidase inhibitor activity (p-value = 7.0^{-3}), negative regulation of viral genome replication (p-value=7.9^{-4}) and p53 signalling pathway (p-value= 2.2^{-2}). Similarly to Cil_4, it presented lower expression of genes related to mature and functional ciliated cells. This cluster was also marked by the DE of the S100 Calcium Binding Protein G gene *S100g* (or calbindin D9K), which is a calcium-binding protein involved in calcium absorption. This protein has been best studied in rat uterus, where its gene expression was found to be dependent on 17 beta-oestradiol, as transcription of this gene was not detected in ovariectomised mice, yet it was upregulated after oestradiol injection and inhibited by progesterone (L'Horset et al. 1990). Curiously, immature cells from 3-week-old rats displayed the same behaviour suggesting that epithelial differentiation was not needed for the upregulation of this gene. Furthermore, *S100g* was also expressed in the rat oviduct epithelium, where it was also regulated by oestrogens and demonstrated a role in the egg transport after ovula-

tion (Ríos et al. 2011). In this way, *S100g* transcripts are first detected at proestrus and peak at oestrus in rats, following the increase in oestrogen. Expression of this gene decreases at metestrus and is not detected at dioestrus, following the rise in progesterone levels. However, this pattern seems to be species-dependent given that in pigs and cows, progesterone is the one upregulating this gene (Choi et al. 2005). The same pattern is seen in the mouse uterine horns, where *S100g* is not under the regulation of oestrogen, but progesterone. In the mouse uterine horns, a basal level of *S100g* expression is detected, increasing at physiological times where progesterone levels are high, such as early pregnancy and implantation (Tatsumi et al. 1999, Nie et al. 2000). In these studies, adult females with removed ovaries increased the expression of this gene when exposed to progesterone, whereas oestrogen had no effect. Surprisingly, combination of both hormones stimulated the highest *S100g* expression (Nie et al. 2000), which could be explained by the positive regulation of the progesterone receptor by oestrogen. Upregulation in this gene expression was also observed during metestrus and dioestrus, while only basal levels were observed at proestrus and oestrus (Nie et al. 2000).

Regarding Cil_Mapk10, it displayed mature and functional ciliated-related genes when compared to secretory cells, and, interestingly, also showed an enrichment in lipid biosynthesis processes with DE of a phosphate acyltransferase (*Agpat3*), a fatty Acyl-CoA reductase (*Far2*) and the monoglyceride lipase (*Mgll*), a serine hydrolase. When DE analysis between Cil_Mapk10 and ciliated cells was run, the profile did not extensively resemble that of secretory cells. The repeated presence of lipid metabolic processes showed up, indicating that these cells display a high lipid metabolic rate like secretory cells, but with a different profile of effectors. Also, there was an enrichment in the regulation of the rho protein signal transduction (p-value = 1.0^{-4}) and the protein-tyrosine/dual specificity phosphatase (p-value= 3.5^{-6}). It was also noticeable that this cluster differentially expressed the mitogen-activated protein kinase 10 (*Mapk10*) and the stem cell antigen-1, Sca-1 (*Lya6*) (Figure 26, pg 95). The former has been reported to promote growth and migration in cancer cell lines models, being negatively regulated by the miRNA miR-205-3p, which was defined as a tumour suppressor (Qiao et al. 2020). The same study also found an inverse expression of both genes in clinical samples from patients. Sca-1 is a protein of unknown function, marker of mouse activated lymphocytes and stem cells, such as hematopoietic stem cells and resident-stem cells in a number of tissues (Holmes and Stanford 2007). Furthermore, it is associated with stemness in leukaemia, where it has been shown to promote tumour growth (Hsu et al. 2014), and breast cancer, where it promotes cell proliferation and migration (Batts et al. 2011).

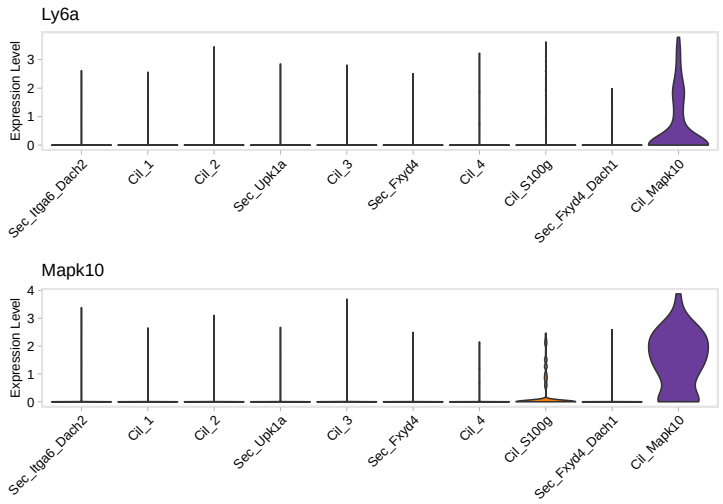


Figure 26. Gene expression of the molecular markers characterising the intermediate cluster Cil_Mapk10
Violin plots of the expression of mitogen-activated protein kinase 10 (*Mapk10*) and stem cell antigen-1, Sca-1 (*Lya6*) in the intermediate cluster Cil_Mapk10. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

Furthermore, in order to decipher whether some epithelial identities are more prone to regulation by the hormonal environment, the distribution of the expression of several hormonal receptors was surveyed (Figure 27, pg 96). This analysis showed that the secretory clusters Itga6_Dach2, Fxyd4_Dach1 and the intermediate Cil_Mapk10 displayed a high expression of the oestrogen receptor alpha. Fxyd4_Dach1 was enriched for the expression of not only this receptor, but also, for the progesterone and the thyrotropin and thyrostimulin receptors. These thyroid-related hormones are members of the same heterodimeric glycoprotein family of LH, FSH and CG. Furthermore, the Itga6_Dach2 cluster was the only cluster expressing the androgen receptor, which is activated by the binding of testosterone. In summary, it seems that these two secretory identities are the most hormone-responsive or prone to be regulated by hormones.

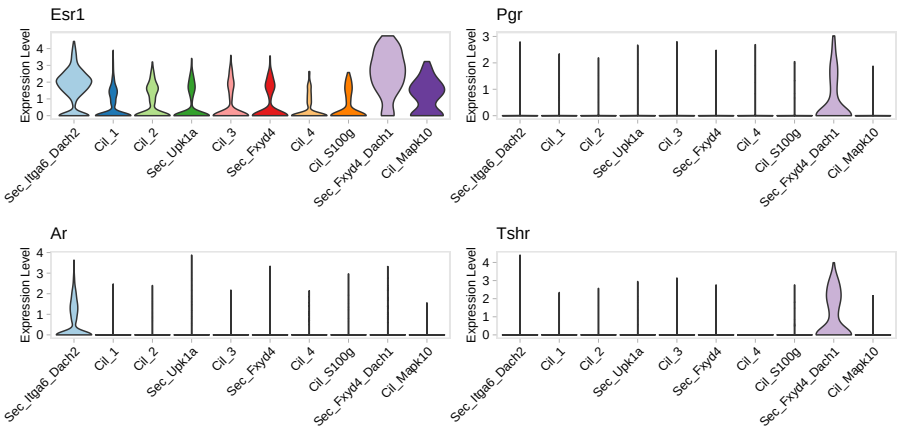


Figure 27. Gene expression of hormone receptors in the oviduct epithelial subclusters

Violin plots of the expression of hormone receptors in the individual epithelial clusters. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

2.7 An alternative single-nucleus RNA sequencing method validates and improves epithelial resolution and bypasses the stress response and the need for fresh samples

Taking everything together, scRNA-seq has provided novel and important insights into the biology behind the oviduct cellular system. However, as with every scientific approach and, in particular, data-driven results, alternative methods and different techniques should be used to validate these results. In this section, validation of the scRNA-seq results was attempted with another next-generation sequencing method, single-nucleus RNA sequencing (snRNA-seq). In this approach, RNA molecules from the nucleus are the ones primarily isolated and sequenced, while with scRNA-seq, all RNA molecules existing in the cell can be captured and processed. This method does not require FACS sorting and introduces an extra amount of flexibility since fresh, archived or snap-frozen samples can all be used as the starting material, whereas in standard scRNA-seq only fresh material can be used. Due to this feature, use of snRNA-seq with frozen oviduct samples was hypothesised to overcome the stress-response bias described in chapter 1, which originates during the cell dissociation step of the scRNA-seq protocol. To test all of the above, four single-nucleus RNA libraries were constructed from unsorted nuclei extracted from snap-frozen C57BL/6 mouse oviducts using the 10X Chromium 3' v2 protocol (Table 5, pg 97).

Sample	10x Genomics chemistry (*)	Estimated Cell Number(*)	Median Genes/Cell (*)	Median UMIs/Cell (*)	Sequencing Saturation (%) (*)	Number of High-quality (HQ) Cells (%)	Median Genes/ HQ Cell	Median UMIs/ HQ Cell
N1	v2	2.468	1.238	1.939	75.8	2090	1157.5	1758.5
N2	v2	1.450	1.483	2.386	91.7	1296	1397.5	2159
N3	v2	2.390	1.400	2.265	81.0	1906	731	1149
N4	v2	5.729	1.241	1.874	77.7	5117	1165	1718

Table 5. Technical metrics per single-nucleus RNA sample from adult mouse oviducts
Metrics marked with an asterisk (*) were part of the cellranger count output when aligning introns and exons to the mouse reference genome. Metrics without asterisks were calculated after the quality control steps of cell filtering by metrics and multiplet status. The abbreviation HQ stands for High Quality.

These libraries were analysed in a similar fashion to the scRNA-seq ones, using cellranger count and testing exon alignment versus exon plus intron alignment. For the nucleus-based approach, the difference in performance between the two methods was of two-fold-increase in the number of detected genes per cell when the combinatorial exon and intro alignment was used. The sharp difference in yield using one type of alignment or another, could be explained given the comparative distribution of the read types captured by each method (Figure 28, pg 98). Whereas the majority of reads in scRNA-seq libraries aligned to exonic regions of the genome, a higher percentage mapped to intronic regions in snRNA-seq libraries, which can be explained by the fact that, when isolating transcripts from the cell nucleus, a higher proportion of them are immature RNA molecules that have not undertaken splicing processes compared to isolation of transcripts from the whole cell, in which the majority of the transcripts have been processed. However, it is also to note that

the new v3 chemistry of 10X Genomics seemed to capture a proportion of intronic reads similar to those of snRNA-seq (see Figure 28, samples S1, S2, S6, S8).

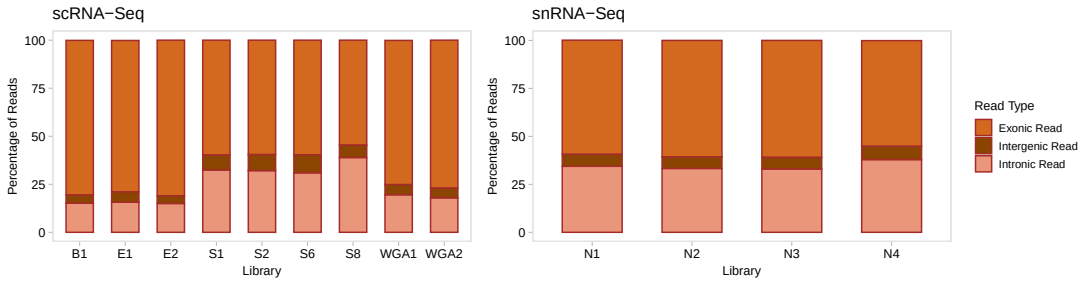


Figure 28. Distribution of read types per sample in each transcriptomic sequencing technique
Bar charts depict the distribution of read types (exonic, intergenic and intronic) per sample when each transcriptomic technique was used.

A total of 10.409 nuclei passed the quality controls and were retained for subsequent analysis. Cell lineage annotation of this dataset revealed differences between the cell-based and the nucleus-based methods (Figure 29A, pg 100). The nucleus-based method yielded a clear majority of epithelial cells (65,38%), whereas in the cell-based method, the percentage of recovered epithelial cells had been a technical limitation even after FACS enrichment (70% epithelial sample composition after FACS enrichment versus 19,98% in scRNA-seq data). Interestingly, cell types such as the immune and mesenchymal ones were missing or extremely underrepresented. This could be due to a previously detected discrimination against small cells in this technique, where immune cells had to be enriched using supplementary scRNA-seq (Litviňuková et al. 2020). Compared to epithelial cells, mesenchymal and the majority of immune cells present a very reduced size, which could be the explanation for their loss in the snRNA-seq data. In line with this, the small number of detected immune cells (0.27%) in the snRNA-seq dataset present macrophage markers, a population of big-sized immune cells.

The epithelial overrepresentation achieved in these unsorted samples offered a unique opportunity to further explore and gain deeper understanding of this phenotype. First, unsupervised clustering of the epithelial type was undertaken, achieving the identification of the secretory and ciliated subtypes (Figure 29B, pg 100). In this case, *Pax8* was not a dropout, but a reliable secretory marker, and another ciliated marker, *Tmem212*, with higher expression than *Foxj1*, was selected as ciliated marker. Second, in order to identify and validate whether the same epithelial identities had been captured by snRNA-seq, a gene signature for each epithelial cluster detected in the scRNA-seq dataset was generated, based on their best differentially-expressed markers. This signature was then applied and scored to the snRNA-seq dataset, treated as a query (Figure 29C, pg 100). The signature performed satisfactorily and enabled the identification of all previously described clusters and a bad-quality one (cluster 4) that did not correlate with any signature. This correlation analysis revealed that snRNA-seq increased the detection resolution of the epithelial subtype clusters. This dataset broke down the initial secretory Itga6_Dach2 and Fxyd4_Dach1 clusters into four and three different subclusters, re-

spectively. It also detected the Upk1a and Fxyd4 clusters. Regarding the ciliated phenotype, fully mature cells were identified as Cil_1 in two single-nucleus clusters (cluster 2 and 14) and the Cil_4 phenotype that was proposed to be exhausted or immature, was also identified, in this case, in close proximity to some secretory clusters. Furthermore, the two intermediate clusters Cil_S100g and Cil_Mapk10 were also identified as clusters 13 and 11, respectively. The proportions of the detected epithelial clusters also varied in the snRNA-seq dataset.

Last, to test the hypothesis that this technique would avoid the cell stress response, the expression of the IEGs *Fos* and *Junb* was assessed. These genes were highly upregulated in the scRNA-seq dataset, however, they were not expressed in this dataset (Figure 29D, pg 100). This lack of expression is in line with their absence in undisturbed oviducts demonstrated by qPCR analysis in chapter 1, suggesting that snRNA-seq enables the capture of gene expression profiles resembling the naïve oviduct.

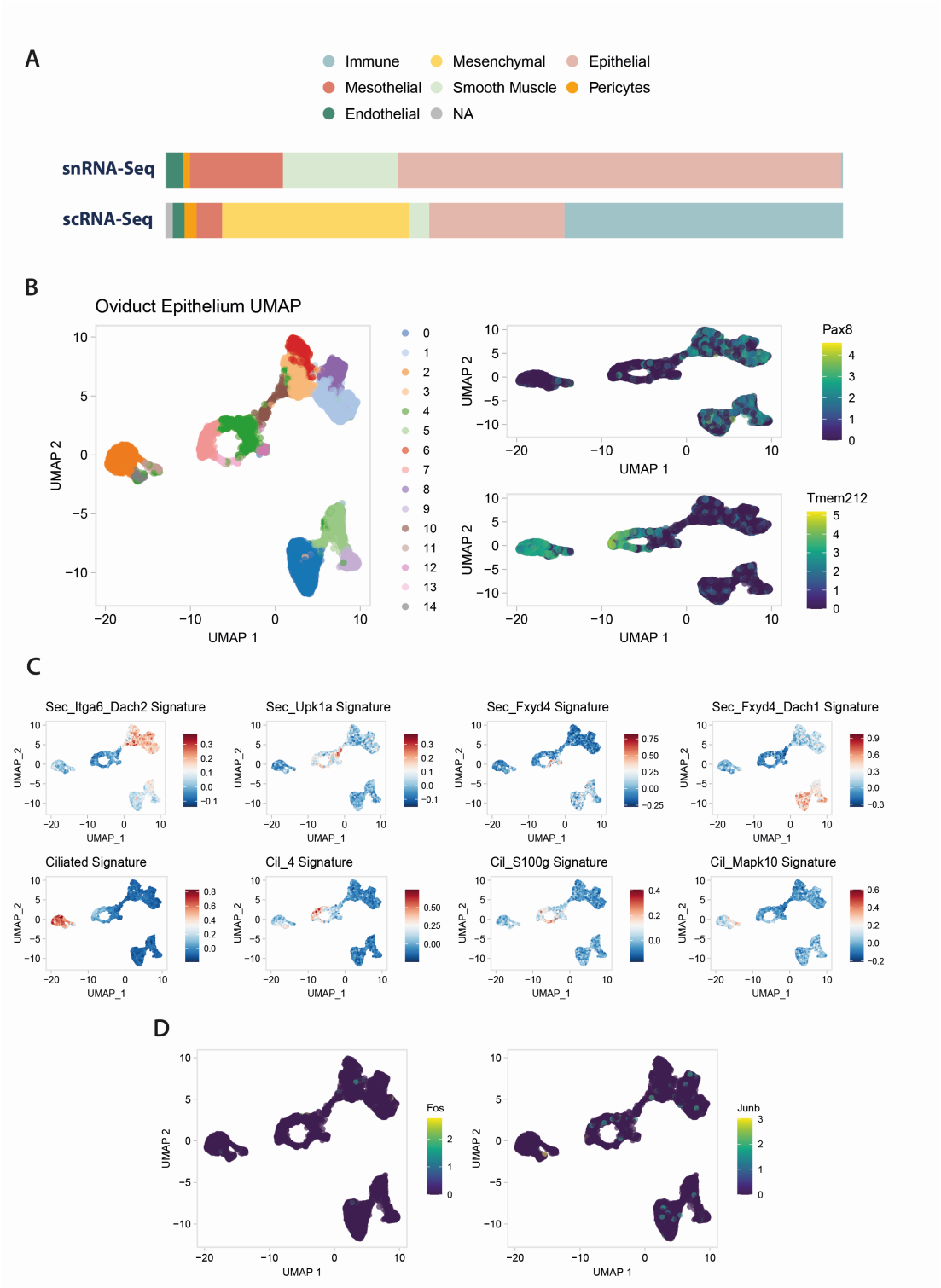


Figure 29. Single-nucleus RNA-Seq characterisation of the oviduct cell-lineage abundances, validation of the scRNA-seq epithelial identities and absence of stress response

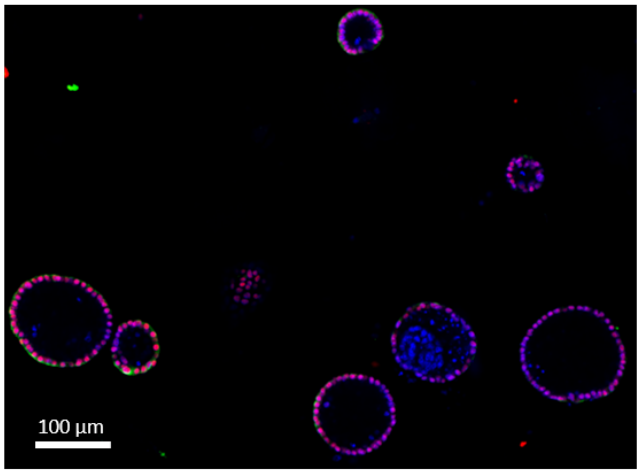
A) Comparison of the detected oviduct lineages abundances in the two transcriptomic methods B) UMAP plot of single oviduct epithelial nuclei coloured by individual clusters (left) and the expression of a secretory marker, *Pax8*, and a ciliated marker, *Tmem212* (right). C) scRNA-seq epithelial clusters scores shown in the UMAP embedding of single epithelial nuclei. D) UMAP plot of single epithelial nuclei showing negative expression of the IEGs *Fos* and *Junb*. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

2.8 Further validation is accomplished by 3D organoid culture, immunohistochemistry and fluorescence microscopy

Further validation experiments were carried out using alternative techniques relying on cellular and molecular biology. First, organoids culture from oviduct epithelial cells was successfully achieved using a concentration of 25.000 cells/50 µl Matrigel. With this approach, small organoid formation was observed as early as culture day 1 and a range of organoid sizes (25-400 µm) were detected after 3 weeks of culture (Figure 30A, pg 102). This time point was chosen to allow secretory cells to differentiate into ciliated cells. Interestingly, protrusions resembling the oviduct folds, characteristic of mature and fully differentiated organoids, were also detected at this time point (Figure 30B, pg 102). To test whether the epithelial differentiation level described in the transcriptomic data was achieved and detected in this model, immunofluorescence against Pax8 and Tubb4 was carried out. Both secretory ($Pax8^+$) and ciliated ($Tubb4^+$) cells were detected in the organoids and, unexpectedly, the vast majority of detected ciliated cells were also positive for the secretory marker Pax8. This phenomenon could have a technical or a biological origin and warrants further investigation, even more, in light of the fact that a limited number of ciliated and, even less, double positive cells, was previously reported in human and mouse oviduct organoids (Paik et al. 2012). Although the presence of ciliated and secretory cells was observed, the unconventional ciliated phenotype and project time constrains limited the use of this *in vitro* model for validating the existence and role of the epithelial subpopulations and, concretely, the secretory and intermediate subclusters.

However, other experiments were run to partially meet this goal in tissue sections from snap frozen mouse oviducts. The first selected cluster to validate was the Sec_Fxyd4_Dach1. This cluster has previously been shown to express *Pax2* as a marker and, also, Cd13, a membrane protein coded by the gene *Anpep* (Figure 31A, pg 103). Antibodies for both markers were tried in oviduct sections and exhibited different expression patterns. Whereas epithelial cells positive for Cd13 were only found on the oviduct section morphologically corresponding to the uterotubal junction, epithelial cells positive for Pax2 were spread along the entire oviduct. Most cells in the proximal oviduct were positive for this transcription factor but positivity was reduced towards the end of the ducts, which could be due to the increase in ciliated cells (negative for Pax2) or an actual decrease of $Pax2^+$ secretory cells. To understand which hypothesis was correct, serial oviduct sections were stained for Pax8 and Pax2, and revealed that all cells that were $Pax2^+$ were also $Pax8^+$, but not all secretory cells were $Pax2^+$. Furthermore, double staining for Cd13/Pax8 or Cd13/Pax2 in serial sections showed the presence of secretory Fxyd4_Dach1 cells in the uterotubal section (Figure 31B, pg 103).

A



B

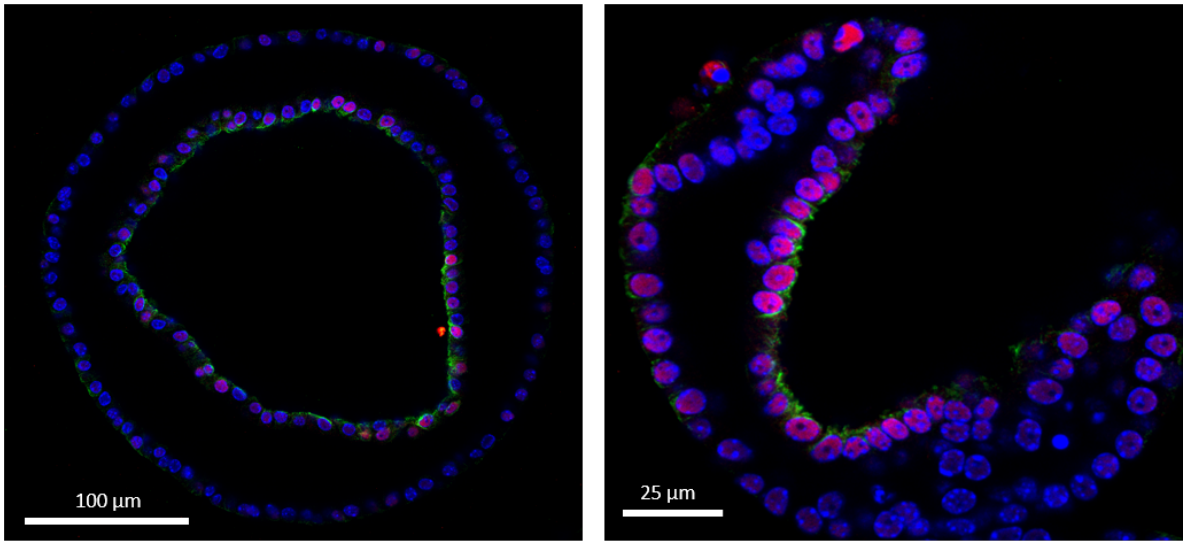


Figure 30. Immunofluorescence of oviduct epithelial organoids
A) Immunofluorescence of oviduct epithelial organoids after three weeks of culture at 10X magnification showing the presence of Pax8 (magenta), Tubb4 (green) and DAPI (blue). B) Immunofluorescence of oviduct epithelial organoids after three weeks of culture showing epithelial protrusions at 25X (left) and 40X magnification (right).

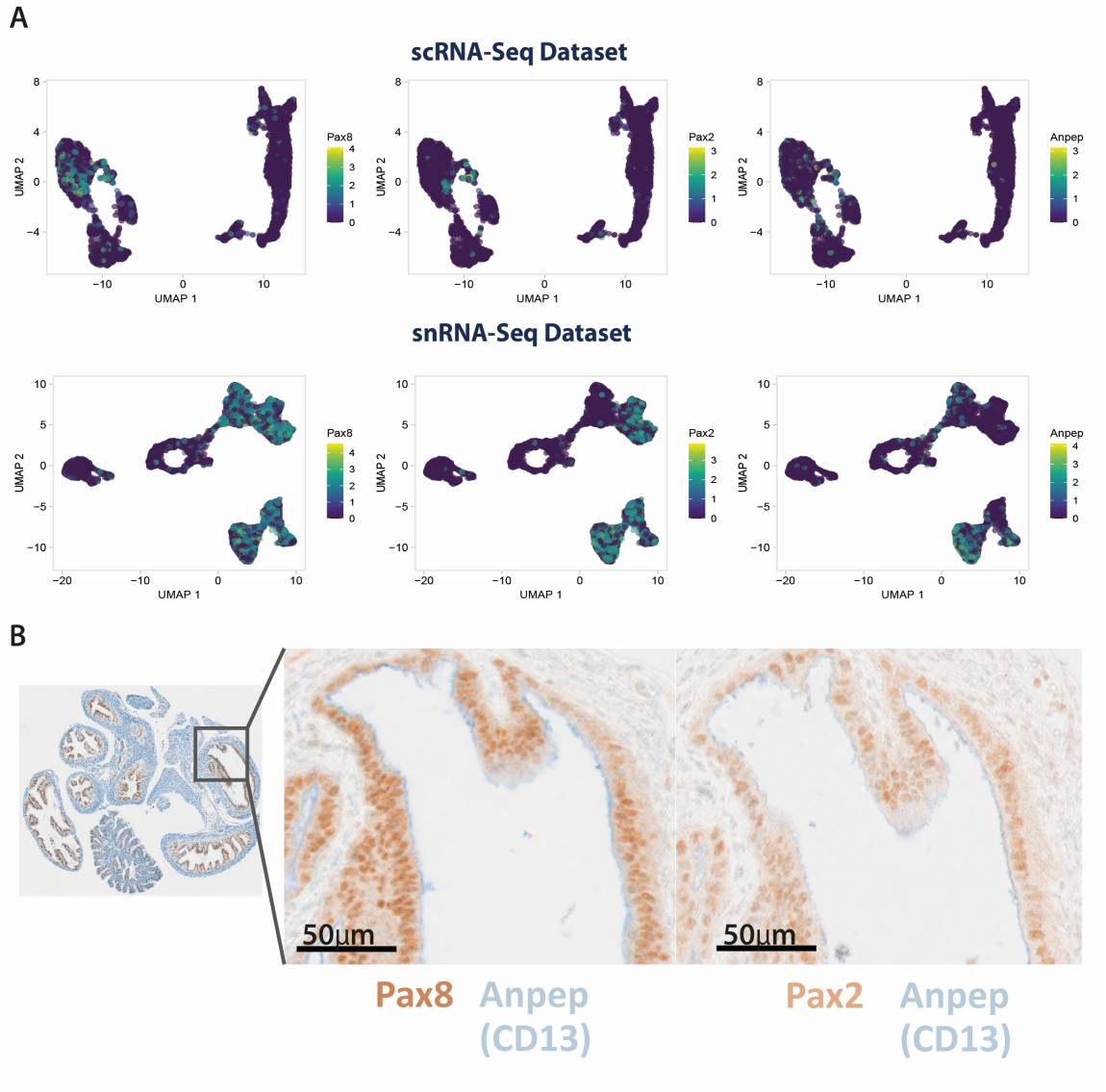


Figure 31. Validation of the Fxyd4_Dach1 epithelial secretory subcluster by IHC

A) UMAP plot of single oviduct epithelial cells (upper row) and nuclei (lower row) coloured by the expression of *Pax8* (left), *Pax2* (middle) and *Anpep* (Cd13) (right). Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. B) Dual immunohistochemistry of serial snap-frozen adult oviduct sections at magnification 20X. On the left, Pax8 is brown and Cd13 is blue and, on the right, Pax2 is brown and Cd13 is blue.

2.9 The mucosal stroma is mainly comprised of mesenchymal, vascular, smooth-muscle and mesothelial cell populations

scRNA-seq analysis showed that the stromal compartment is comprised of thirteen populations of mesenchymal and endothelial cells, pericytes and, the external lineages wrapping the oviducts, smooth-muscle cells and mesothelium. The principal component of the oviduct stroma are mesenchymal cells (almost 30% of the total sample), which are organised into eight populations, of which, MesCycling, is a cell cycling population (Figure 32, pg 105). These populations are positive for the

markers *Dcn*, *Col1a1* and *Col1a2* and share a high degree of similarity, which unveiled shared activated gene programmes. Among these mesenchymal markers, the hormone-regulated expression of midkine (*Mdk*) and pleiotrophin (*Ptn*) were found in this type too. These proteins are secreted growth factors of the same family, regulated by progesterone and oestradiol. They have been reported to have a role in the reproduction ability of female mice, although their expression had only previously been reported in ovarian and uterine epithelium (Muramatsu et al. 2006, Fischer et al. 2018).

Within the mesenchymal clusters, Mes1 and Mes3 are homogeneous clusters that share the expression of Insulin Like Growth Factor Binding Protein 3 (*Igfbp3*), a protein that prolongs the half-life of insulin-like growth factors (IGFs). Mes1 and Mes3 also express C-Type Lectin Domain Family 3 Member B (*Clec3b*) and BMP Binding Endothelial Regulator (*Bmper*), a BMP inhibitor. A differentiating feature of Mes3 versus Mes1 is the expression of the carboxypeptidase Z (*Cpz*). More divergent clusters are: (1) Mes4 with the expression of Anoctamin 4 (*Ano4*) and CUB And Sushi Multiple Domains 1 (*Csmd1*); (2) Mes5 with the expression of the palmitoleoyl-protein carboxylesterase *Notum* and Dickkopf WNT Signaling Pathway Inhibitor 2 (*Dkk2*), both of them inhibitors of the canonical Wnt pathway, the relaxin receptor *Rxfp1* and the transcription factor ETS Variant Transcription Factor 1 (*Etv1*); (3) Mes6 with the expression of the voltage-gated potassium channel *Kcnd2*; (4) Mes7 with the expression of the Leucine Rich Repeat Containing G Protein-Coupled Receptor 6 (*Lgr6*), the GDNF Family Receptor Alpha 1 (*Gfra1*), the growth factor *Fgf12* and the progesterone receptor *Pgr*.

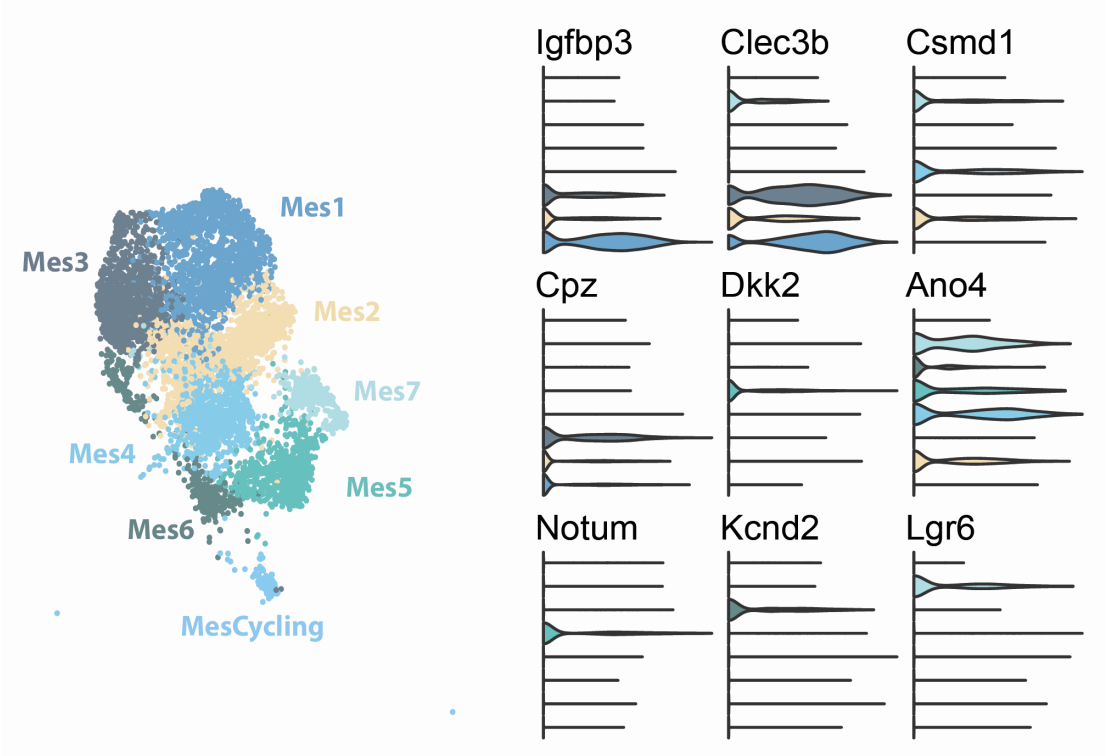


Figure 32. Single-cell RNA-Seq characterisation of the oviduct mesenchymal lineage
UMAP plot of single oviduct mesenchymal cells (left) coloured by the discrete cluster identities. On the right, violin plots showing the expression of molecular markers of the single mesenchymal clusters.

Mesothelial cells also accounted for one cluster that presented specific markers, such as *Upk3b* and *Upk1b*. They also showed intermediate features between epithelial and mesenchymal cells, for example, with the expression of mesenchymal markers and keratin, the intermediate filament found in epithelial cells (Figure 19, pg 79). Smooth muscle cells were less prominent and only one cluster, accounting for less than 3% of the total sample, was detected. They were characterised by the expression of pan-smooth muscle cell markers such as *Cnn1*, *Acta2* and *Tagln*. Furthermore, also endothelial cells and pericytes were detected, as expected due to the presence of lymphatic and blood vessels in the oviducts. One unique identity characterised pericytes by the expression of smooth muscle cell markers as well as phospholamban (*Pln*), *Ctnna3* and *Tinagl1*. On the other hand, endothelial cells were represented by two clusters, which corresponded to the lymphatic (31.76%) and blood (68.24%) vessels draining and feeding the oviducts (Figure 33, pg 106). Previously, comprehensive scRNA-seq analysis on murine endothelial cells by Kalucka et al. 2020 revealed an extensive heterogeneity for this type across different tissues, a characteristic that was also found in this project. Although both clusters were the same cell type and shared the markers *Egfl7* and *Pecam1*, they were well differentiated by their specific gene signature and separated in the oviduct tSNE (Figure 18A, pg 78) and UMAP (data not shown), highlighting their functional difference. Further insights into the specifics of the oviduct endothelial lineages could be achieved by comparative analysis against other endothelial identities found in other tissues.

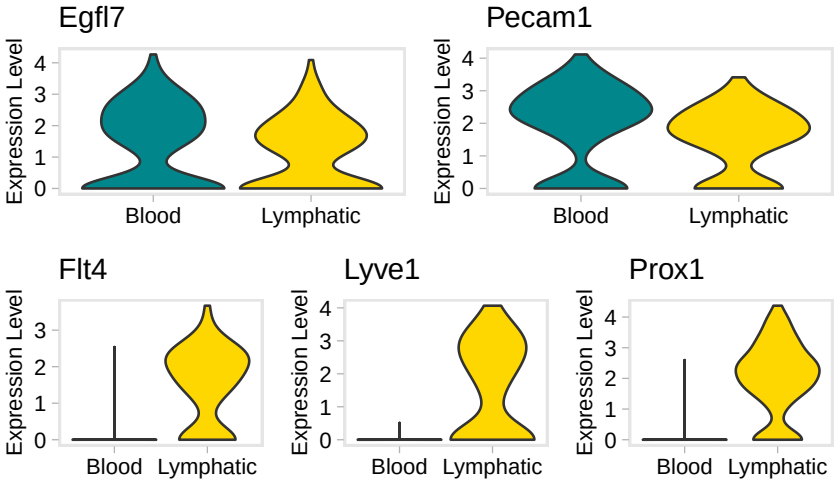


Figure 33. Identification of the two oviduct endothelial cell clusters
Violin plots depicting the expression level of the generic endothelial markers *Egfl7* and *Pecam1* and the lymphatic vascular vessel markers *Flt4*, *Lyve1* and *Prox1*. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

2.10 The single-cell immune landscape in the adult murine oviducts reveals a heterogeneous population of macrophages, dendritic cells, T and B lymphocytes and natural killer (NK) cells

The scRNA-seq data revealed a complex and miscellaneous group of leucocytes populating the mouse oviducts. Members of both, the myeloid and lymphoid lineages, were detected (Figure 34, pg 107). The former accounted for fourteen clusters and the latter for four clusters (Figure 35A, pg 108). Furthermore, two clusters were removed from subsequent analysis given their bad quality. Within the lymphoid lineage, B cells, T cells and NK cells were detected (Figure 35B, pg 108). B cells were identified by the expression of the Immunoglobulin Kappa Constant gene (*Igkc*), which codes for the constant region of the antibodies light chain. NK cells were marked by the expression of the Natural Killer Cell Granule Protein 7 (*Nkg7*) and T cells by subcomponents of the CD3 complex (*Cd3d*, *Cd3e* and *Cd3g*). Furthermore, nuocytes were distinguished in the immune repertoire by the absence of markers for the other lymphoid cell types and the expression of the transcription factor *Gata3*, which has been related to nuocyte's development and maintenance (Hoyler et al. 2012, Yagi et al. 2014). These cells are relatively newly-discovered lymphoid immune cells with key roles in innate type 2 immunity (Neill et al. 2010, Moro et al. 2010, Wong et al. 2012). The most common lymphoid cells in the oviducts were NK cells, closely followed by nuocytes, which was surprising, given that this last cell type is reported as rare.

Among the myeloid cells, all of them were phagocytes: neutrophils, dendritic cells and macrophages. The vast majority of these cells corresponded to macrophages, with nine groups identified, which unveiled an unexpected large and heterogenous group within the oviducts. The rest of the myeloid cells comprised a big group of dendritic cells expressing the myeloid lectin-like receptor *Clec9a* and

the serotonin receptor *Htr7*, plasmacytoid dendritic cells (*Siglech*$^{+}$) and two groups of neutrophils, one standard (*Cd177*$^{+}$), and one activated (*Cxcr2*$^{+}$) (Figure 35C, pg 108). Furthermore, an independent cell cycling cluster was identified in what seems a subset of the M_Cd209, although cycling cells were also detected as part of clusters of T cells, nuocytes, M_C1qb and dendritic cells. This finding could represent an enhanced proliferative capacity of these groups.

Regarding macrophages, a large homogeneous group was distinctively classified by the expression of the receptor *Cd209a*, that is involved in pathogen recognition (Figure 35D, pg 108). Three other clusters showed an elevated expression of *C1qb* (a component of the complement system). Two of these complement-activated clusters of macrophages only differed from the third in the expression of few genes such as the chemokine *Pf4*, in one, and *Pmepa1*, a negative regulator of the androgen receptor and TGFβ signalling, in the other. This suggests that the cluster M_C1qb could represent a resting macrophage state, whereas the M_C1qb_Pf4 and M_C1qb_Pmepa1 could depict other cellular stages, more likely an immunoreactive/pro-inflammatory one. Less abundant macrophages are the other five clusters, positive for the adipokine *Retnla*, melanoregulin (*Mreg*), a G protein-coupled receptor (*Adgre4*), the chemoattractant *Saa3* and, finally, the rare cluster M_Cd300e, which differentially expresses the gene *Cd300e*, a cell surface glycoprotein (Figure 35D, pg 108).

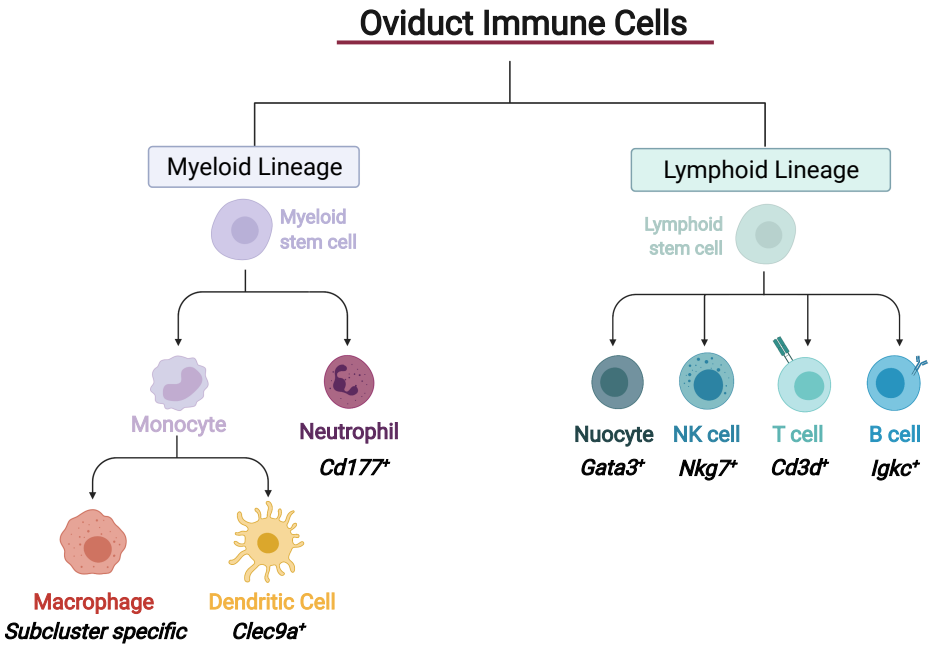


Figure 34. Diagram of the detected oviduct immune cell repertoire
This diagram shows all the immune cell types detected in the mouse oviduct scRNA-seq dataset with bright colours. Each detected cell type is associated to an oviduct-specific marker gene. Those with faint colours and labels and lacking genetic markers were not detected, but have been added for lineage-clarification purposes. Created with Biorender.

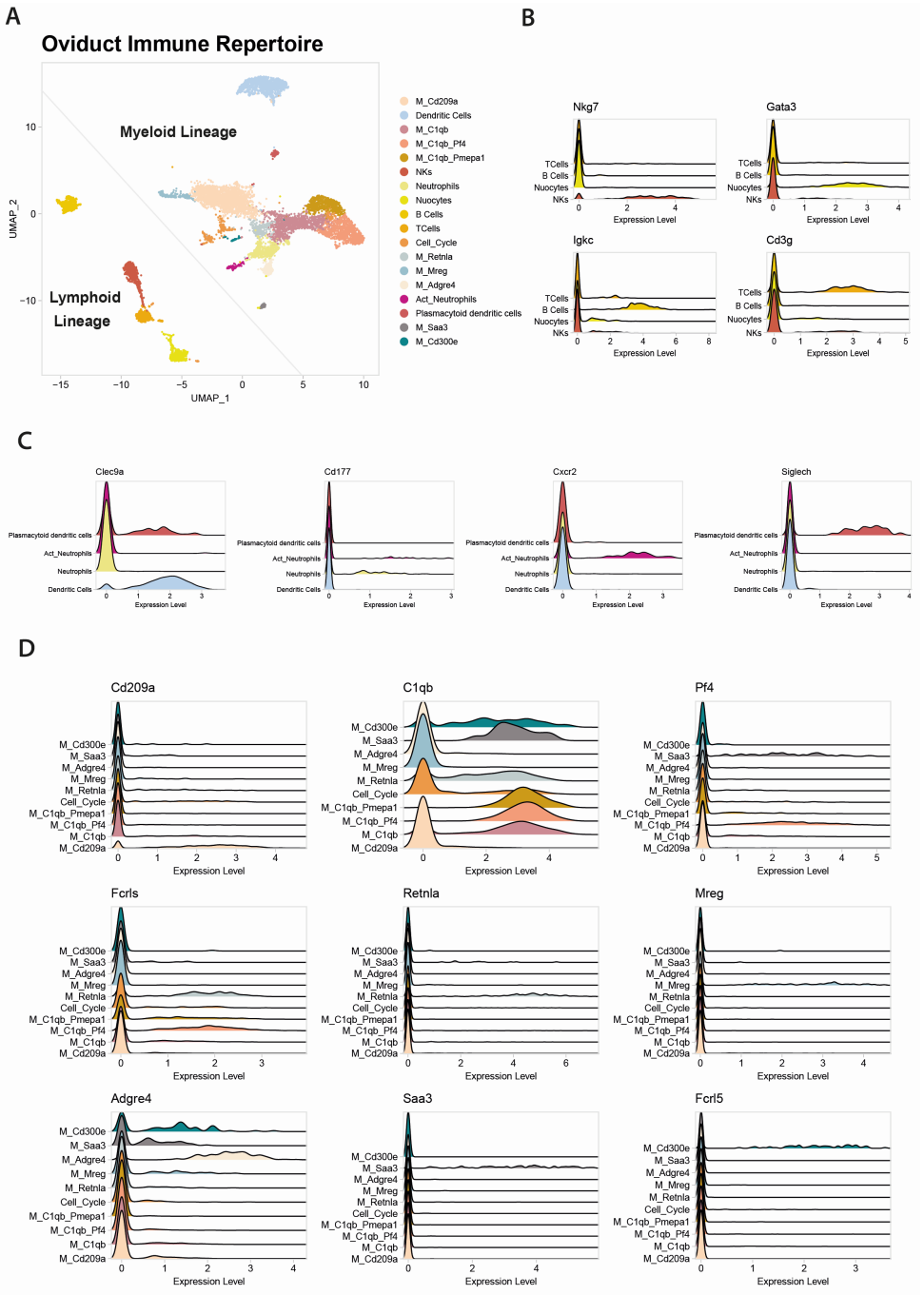


Figure 35. Identification and further classification of the oviduct immune repertoire

A) UMAP plot of the oviduct immune repertoire coloured by individual clusters and specifying the myeloid and the lymphoid lineages. B) Ridge plots showing the expression of molecular markers for the lymphoid identities. C) Expression of novel DEGs for all myeloid clusters, except macrophages. D) Expression of molecular markers characterising the individual macrophage populations. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

Compelling evidence has shown that cells of the lymphoid lineage are highly conserved across the organ-specific immune repertoires versus the heterogeneity pre-

sented by the myeloid lineage, especially by macrophages (Litviňuková et al. 2020, Gautier et al. 2012, Davies et al. 2013). A related and contemporary discovery in the immunological field is the dual origin of macrophages, which used to be considered the exclusive product of hematopoietic precursors. In this traditional theory, all macrophages would follow the mononuclear phagocyte system, where a bone marrow progenitor differentiates into blood monocytes that circulate in the body and infiltrate specific tissues, subsequently differentiating into tissue-resident or shortly-infiltrating macrophages. Nowadays, it has been proven that some tissue-resident macrophages can also derive from early embryonic primitive populations from the yolk sac before hematopoietic stem cells arise (Schulz et al. 2012). These macrophages then travel and seed different tissues, being able of local self-renewal and independent maintenance. Whereas some tissue-resident macrophages, such as the microglia, are mainly derived from the embryonic lineage (Ginhoux et al. 2010), others, such as Langerhans cells, show a mixed origin. Tissue-resident macrophages are also extremely different depending on the tissue they reside in or act upon, heterogeneity that is also reflected in their roles, which range from the standard homeostasis maintenance and immune surveillance to tissue-specific functions (Davies et al. 2013).

This reported heterogeneity in function and origin, along with the unexpected amalgam of phenotypes found in the oviduct scRNA-seq data, prompted further study of macrophages. Virtually all the tissues in the body contain tissue-resident macrophages, but this class had not been described in the oviducts before. In this project, several novel identities have been identified, deserving functional exploration and validation. However, there is still an outstanding question in the field, which these data could help answering: is it possible to define the developmental lineage of the tissue-resident macrophages in the oviducts?

To this end, the first approach was to investigate the expression of a classical triad of yolk-sac derived macrophages markers (Figure 36A, pg 110). First, the F4/80 receptor encoded by the gene *Adgre1* was investigated. This marker is highly expressed by a number of embryonic-lineage tissue macrophages such as some Langerhans cells, Kupffer cells, microglia, pancreas, lung, kidney, peritoneal and red pulp macrophages (Schulz et al. 2012, Yona et al. 2013). This marker is lowly expressed by hematopoietic-derived tissue macrophages. Second, the expression of the Mer tyrosine kinase receptor (*Mertk*) and the receptor *Csf1r*, preferentially expressed by tissue-resident macrophages, were also surveyed (Gautier et al. 2012). The third analysis I carried out, was to apply a score signature of 39 genes that Gautier et al. 2012 developed, to distinguish mouse tissue-resident macrophages from those arising from the peripherally-infiltrating lineage (Figure 36B, pg 110). When that signature was applied to the oviduct immune repertoire, the best scores were achieved by M_C1qb, M_C1qb_Pf4, M_C1qb_Pmepa1, M_Saa3 and M_Cd300e, which were the same clusters mainly positive for the marker triad. Whereas this analysis proposes an attractive starting point to confirm the identity of embryonic-lineage macrophages in the oviducts, lineage-tracing and fate-mapping experiments are necessary to definitively conclude these findings. Moreover, it would not be unexpected to find that these oviduct macrophages populations are of mixed origins, as a number of tissues have been described to contain macrophages of both origins (Schulz et al. 2012, Yona et al. 2013).

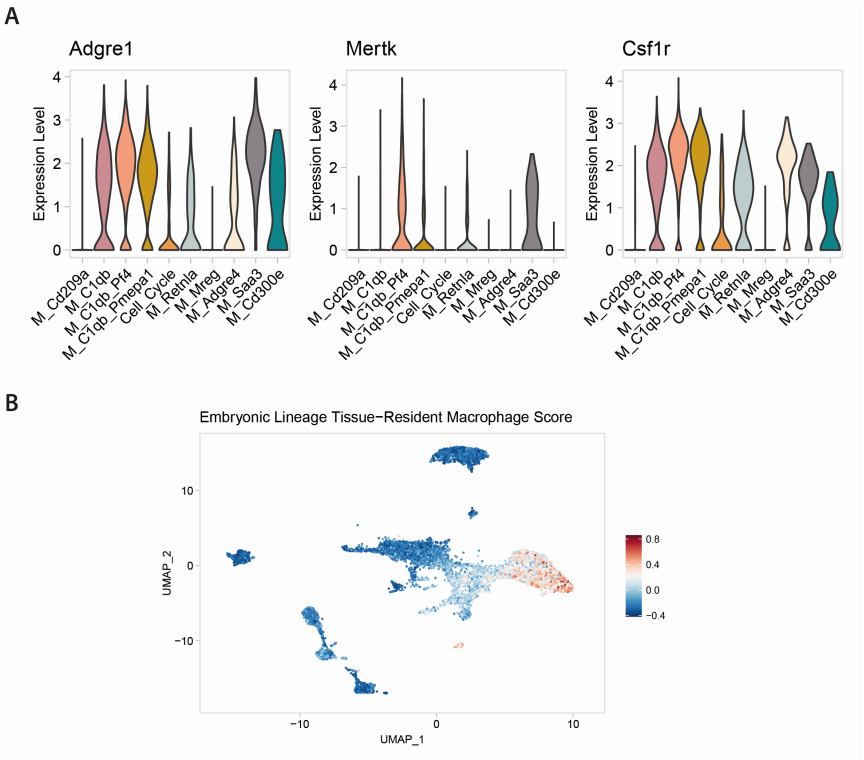


Figure 36. Analysis of the tissue-resident origin of oviduct macrophages
A) Violin plots detecting the expression of genes related to the embryonic macrophage lineage. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation. B) Tissue-resident macrophages signature scores shown in the UMAP plot of single immune cells.

2.11 Conclusion

In this chapter, a thorough and comprehensive analysis of the single transcriptomes of almost 35k cells of the adult mouse oviduct is presented, while also addressing the technical stress response presented in chapter 1. Moreover, an initial validation of intriguing evidences extracted from these data has been carried out. This investigation has unravelled a broad number of molecular cell states from all the cell lineages comprising the oviducts, which has provided a major insight into the biology of this healthy mouse organ for the first time in the field. The detection of quantitative attributes measured by scRNA-seq discovered and enabled the cataloguing of more than 40 cellular and molecular signatures belonging to the epithelial, mesenchymal, smooth-muscle, endothelial, mesothelial and immune lineages, refining the existing classifications and defining novel identities. These data not only serve as a single-cell atlas of the oviducts, but it holds the potential of fuelling further research and discoveries in the field. These findings represent a framework for the study of: (1) homeostatic and pathological processes in the oviducts; (2) cyclic hormonal-dependent physiological states and transcriptional programmes; (3) developmental lineages, which could help deciphering the early and late stem cell identities maintaining this organ from conception to adulthood; (4) the regulatory circuitries; (5) the cell-to-cell interaction landscape.

Despite the efforts with the FACS enrichment of the elusive epithelial subtypes to represent 70% of the sample before creating the scRNA-seq libraries, they ultimately represented 20% of the final data. This sharp shift from the originally sorted oviduct sample might be due to the preferential recovery or survival of immune and mesenchymal cells after FACS and during cell diffusion in the narrow channels of the 10X chip. Alternatively, an epithelial death prompting their loss during the initial low-quality cell filtering steps during data analysis could be another plausible explanation. Although this denotes a profound intrinsic technical bias of scRNA-seq affecting epithelial representation, it was overcome by the use of nuclear transcriptomic data. The use of snRNA-seq not only refined the capture and detection of epithelial clusters by scRNA-seq, but also bypassed the necessity of using fresh samples, which enables the retrospective use of archived or frozen samples. This method was also able to avoid the transcriptomic stress signature detected with scRNA-seq and has been further reported to preferentially allow the sequencing and capture of big-sized cells and difficult-to-dissociate samples that could not be captured with scRNA-seq and isolated, respectively. Therefore, this method provided a complementary line of evidence validating the epithelial subclusters described with scRNA-seq and, because of its advantages, it can be suggested as a better method to employ in cases where immune and mesenchymal profiling is not required, given its higher ability to offer a better and more accurate sampling of cells in their native environment and unperturbed transcriptomic state. If the loss of the immune and mesenchymal components was a drawback for the experimental goal, it could still be used as a complementary tool in combination with scRNA-seq.

These transcriptomic technologies detected not only cell types and canonical molecular markers previously reported, but also novel cell clusters and marker genes, found through differential expression analysis. The discovery of these new markers facilitates the genesis of new cell type isolation/identification strategies through the design of antibodies panels against DE proteins. This could also help in tracing a cell type of interest and studying its functions either *in vivo*, through the creation of transgenic mouse lines, or *in vitro*, through the establishment of cell culture models. To this end, the successful establishment of organoids lines in this project, a three-dimensional oviduct system model, could propel its systematic and unlimited study, offering the opportunity to establish genetically-modified organoids lines and carry out functional analyses. An example of an interesting experiment that could follow the development of an epithelial cluster-isolating strategy is the study of the potency/proliferative potential of the individual epithelial subtypes (with particular attention to the secretory and intermediate clusters). To this end, isolated epithelial cells of one type could be plated to measure their organoid formation and differentiation capacity and/or tested in tissue reconstitution assays. On this line, it is to note that the only lineages that produced discrete clusters of cells undergoing the cell cycle in this project were the immune and mesenchymal and not the epithelial. This might be due to a lower proliferation rate that impaired the capturing of dividing cells or to the fact that, being the third most abundant cluster after the aforementioned, the number of cells was not enough to capture proliferating cells.

Moreover, the cell and molecular elements necessary for the orchestration of an immune response in the oviducts had a privileged place in this dataset. On

one hand, a mature repertoire of immune cells was described in the oviducts, unveiling functional and active innate and adaptive immune environments. On the other hand, a gene transcriptional programme governing the first line of innate immunity carried out by the oviduct epithelium has been described in this chapter. Regarding the identification of the guardians of the oviducts, a special focus was given to macrophages, given the unexpected number and heterogeneity found. Among the macrophages identities, bioinformatics analyses have highlighted the potential progeny candidates of the embryonic yolk-sac lineage. Further questions of interest in this field could be the correlation between macrophage identity and function in the oviducts, the underlying biology and the extent to what the developmental origin of macrophages could determine their functions. Furthermore, the epithelial secretory subtype and, more concretely, the Itga6_Dach2 cluster, have been highlighted as the cell identities with a putative main immune role, although ciliated cells also displayed immunogenic potential. Epithelial layers are known to be the first protective barrier against pathogens and insults and they play a paramount role in pathogens detection and the initiation and modulation of immune responses through cytokine release (Sinha et al. 2016). This has been demonstrated for epithelial populations in many organs, such as the lung and the trachea (Sinha et al. 2016), however, this is the first time that this role is thoroughly investigated and described in the oviducts and further mechanistic insights into the immune regulation exerted by the oviduct epithelium should be carried out in order to validate and expand on these data.

Taking everything together, this chapter has provided important insights into the cellular and molecular landscape of the oviducts, their functionality and the governing biology. However, there are still several outstanding questions in the field such as the intrinsic similarities and differences between the mouse oviduct and the human fallopian tube and the cellular interaction landscape governing the organ homeostasis, that will be the focus of the next chapter.

3 Chapter 3. Cross-species comparison of cell populations and interactions in the oviduct and fallopian tube

3.1 Introduction

Mouse models have successfully been utilised for decades to recapitulate and model human biological processes and diseases (Rydell-Törmänen and Johnson 2019). Although other animal models exist, mouse lines have extensively been used and are better understood due to a number of features. Mice are physiologically and genetically similar to human beings and spontaneously share a range of diseases that humans endure, whether they are of microbial or genetic origin. Mice are also small, with a high breeding rate, which is carried out in controlled conditions, allowing the generation of genetically identical strains (inbred strains). The availability of such strains enables higher experimental reproducibility and drastically reduces the inter-individual variability seen in humans. Mice also have a short lifespan, whose phases can be easily correlated to the different stages in human life. Furthermore, they can be genetically manipulated to generate a vast range of diseases models and, currently, there exists an extensive stock of different mouse lines, easily accessible to researchers. The use of mice allows much easier and flexible research in fields such as embriology, that would be strongly limited if performed only on humans. Nevertheless, while the advantages are huge, it is wise to bear in mind, that despite the many similarities, important differences exist and mice might respond very differently from humans when exposed to identical experimental situations. On this line, it has been reported that a big number of diverging responses are due to overlooking several parameters, such as the mouse line and strain, controls and statistics. Therefore, a best practice guideline in mouse experimental setups has been proposed in order to reduce these biases (Justice and Dhillon 2016).

With a special focus on the human fallopian tubes, several mouse models have been developed to mimic a number of diseases affecting these organs. For example, mouse models bearing HGSOC demonstrated that this tumour can originate in the oviduct secretory epithelium of one mouse line with *Brca*, *Tp53*, and *Pten* mutations (Perets et al. 2013) and another with *Pten* and *Dicer* mutations (Kim et al. 2012). Furthermore, the first model was able to describe serous tubal intraepithelial carcinoma lesions as the precursor of HGSOC, an analogous scenario to the human fallopian tube (Perets et al. 2013). It is important to note that, despite the extended use of the mouse as a model, these animals do not develop this disease spontaneously, while recent data have highlighted another species, laying hens, in which this could take place without genetical manipulation (Paris et al. 2021).

When it comes to comparisons between the mouse and human biology, scRNA-seq has proven a very useful tool to carry out this analysis from a high-throughput and unbiased perspective. To date, many single-cell transcriptomic studies have provided new insights into the mouse-human similarities, improving our understanding of what aspects of human biology can be well modelled in mouse. Importantly, inter-species differences and similarities have been found not only at the cellular level, but also at the genetic and regulatory levels for a number of organs and biological systems. Among these organs and systems are the enteric nervous system (Drokhlyansky et al. 2020), prostate (Crowley et al. 2020), pancreas (Baron

et al. 2016, Butler et al. 2018) and brain (Welch et al. 2019).

To date, no oviduct versus fallopian tube comparison has been carried out, among other reasons, because of the lack of data on murine oviducts. In this chapter, fallopian tube scRNA-seq data are analysed and, subsequently, a comparative study across the human and mouse species is carried out, taking advantage of the comprehensive murine cellular atlas that this project has generated. Furthermore, once the mouse and human landscapes have been well defined, the single-cell data analysis takes one step beyond and explores the cell-to-cell interactions established in both systems.

3.2 The physiological single-cell transcriptomic landscape of the human fallopian tubes

To carry out single-cell resolution studies in the human fallopian tube, a different approach was taken thanks to the availability of published datasets. In the human context, previous research projects had applied single-cell RNA sequencing to the fallopian tubes of normal individuals (Hu et al. 2020: epithelial cells, Dinh et al. 2021: comprehensive), so these data were downloaded, normalised, integrated and used in this chapter. The first published study was produced in our laboratory and, from its data, this thesis used 1318 good-quality epithelial cells, which had been processed with the SmartSeq2 protocol. From the second study, 38405 normal fallopian tube cells, sequenced with the 10X Chromium Single Cell 3' v2 and v3 protocols, were retained. This is a lower number than the one reported in the original publication due to higher astringency in defining what a good quality cell represents. Furthermore, the current project contributed with 929 fallopian tube cells obtained following the 10X Chromium Single Cell 3' v2 protocol to amount a total of 40645 integrated cells. The libraries produced in our laboratory were then considered as internal data (5.53% of the total data) and the ones obtained from the external publication, as external data (94.47% of the total data). Despite the big difference in input, cells from the in-house data contributed to all the lineage clusters identified in the final dataset (Figure 37, pg 114).

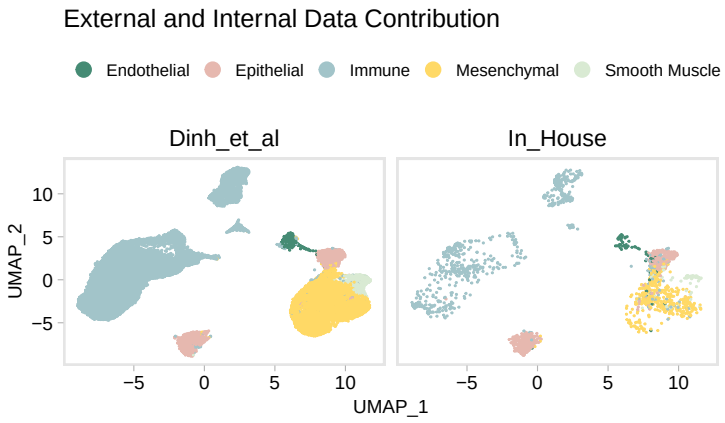


Figure 37. External and internal fallopian tube scRNA-seq data contribution UMAP plot of each dataset contribution to the fallopian tube lineages

Lineage annotation of the fallopian tube single-cell environment revealed a similar lineage distribution to the one observed in mouse. The immune, mesenchymal and epithelial were the fallopian tube dominant lineages, accounting for the 61.46%, 23.67% and 10.91%, respectively, versus the 41.09%, 30.58% and 19.97% detected in the mouse scRNA-seq data. Minority clusters such as endothelial and smooth muscle cells covered the 2.11% and 1.85% of the data, compared to the 1.1% and 3.77% found in mouse. The fact that mesothelial cells were not found in the human dataset suggests that they might account for a very low proportion in human and, interestingly, one of their mouse canonical markers, *UPK1B*, was found to be expressed by secretory and ciliated epithelial cells.

Subsequently, unsupervised clustering was carried out in the dataset and 24 final clusters were identified, although the last three were removed from downstream analysis due to the low number of cells recognised in them: 3, 2 and 2, respectively. Epithelial clusters in human, as in the mouse, were distinguished by the expression of *EPCAM* and broke down into two clusters corresponding to ciliated and secretory subpopulations, detected by the differential expression of *TUBB4* and *FOXJ1*, in the first case, and *PAX8* and *OVGP1*, in the second case. The vast majority of secretory cells showing positivity for *PAX8* were the cells originally sequenced with SmartSeq2, whereas in the ones that were sequenced with the 10X Chromium protocol this gene was not captured, resembling the situation found in the scRNA-seq oviduct dataset. The cells sequenced with SmartSeq2 displayed a higher number of detected genes per cell compared to those sequenced with 10X, which could have been the reason why *PAX8* was not a dropout in this subset. However, as it was highlighted in the previous chapter, it could also be a characteristic intrinsic to the sequencing method, as 10X snRNA-seq was able to capture *Pax8*, while presenting a library complexity similar to 10X scRNA-seq.

Expanding on further differences on the epithelial traits between mouse and human, it was found that while *Krt7* is expressed as a secretory marker in the human fallopian tube, it is ubiquitously expressed in both epithelial subtypes in mouse, with no significant distinction. Additionally, a putative immune role for oviduct epithelial cells, especially secretory cells, had been highlighted in the previous chapter. In human cells, that sharp "immunological" difference between secretory and ciliated was not found and there were no clusters of ciliated cells that showed negativity for the MHC Class I. However, those secretory cells that were sequenced with SmartSeq2 showed expression of genes coding for the MHC Class II, which would suggest a role in adaptive immunity by these cells that was not seen in mouse, where the secretory roles related to innate immunity. Furthermore, other clusters such as mesenchymal, smooth muscle and endothelial cells were detected through the expression of canonical markers conserved in mouse, such as *DCN* and *IGFBP5*, for mesenchymal cells, *PECAM1* and *EGFL7*, for endothelial cells and, *TAGLN* (Transgelin) and *ACTA2* (Actin Alpha 2), for smooth muscle cells.

A notorious difference between the mouse oviducts and the human fallopian tubes was the immune repertoire dominance of macrophages in the former and the dominance of T cells in the latter. Despite the fact that this dominance of innate immune cells in mouse and adaptive immune cells in humans could be a reflection of their specific biological conditions and environmental exposure (i.e. mouse lines are grown under controlled and sterile conditions, whereas humans are exposed to

standard pathogenic risks), it could also be a technical bias denoting a preferential capture of T cells in human due to their smaller size. This is backed by evidence of studies of fallopian tube immune populations that used alternative techniques, such as in Ardighieri et al. 2014. Using a histological approach, these authors found that $CD45^+$ leucocytes are frequent in the adult fallopian tubes and that most of them were $CD163^+$ macrophages, $CD11c^+$ dendritic cells and, also, $CD8^+$ T-cells. Although in the human scRNA-seq data, immune cells of the lymphoid, rather than the myeloid lineage, were the majority, $CD8^+$ T-cells (positive for *CD8A* and *CD8B*) were the most abundant among them, consistently with Ardighieri et al. 2014, and accounted for the 72.81% of lymphocytes.

The elevated number of T cells found in fallopian tubes allowed the identification of several subclusters that could not be identified in mouse, as T cells only comprised one homogeneous cluster. The first group of fallopian tube T cells was characterised by the expression of *CD4*, *IL7R* and complete negativity for the CD8 genes and *NKG7*, which are predominantly expressed in $CD8^+$ cytotoxic T cells and NK cells (Malarkannan 2020). Thus, these cells were identified as $CD4^+$ T helper cells. They show a subclassification into two groups, one positive for the expression of *CCR6* (C-C Motif Chemokine Receptor 6), *CD40LG* (CD40 Ligand) and *S100A4* (S100 Calcium Binding Protein A4), which could denote memory $CD4^+$ T cells, whereas the other presented positivity for the aforementioned genes and a higher expression of *CCR7*, which could be characteristic of naïve $CD4^+$ T cells. No clear effector and regulatory $CD4^+$ T cells were found in this dataset. Regarding $CD8^+$ T cells, no resting $CD8^+$ T cells were found in this dataset, as all the four identified subtypes express receptors for pro-inflammatory cytokines such as *CCR5*. Two of these clusters display a high similarity and an intermediate expression of *NKG7* and *GZMB* (Granzyme B), which mark effector cells. Another group presents a positive expression for the genes *CCL4* and *CCL4L2*, coding for cytokines that attract cells expressing the *CCR5* receptor, which in this dataset are mainly $CD8^+$ T cells too. The last $CD8^+$ T cell cluster differentially expressed *GZMK*, a pro-inflammatory granzyme. Analysis of the $GZMK^+$ $CD8^+$ T cell transcriptomic signature showed a striking resemblance to a very recently described $CD8^+$ T cells, termed as age-associated T cells (Mogilenko et al. 2021). This name originates from their similarity to terminal exhausted T cells, characteristic of chronic-inflammation mouse models. In Mogilenko et al. 2021, a twin cluster for these age-related identities in $CD8^+$ memory T cells was also identified and linked to an exhausted profile, however, $CD8^+$ memory cells were not found in this subset and, therefore, the $GZMK^+$ cluster could only correspond to the previous age-associated $CD8^+$ T cells (Galletti et al. 2020).

Regarding NK cells, three clusters could be distinguished by the high expression levels of *NKG7*. One was positive for *GNLY* (Granulysin), *KRT81* (Keratin 81), the transcription factor *ZNF683* (Zinc Finger Protein 683) and *ITGA1* (Integrin Subunit Alpha 1), where the last two have been previously identified as resident markers of tissue-resident NK cells in the human lung (Marquardt et al. 2019). Another NK cluster shared the markers *KRT81*, *AREG* (Amphiregulin) and *XCL2* (a chemokine for lymphocytes) with the first cluster, yet it was negative for *GNLY* and the tissue-resident markers, and was highly positive for *GZMK*. The last group differentially expressed *NKG7*, *GNLY* and *GZMB*, which could denote an enhanced

cytotoxic capacity. Interestingly, a cluster of T cells expressed typical $CD8^+$ T cell markers, such as CD3 and CD8 genes, but also, NK-related genes, such as *NKG7*, *KLRB1* and *KLRD1* (Killer Cell Lectin Like Receptors B1 and D1), which set this cluster as NKT cells. Furthermore, to complete the immune repertoire of the fallopian tube, a cluster of mast cells, which was not seen in the mouse, is identified by the expression of *TPSAB1* and *TPSB2* (Tryptases Alpha/Beta 1 and 2). Mast cells are mainly tissue-resident cells, although they can be found in the bone marrow at a very low proportion (0.008%) (Teodosio et al. 2015). They are involved in a variety of immunological processes, such as inflammation and allergic reactions. Finally, macrophages were also identified by the exclusive expression of *LYZ* (Lysozyme) and the expression of *FCER1G* (Fc Fragment of IgE Receptor Ig) and *TYROBP* (Transmembrane Immune Signaling Adaptor), which are also found in NK cells.

3.3 Inter-species comparisons of the oviduct and fallopian tube cell types

Once the single-cell landscape of the human fallopian tubes had been initially explored and characterised, this project sought to further understand the differences and similarities across the oviduct and the fallopian tube by establishing a comparative cellular and biological correlation analysis. To this end, the average of the normalised expression of individual genes within the different cell clusters was calculated in the mouse and human datasets to yield individual cluster-specific transcriptomic signatures. Next, human genes were converted into their mouse one-to-one orthologs and only genes present in both datasets were kept, amounting for a total of 14379 genes.

Then a correlation analysis method was chosen. Among these, there are parametric and non-parametric tests, such as Pearson, for the first instance, and Kendall tau and Spearman rho, for the second. A parametric test is based on the hypothesis that the tested variables follow a linear dependence and a normal distribution, which is not usually the case for RNA-Seq gene expression, normally following a poisson (Marioni et al. 2008, Wang et al. 2010) or a negative binomial distribution (Anders and Huber 2010). To test the suitability of this method for this project's data, visualisation of pairwise comparisons between different clusters using the Pearson's method was first executed. They showed that regardless of the similarity among clusters, a linear covariate was followed among the cell identities between and within species (Figure 38, pg 118).

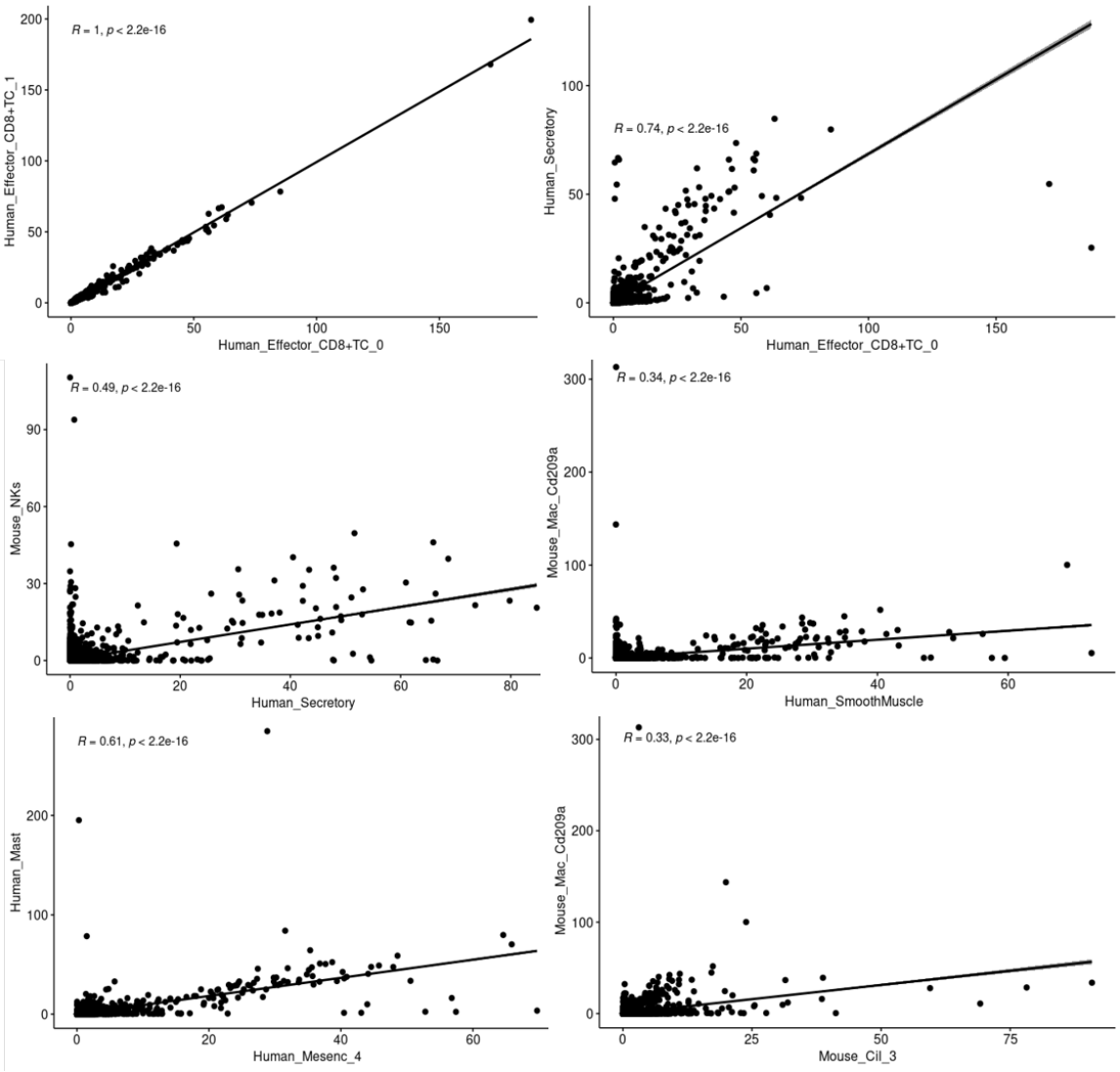


Figure 38. Pearson linear correlation tests between the transcriptomes of individual cell clusters

The linear dependency among the gene expressions of different inter and intra species and inter and intra lineage clusters was measured applying the Pearson's correlation method. Gene expression values are calculated as the average of the normalised expression of individual genes within a specific cell cluster. The normalised expression of a gene expressed in a cell is calculated by dividing that gene counts by the total counts for that cell and multiplying by 10,000.

Subsequently, a Shapiro-Wilk normality test was run with the data from each variable (cluster) to test whether or not they follow a normal distribution. The null hypothesis of this test assumes that the data are normally distributed, and the results for the scRNA-seq clusters showed that a clear trend was not followed. Some clusters failed to reject the null hypothesis (when p-value is higher than 0.05, meaning that the distribution of the data is not significantly different from a normal distribution) and some rejected it (when p-value is lower than 0.05, meaning that the distribution of the data is significantly different from a normal distribution). To further confirm whether the Pearson's test should be discarded, a second visual method to test normality, Q-Q plots (quantile-quantile plots), was applied. These

graphs represent the correlation between a sample and the normal distribution, and to assume that the data follow a normal distribution the dots should fit the line appropriately, a case, which was not seen for any of the clusters.

Taking everything together, the Pearson's method was rejected and the non-parametric Spearman test was chosen as the correlation strategy. When plotting correlations between the different clusters, it became apparent that the expression of a high percentage of the 14379 genes could be considered as background expression. Therefore, to focus the comprehensive correlation analysis between the species only on variable genes, the 2000 most highly variable genes across mouse cells and across human cells were identified, with the human genes converted into their mouse one-to-one orthologs. Among these genes, those common to both species were selected (271 genes) and only the expression of the first 50 most highly variable genes across all the clusters in both datasets were kept. Then, a Spearman correlation test was applied to these data and a correlation matrix for all the human and murine clusters was generated (Figure 39, pg 120).

Despite the adopted filtering steps, a species-specific signature can still be detected given that some cell types show a high correlation to different cell types within the same species. This phenomenon would require further investigation given the possibility of being produced by either a biological or a technical effect. However, even in these cases, a correlation gradient is found. For example, fallopian tube mesenchymal and smooth muscle cells show a very high correlation among themselves, medium-high correlation with human endothelial and epithelial cells and lower correlation with fallopian tube immune cells. The human cell type that showed the lowest correlation values against the same-species clusters were the two macrophages groups. Furthermore, the correlation matrix also allows for the identification of those clusters showing the best correlation values across species. On one hand, oviduct epithelial cells showed homogeneous and similar correlations with smooth muscle, mesenchymal, endothelial and epithelial cells of the fallopian tubes. However, mouse endothelial cells showed good correlation with all the aforementioned fallopian tube cells, except for the epithelial type. All the mouse macrophages types highly correlated to the fallopian tube macrophages without distinction between the two groups. The mouse macrophages types that showed the highest correlations belonged to the C1qb type and the Cd209a type. The highest-correlated cell types across the two species were the human and mouse mesenchymal cells and the human and mouse macrophages.

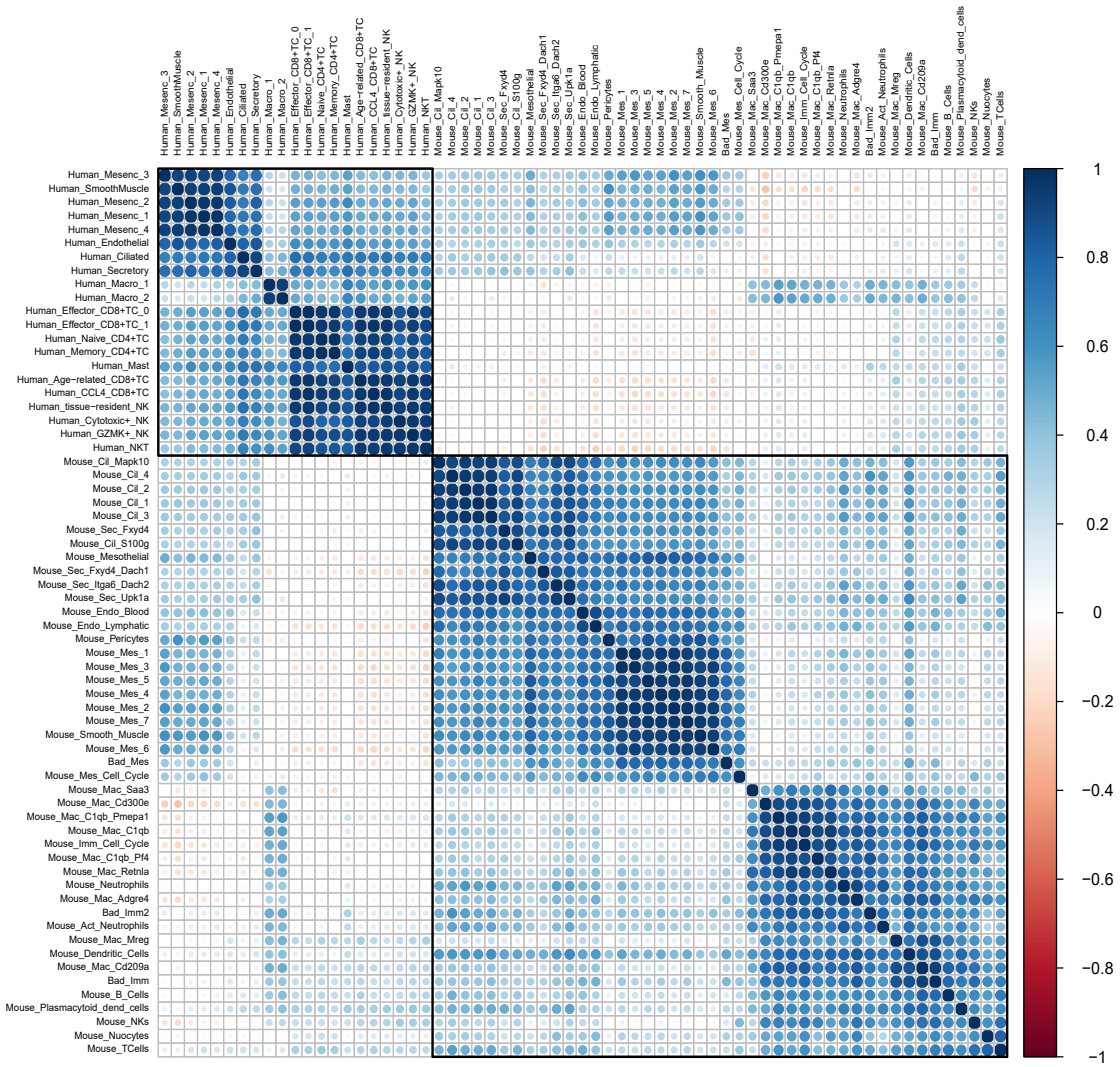


Figure 39. Human-mouse cluster Spearman's correlation matrix based on the expression of the 50 most highly-variable genes
All individual cell clusters in the single-cell mouse dataset and the single-human dataset are compared and assigned a correlation coefficient, which colours by the legend. Positive correlations present a blue colour, no correlation presents a white colour and negative correlation presents a red colour. The circle sizes show the absolute value of corresponding correlation coefficients.

3.4 Characterisation of the complex cell-to-cell interactions network in the oviduct and fallopian tube

To gain further insights into these organs, this project aimed to identify and characterise the main cell-cell crosstalk coordinating different biological aspects in the human fallopian tube and the oviduct environment. To this end, the algorithm CellPhoneDB was applied to both datasets. CellPhoneDB is based on a manually-acquired database of ligand-receptor interactions, which takes into account the sub-

units of receptor complexes and detects enriched pairwise interactions (Efremova et al. 2020).

In mouse, the results showed an interactive landscape that was unquestionably dominated by mesenchymal cells, which established communicative connections with practically all members of the same and other lineages, followed by macrophages, which also displayed broad and numerous communications (Figure 40, pg 122). The highest number of mesenchymal interactions were established with other mesenchymal cells, yet communication was also large with members of the immune group, predominantly macrophages, dendritic cells and neutrophils, as well as, endothelial, pericytes and the most interactive epithelial clusters. The majority of interactions were established by a bad-quality mesenchymal cluster with no special features or DE genes and by the cluster Mes_3, differentially expressing the markers *Clec3b* (tetranectin) and *Cpz*, a carboxypeptidase, which has been related to modulation of the Wnt pathway.

Surprisingly, T cells and the majority of epithelial cells accounted for the clusters with less absolute interactions, both within the same and external lineages. As expected, the few T cells interactions were primarily established with macrophages, representing the antigen-presenting processes necessary for the activation of the adaptive immunity. However, the second most interactive group for T cells were all the mesenchymal clusters, except Mes_4. In the case of the epithelium, the clusters Cil_S100g, Sec_Fxyd4 and Sec_Upk1a were more inert than T cells, with only a minimum number of cross talks found. In contraposition, the clusters Sec_Itga6_Dach2, Fxyd4_Dach1 and Cil_Mapk10 were termed as the epithelial "ambassadors" for the epithelial lineage, as they established an intense communication network with almost all members of the same and different lineages, especially with mesenchymal cells, the activated C1q macrophages and Mac_Saa3. The findings regarding the strong communication of these epithelial cells with those specific macrophage identities was very interesting, as they are the subtypes that accounted for the best scores for the tissue-resident macrophage signature applied in chapter 2. Furthermore, all ciliated clusters, except Cil_2, exhibited a rich and almost exclusive communication network with mesenchymal cells and a very poor one with themselves and secretory cells.

Taking all the data together, several scenarios in the oviducts system can be suggested: (1) as it has been highlighted throughout this project, the T cell function and, thus, the adaptive immunity in mouse is not very active, in contraposition to the high communication network exhibited by macrophages; (2) mesenchymal cells might be a key regulatory component for the oviduct homeostasis, including immune processes; (3) mesenchymal and ciliated cells might be involved in mono-directional or reciprocal regulation; (4) Sec_Itga6_Dach2, Fxyd4_Dach1 and Cil_Mapk10 are putative important clusters for the epithelial or other functions; (5) unexpectedly, secretory and ciliated cells do not display cross-talk.

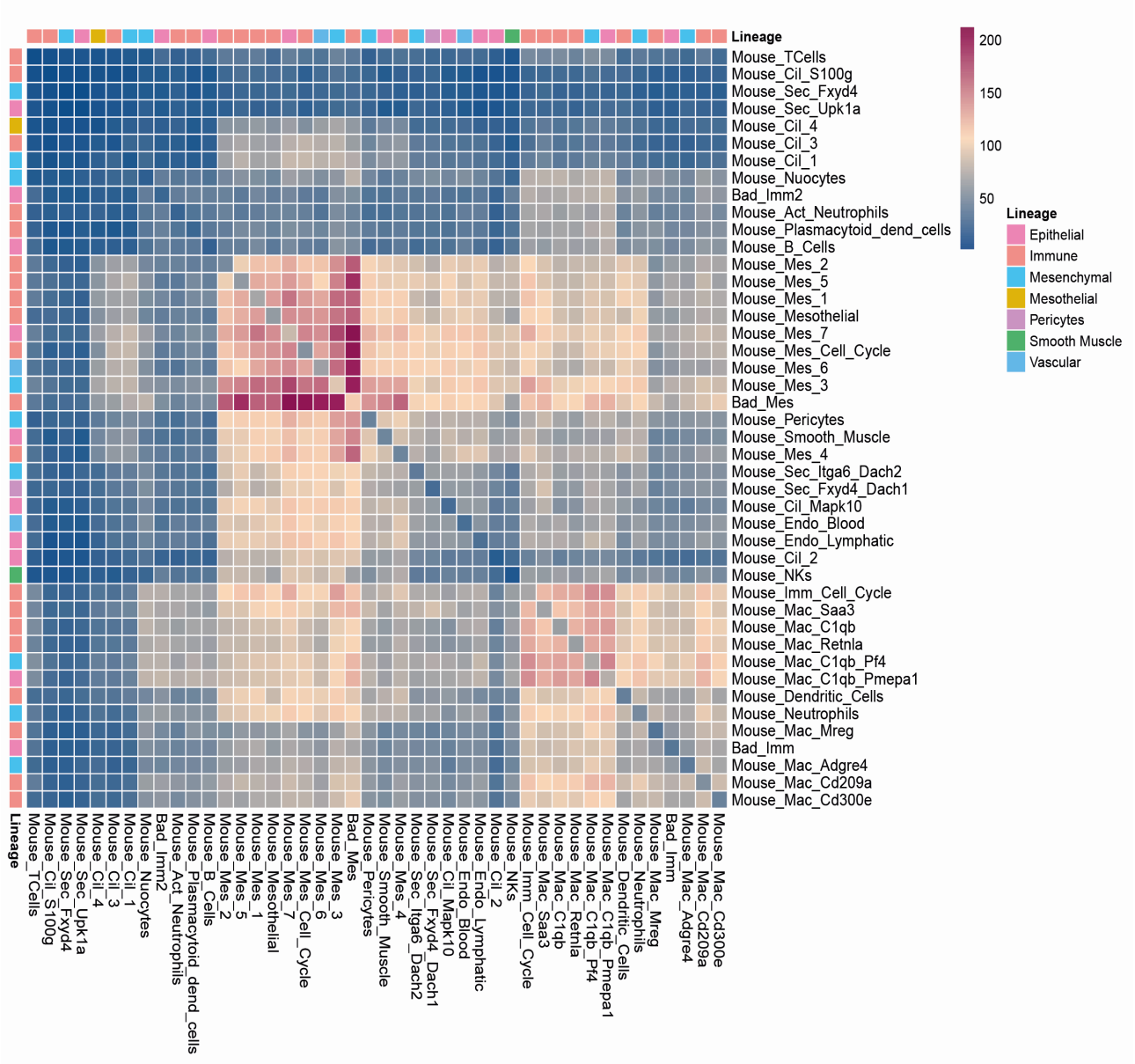


Figure 40. Heatmap showing the total number of interactions between oviduct cell types and their associated lineages

In the human ligand-receptor interaction landscape, the most interactive identities are the two macrophage clusters and the mesenchymal cluster Mesenc_3, closely followed by the other mesenchymal clusters, the immune NK clusters and some of the CD8$^+$ T cells (Figure 41, pg 123). This scenario was surprising given that T cells and NK cells account for the predominant identity found in the fallopian tube scRNA-seq data and the innate immunity-related macrophages are a minority. Another interesting finding related to Mesenc_3, a cluster marked by the exclusive expression of *ADH1B*, an alcohol dehydrogenase, *C7*, a member of the complement pathway of the innate immune system, *ANGPTL1* (Angiopoietin Like 1), involved in the formation of blood vessels and, also, the differential expression of *PDGFRL* (Platelet Derived Growth Factor Receptor Like) and *FBLN5* (Fibulin 5), a secreted extracellular matrix protein promoting adhesion to endothelial cells. These DE genes suggest a function related to endothelial cells for this cluster and, ef-

fectively, Mesenc_3 accounted for the cell type with which endothelial cells establish the highest number of interactions within the fallopian tube.

As in the mouse situation, the majority of interactions were established by immune and mesenchymal cells, while the epithelial type only presented a moderate number of them. Within the epithelial types, the secretory one established the highest number of connections, mainly with the two macrophages, Mesenc_3, the age-related $CD8^+$ T cells, the other mesenchymal clusters, the cytotoxic and the tissue-resident NK cells and NKT cells. These last three identities, along with Effector_$CD8^+$ TC_1, CCL4_$CD8^+$ T cells and the $GZMK^+$ NK cells, were the most interactive clusters within the immune repertoire, in contraposition to the more inert Naïve_$CD4^+$ TC, Effector_$CD8^+$ TC_0, age-related $CD8^+$ T cells and Memory_$CD4^+$ TC. Furthermore, an additional shared similarity that was detected after this analysis, is that macrophages account for the most interactive immune cell type in both mouse and human.

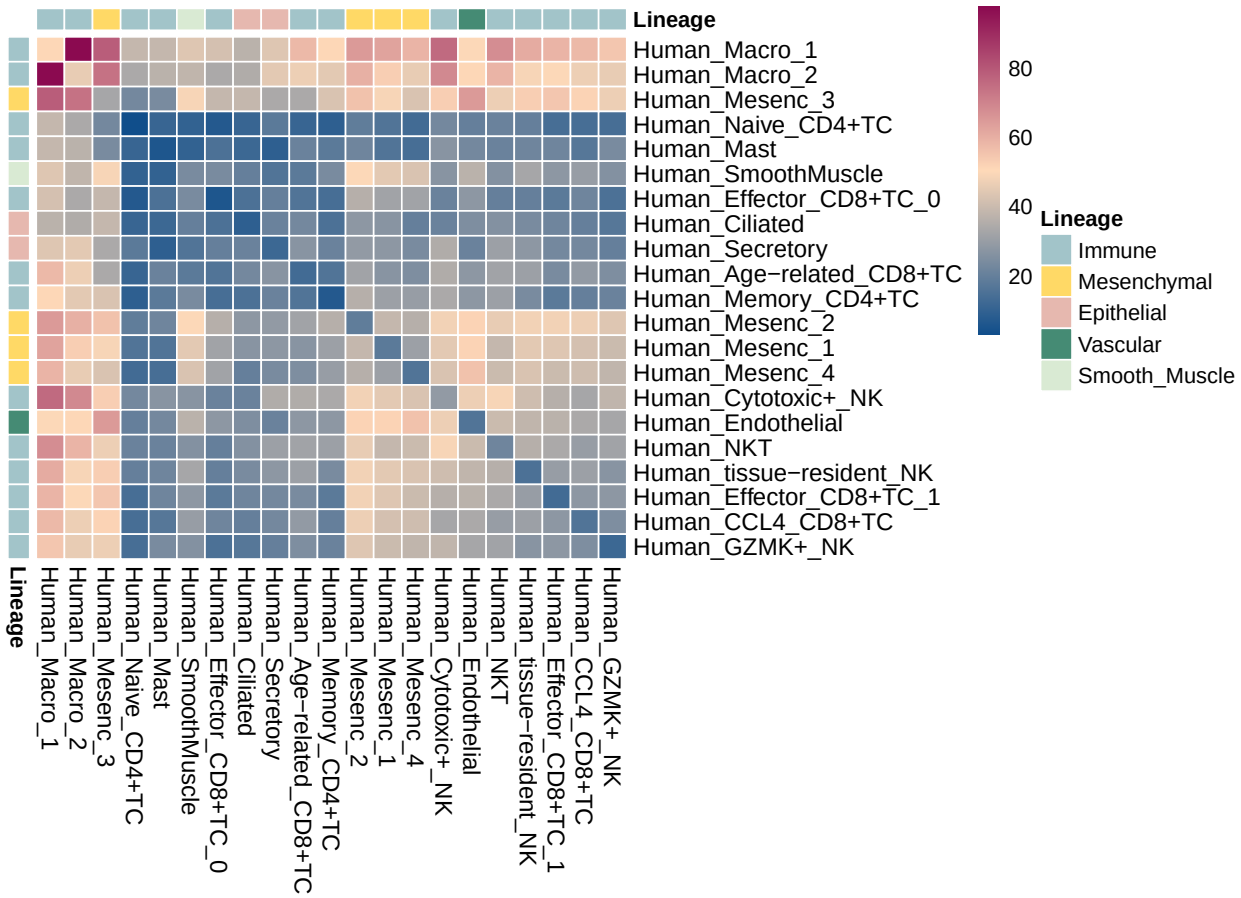


Figure 41. Heatmap showing the total number of interactions between fallopian tube cell types and their associated lineages

3.5 Conclusion

The main significance of this study relies on the integration and characterisation of a comprehensive fallopian tube single-cell atlas, capturing the broad heterogeneity and remarkably improving our understanding of the organ. Furthermore, this analysis allowed the direct comparison between the different transcriptomic identities, states and cross-talk networks that define the fallopian tube and the oviduct. For this, the individual clusters among the human and the mouse were manually compared, revealing a highly-conserved gene expression of a few canonical markers among equivalent cell clusters in the fallopian tube and the oviduct. However, this comparison also unravelled core differences regarding the immune landscape, where the human one showed enrichment for cells in charge of orchestrating adaptive immune responses (T cells), in contraposition to the innate murine signature (macrophages). When the role and interactions of these cell types were analysed by population-level interaction studies run with CellPhoneDB, macrophages proved to be the most interactive immune cell type in the oviduct, as expected. Nevertheless, and against all odds, they also accounted, not only for the most interactive immune cell identity in the fallopian tube, but for the most interactive cell type of all the fallopian tube populations.

Moreover, an in-depth analysis relying on an averaged expression profile approach of highly variable genes across the two species (50 genes) revealed further information about the most correlated cell types across the two species. This analysis also revealed certain species-specific background gene expression, which might be due to an actual biological reality or a technical bias of the studies, which were independently run on each species. Even if this was a biological effect, the contribution of these species-specific genes to the variance and phenotype should be measured and analysed. This is because there could be the possibility that their contribution to the cell type phenotype could still be smaller than that of the few conserved markers across cell types, which could be important phenotypical drivers. Furthermore, a good strategy to quantify the degree of technical influence on the obtained results would be the design of scRNA-seq experiments with dual sequencing of cells coming from the two species in the same sample, which would be possible thanks to cell hashing techniques. This technique is based on the differential tagging of cell groups of interest enabling multiplexing of diverse and heterogeneous samples (including those originating from different species) as well as reliable doublet identification in scRNA-seq experiments. The most popular methods are cell conjugation with a barcoded antibody or lipid and direct tagging of the cell protein with an oligonucleotide sequence (Stoeckius et al. 2018, McGinnis et al. 2019b, Gehring et al. 2020). Furthermore, and despite the low correlation found, this comparative analysis highlighted the existence of a more prominent mesenchymal-like profile in human secretory epithelial cells than in mouse secretory cells. Additionally, the mouse and the human communication landscape was also presented in this chapter, enabling the identification of cell-to-cell relationships that could be key for these organs' homeostasis.

Finally, in the last part of this project, I look at how the new insights into the biology of these organs can be used to better understand neoplastic conditions affecting the fallopian tube. More specifically, further bioinformatic analyses, based

on the scRNA-seq foundation and results generated in the previous chapters, are applied to HGSOC transcriptomic data in the following chapter.

4 Chapter 4. scRNA-seq, microarray and bulk RNA-Seq analyses discover a TGFβ signalling signature conserved in oviducts, fallopian tubes and high-grade serous ovarian cancer

4.1 Introduction

In the last decade, the revolutionary progress of bulk high-throughput sequencing techniques (and, more recently, their single-cell counterparts) have generated a big amount of cancer genomic and transcriptomic data. This superlative data yield is now stored in big cancer databases, which can be interrogated and explored to unravel complex aspects of cancer biology with, potentially, enormous benefits in terms of disease understanding and, hopefully, clinical management. Probably, the most well-known among these repositories is The Cancer Genome Atlas (TCGA) database, a large-scale project that, hitherto, has collected comprehensive data of individual tumours at the DNA, RNA, protein, epigenetic and clinical levels. Naturally, this rich repository has been critical and continues fuelling major discoveries in the cancer biology field, whose hallmark is its intrinsic complexity and heterogeneity (Cancer Genome Atlas Research Network et al. 2013, Tomczak et al. 2015, Hoadley et al. 2014).

The development of this strong data core, along with the advent of single-cell sequencing techniques, has proven as an opportunity to adopt new strategies. For example, scRNA-seq has prompted the genesis of bioinformatics tools to ascertain the cell type composition within a bulk sample, technically termed as deconvolution or decomposition (Wang et al. 2019, Baron et al. 2016). This approach is very useful, taking into account that, unlike bulk RNA-Seq, scRNA-seq remains highly expensive, making its use unfeasible to systematically interrogate a large number of samples. Such technical combinations offer unique and new insights into functional genomics studies, improving our understanding of diseases that have historically been challenging to study and model, such as ovarian cancer. This tumour type is a combination of heterogeneous diseases with molecular and genomic features such as *TP53* mutations, DNA copy number alterations and aberrant methylation events (Cancer Genome Atlas Research Network 2011). At the clinical level, ovarian cancer is often characterised by an advanced stage at diagnosis, given that the early symptoms of this disease are highly unspecific. This situation has motivated large scientific efforts to decipher the cell of origin and the conditions promoting this tumour's development, in order to develop early-screening strategies.

Formerly, the big steps in the investigation and the progress in this field had largely relied on strategies targeting mRNA transcripts from millions of pooled cells such as gene-hybridization microarray platforms and next-generation bulk RNA-Seq. Thanks to microarray studies and analyses of thousands of HGSOC transcriptomes, its first molecular classification was generated in 2008 (Tothill et al. 2008) and validated in 2011 (Cancer Genome Atlas Research Network 2011). However, ovar-

ian cancer is one of the pathologies that has currently benefited from the integrative bulk and single-cell approaches, which have enabled breakthroughs such as the more advanced and updated classification of HGSOC patients and the novel identification of a poor-prognosis cohort with upregulated expression of EMT genes (Hu et al. 2020). Thus, elegantly demonstrating that the old bulk RNA-Seq and microarray techniques have not become obsolete, still hold important value and carry new exciting research opportunities, thanks to these novel approaches.

In this final chapter, data mining using the generated scRNA-seq data and the population-level interaction analysis identified several genetic identities and pathways highly enriched in normal fallopian tubes and oviducts. Of these, the TGFβ signalling is additionally explored in HGSOC transcriptomic data, revealing two TGFβ ligands differentially expressed in tumours presenting a high EMT signature. Next, a retrospective analysis of the scRNA-seq datasets detected their putative secreting sources and a positive correlation between these cell types and the TGFβ ligands, as well as with the EMT signature in HGSOC.

4.2 TGFβ interactions are significantly enriched in oviduct and fallopian tube homeostasis

The aim of this part of the project was to expand beyond the quantitative analysis of the interaction landscape in the oviduct and fallopian tube, to shed light on the signalling pathways representing key hallmarks. Furthermore, it was reasoned that if one of those molecular pathways was found to be conserved across species and pathological states, it would highlight its importance and offer a novel candidate for future mechanistic studies.

To this end, further systematic analysis searching for statistically-significant molecular interactions across the oviduct and the fallopian tube was executed using CellPhoneDB. This revealed an enrichment for interactions with members of the NOTCH, WNT, BMP and TGFβ signalling pathways. The functions of the first three pathways have previously been studied within the fallopian tube and oviduct context. The NOTCH and WNT pathways have been described as keepers of the growth and differentiation processes in a fallopian tube organoid model (Kessler et al. 2015). The same study and others have proposed a specific role for NOTCH in stemness, proliferation and maintenance of the secretory phenotype by blocking its differentiation into ciliated epithelial cells (Zhu et al. 2019, Kessler et al. 2015). With regard to the BMP and WNT signalling pathways, they have been shown to exhibit paramount roles during the embryonic development of the Müllerian ducts, as extensively reviewed in the introduction section. Moreover, Chapter 2 of this thesis presented the existence of oviduct mesenchymal identities which seem to negatively regulate the WNT and the BMP pathways through the expression of *Notum* and *Dkk2* (Wnt inhibitors) by Mes_5 and *Bmper* (BMP inhibitor) by Mes_1 and Mes_3. Additionally, the expression of many of those pathways' ligands and receptors showing statistically-significant interactions in the CellPhoneDB analysis was detected in the mouse and the human scRNA-seq datasets.

Despite the amount of information about the roles of the previous pathways in the oviduct/fallopian tube biology, TGFβ signalling has been less explored in

this context. This pathway works through hetero-tetrameric serine-threonine kinase receptor complexes composed by a combination of two type I (TGFβR1, signal propagators) and two type II (TGFβR2, activators) receptors, although TGFβR3 can also be found retaining ligands. The type II receptors phosphorylate type Is, which then propagate the signal through phosphorylation cascades. The ligands binding to these receptors can be from the TGFβ and the BMP subfamilies. The overview of the TGFβ interactions among discrete clusters of the mouse oviducts showed a large representation of this molecular pathway (Figure 42, pg 129). The most represented receptors were *Tgfβr1* and *Tgfβr3*, with *Tgfβr2* in last place. The first two receptors were mainly bound by *Tgfβ1* ligand, produced by macrophages, neutrophils and nuocytes in the case of *Tgfβr3*, and endothelial, dendritic cells, B cells and activated neutrophiles in the case of *Tgfβr1*. Furthermore, *Tgfβr3* was used by several mesenchymal identities to bind to the *Tgfβr1* of other lineages.

It seems that a major communication axis involving immune cells in the oviduct is mediated by the secretion of *Tgfβ1*, except for the macrophage cluster Mac_Saa3, whose main secreted ligand is *Tgfβ2*. Interestingly, Mac_Saa3 is the macrophage cluster accounting for the highest number of TGFβ interactions with other clusters. It is able to establish connections with all lineage members primarily through the expression and putative secretion of *Tgfβ2* which binds to *Tgfβr1*, *Tgfβr3* and, in less instances, to *Tgfβr2*. Notably, it is the only cluster showing statistically-significant TGFβ interactions with the secretory epithelial clusters Sec_Itga6_Dach2 and Sec_Fxyd4_Dach1. Although the vast majority of immune cells use *Tgfβ1* as a ligand, the main binding interactions are to different receptors depending on the receiving cell type. In this way, the *Tgfβ1* ligand produced by macrophages, neutrophils and nuocytes binds to *Tgfβr3* expressed in endothelial cells, mesenchymal, mesothelial cells, Cil_Mapk10, Mac_Adgre4 and Mac_Saa3. Furthermore, mesothelial cells produce the growth factor *Fgf1*, which seems to establish a connection with mesenchymal, macrophages, Cil_Mapk10 and ciliated cells expressing *Tgfβr3*. Other cell types, such as the macrophage clusters Mac_C1qb_Pmepa1, Mac_Adgre4 and Mac_Cd300e, use the receptor *Tgfβr1* to receive *Tgfβ1* from: (1) all the immune cells, mes_5, mes_7 and endothelial cells in the Mac_C1qb_Pmepa1 case; (2) from the majority of macrophages, nuocytes and neutrophils in the Mac_Adgre4 case; (3) from neutrophils in the Mac_Cd300e case. The macrophages Mac_C1qb_Pmepa1 and Mac_Adgre4 seem interesting identities that would deserve further study given the amount of potential TGFβ regulation that they receive. In the specific case of Mac_C1qb_Pmepa1, they receive it from all immune cells, some mesenchymal and all endothelial cells, while Mac_Adgre4 receives it from a large number of immune cells.

Regarding the oviduct epithelium, the Cil_Mapk10 demonstrated to be an interesting cluster for its ability to receive TGFβ ligands from a large number of clusters, which potentially could suggest that it is the epithelial cluster that most likely could activate the TGFβ signalling. It is followed by ciliated, Sec_Itga6_Dach2 and Sec_Fxyd4_Dach1 clusters, which also displayed TGFβ reception. These secretory identities were not only capable of receiving external TGFβ input, but also were the only epithelial cells able to express the predominantly mesenchymal-produced *Tgfβ3* ligand to establish connections with Mac_C1qb_Pmepa1.

In the fallopian tube, the TGFβ interactions were only dominated by the

TGFβ1 ligand, which was expressed by T cells, NK cells and the macrophage cluster Macro_1 (Figure 43, pg 130). This ligand was received by all the other cell lineages, except epithelial cells, mainly through the *TGFβR3* receptor and the *EGFR* receptor (only in mesenchymal cells).



Figure 42. Overview of statistically-significant TGFβ ligand-receptor interactions using CellPhoneDB on the oviduct scRNA-seq dataset

The p-value is indicated by the circle area and the means of the average expression level of interacting molecule 1 in cluster 1 and interacting molecule 2 in cluster 2 are indicated by the legend colour gradient.

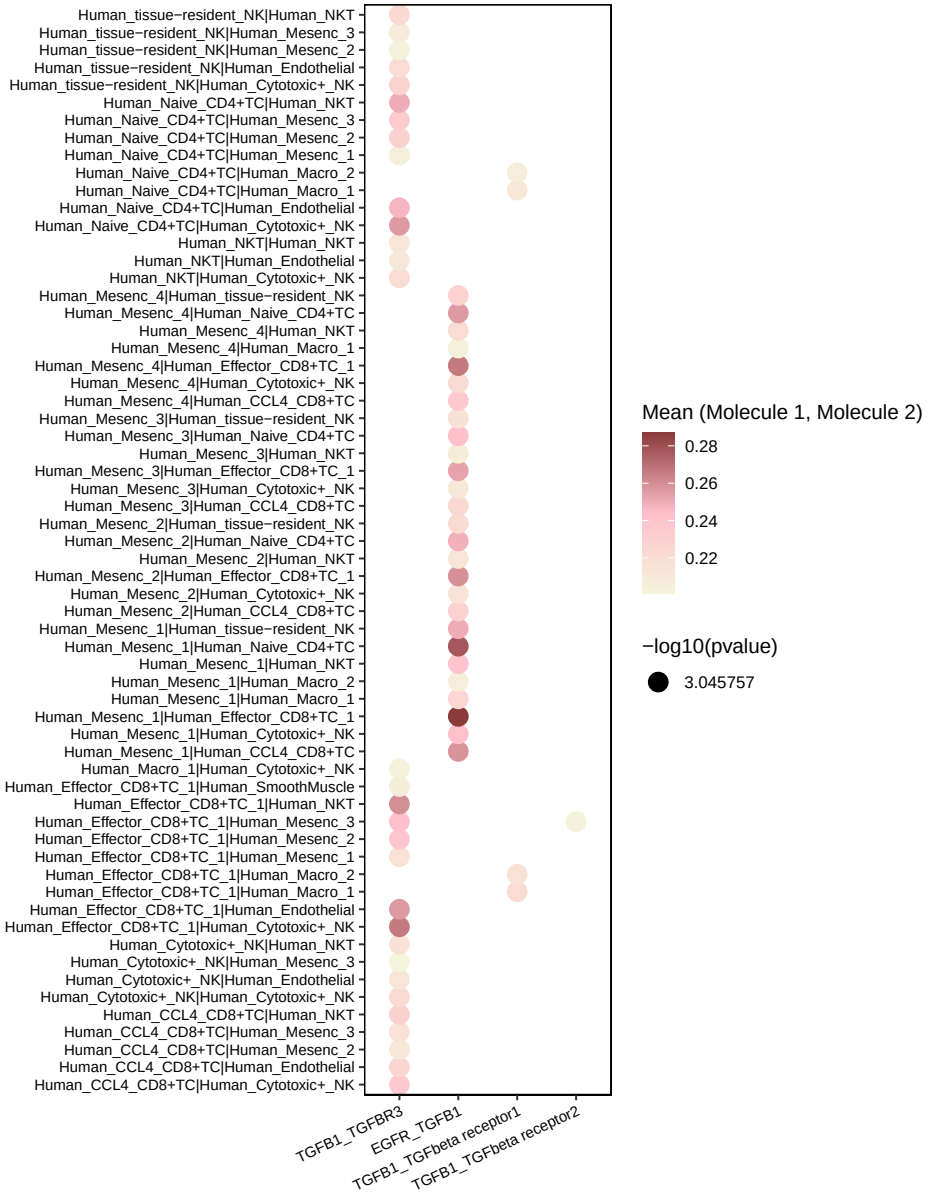


Figure 43. Overview of statistically-significant TGFβ ligand-receptor interactions using CellPhoneDB on the fallopian tube scRNA-seq dataset
The p-value is indicated by the circle area and the means of the average expression level of interacting molecule 1 in cluster 1 and interacting molecule 2 in cluster 2 are indicated by the legend colour gradient.

Taking everything together, it seems that the TGFβ pathway could be an important pathway underlying the oviduct and the fallopian tube biology. TGFβ signalling has been reported to be multifunctional, controlling diverse and also opposite processes, such as proliferation and growth inhibition, pluripotency and differentiation, pre-malignancy suppression and metastasis (Massagué 2012). Given the ample range of functions that have been described for this pathway, it results challenging, based only on gene expression data and without functional validation, to ascertain which role/s this pathway could be playing in the fallopian tube and the oviduct. An extra layer of complexity is also added by the nature of the TGFβ-triggered responses, as they are known to be affected by the environment and different ef-

fects have been reported depending on cell types and states. Among these cell type-specific roles, EMT has extensively been shown to be controlled by TGFβ signalling, through the Tgfβ1, Tgfβ2 and Tgfβ3 ligands, in epithelial cells. Addition of any of these factors to cultured epithelial cells induces an EMT response, which changes their morphology into mesenchymal-like and enhances their migratory capacity, among other effects. However, to the best of my knowledge, only Tgfβ1 has been experimentally demonstrated to trigger EMT in cultured oviduct epithelial cells (Alwosaibai et al. 2017). Based on the data-driven evidence that the TGFβ pathway could be important for the oviduct and fallopian tube biology and its widely proven role in EMT across a number of organs and systems, I aimed to answer the question of whether or not this pathway could potentially be conserved and play a role in HGSOC types marked by an EMT signature.

4.3 TGFβ1 and TGFβ3 are differentially expressed in EMT-high HGSOCs and secreted by immune and mesenchymal cells, respectively, in the normal fallopian tube and oviduct

To carry out this analysis, the transcriptomic profiles of around 1,700 HGSOC tumours were obtained from both TCGA bulk RNA-Seq data (Cancer Genome Atlas Research Network 2011) and microarray data from the Australian Ovarian Cancer Study (AOCS) (Tothill et al. 2008). Subsequently, the tumours were classified according to an EMT signature score developed by Hu et al. 2020, which significantly associated with poor survival in HGSOC. In order to test whether any of the TGFβ ligands was differentially expressed in EMT-high cases, the average expression of these ligands in each tumour was plotted (Figure 44, pg 131), which revealed a high differential expression of the ligands *TGFβ1* and *TGFβ3* in tumours with a high EMT profile.

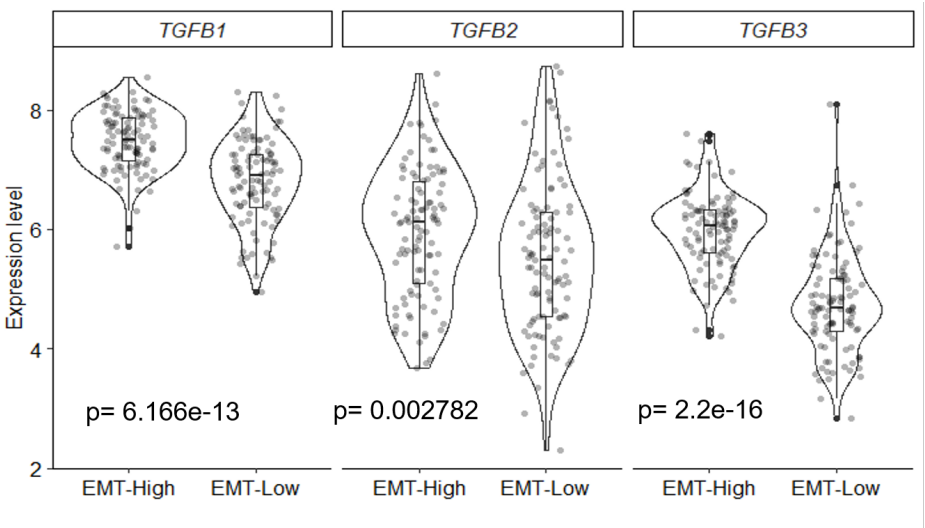


Figure 44. Differential expression of TGFβ ligands in EMT-high versus EMT-low ovarian tumours
Violin plots show expression of *TGFβ1*, *TGFβ2* and *TGFβ3* in low-EMT compared to high-EMT HGSOC tumours from the TCGA dataset. The box and whiskers represent the median, two hinges (first and third quartiles) and two whiskers. The represented p-values were calculated with the non-parametric Wilcoxon test.

Based on these results, the next formulated questions were: What are the cell identities expressing these ligands under the normal organ homeostasis? Could these TGFβ-secreting cell types also be differentially expressed in EMT-high HGSOCs? Although the previous analysis that focused on cell-to-cell interactions of the TGFβ pathway suggested that *TGFβ1* could be primarily expressed and secreted by immune cells and *TGFβ3* by mesenchymal cells, interrogation of the scRNA-seq datasets gave the final answer (Figure 45, pg 132).

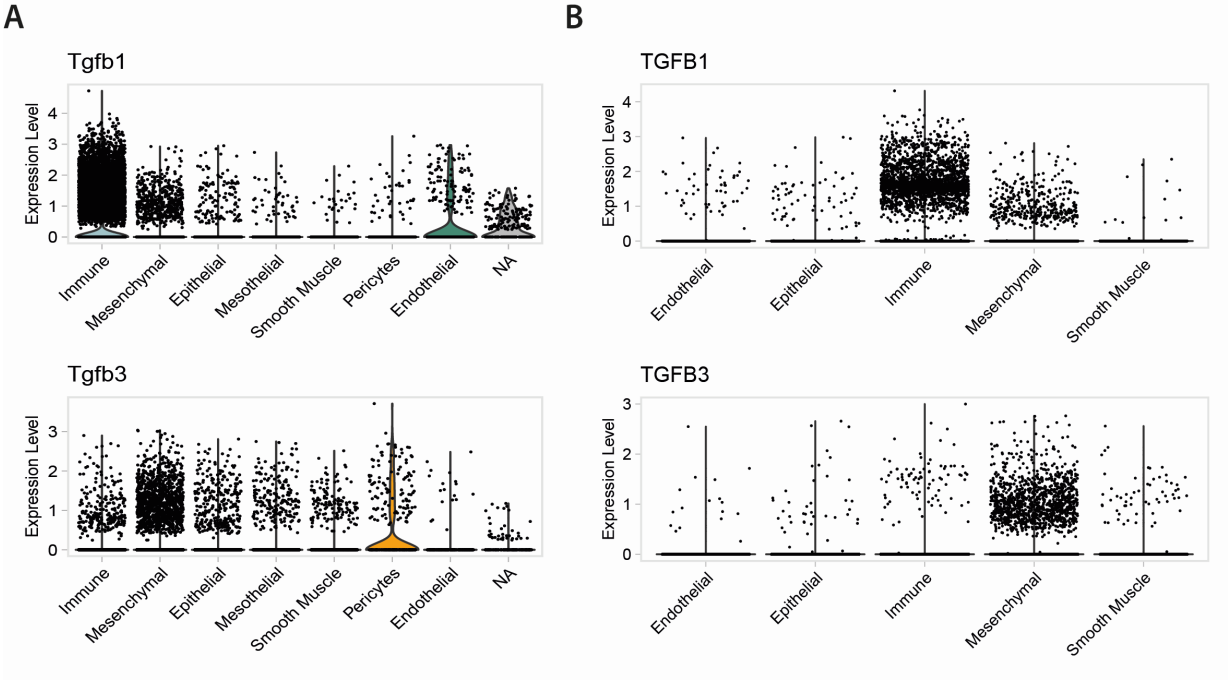


Figure 45. *TGFβ1* and *TGFβ3* expression in the oviduct and the fallopian tube
A) Violin plots show the expression of *Tgfβ1* and *Tgfβ3* in each lineage of the mouse oviduct. B) Violin plots show the expression of the same TGFβ ligands in all the lineages of the fallopian tubes. Gene expressions are calculated by dividing the gene counts for each cell by the total counts for that cell, multiplying by 10.000 and applying a log1p-transformation.

Confirming the theory extracted from the interaction analyses, gene expression analysis showed that *TGFβ1* is primarily produced by immune cells, and less by mesenchymal and endothelial cells, and that *TGFβ3* is primarily produced by mesenchymal cells. This situation was conserved in mouse and human. Within immune cells, the cell types that expressed the highest level of *Tgfβ1* in mouse were macrophages and, in human, T and NK cells. Therefore, the next steps were to identify within the TCGA and AOCS datasets: (1) whether markers for these cell subtypes show a positive linear correlation with their specific TGFβ ligand and (2) whether markers for those cell subtypes show differential expression in EMT-high tumours. If the expression of the molecular markers of these cell types show a good correlation with the expression of their specific TGFβ ligand, it could indicate that they might be responsible for the secretion of these factors in the tumour environment. Moreover, if these cell-specific markers show differential expression in EMT-high conditions, as the TGFβ ligands do, it could suggest a link between

the presence of these specific cells, the expression of TGFβ ligands and the EMT-phenotype of the tumours.

4.4 The expression of TGFβ ligands positively correlates with the expression of markers of their proposed secreting cell types in HGSOCs and the same cell-type markers show higher expression in EMT-enriched tumours

In the first part of this new analysis, correlation between the expression of *TGFβ1* and T and NK markers and between *TGFβ3* and mesenchymal markers was carried out using the Spearman method (Figure 46, pg 134). The results displayed a good positive correlation between the two studied groups, a finding that was reproducible in both, the bulk RNA-Seq dataset and the microarray dataset.

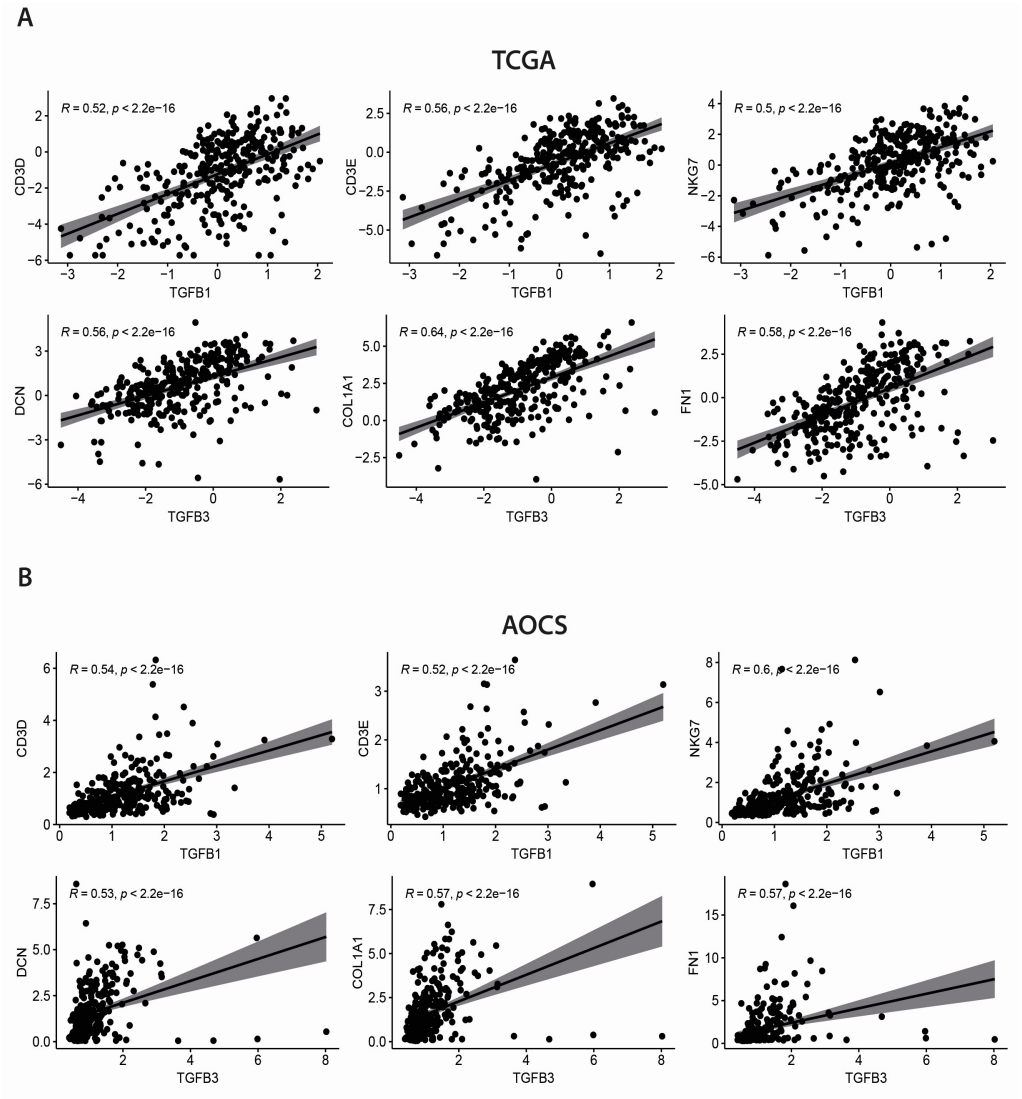


Figure 46. The positive correlation between the expression of *TGFβ1* and immune gene markers and *TGFβ3* and mesenchymal gene markers in ovarian tumours

A) Correlation of *TGFβ1* with the immune markers *CD3E* and *CD3D* (T-cell markers) and *NKG7* (NK cell markers) and *TGFβ3* with the mesenchymal markers *DCN*, *COL1A1* and *FN1* in the TCGA dataset. B) Same as above with the AOCS data. The Spearman's correlation coefficients (R) and p-values are presented on the left corner. The black straight lines represent the linear regression result and the grey regions depict 95% confidence intervals.

The previous analysis was run based on the cell types that were found to primarily express *TGFβ1* and *TGFβ3* in the fallopian tube scRNA-seq data, as the query cancer dataset is of human origin. However, the aim of the next investigation is to demonstrate that by using the mouse-derived results, the same and further complementary conclusions can be reached. The evidence obtained from mouse points at the fact that mesenchymal cells are the main producers of *TGFβ3*, in syntony with the human evidence. Previously, human and mouse mesenchymal cells have been shown to share the expression of canonical markers such as *Dcn*, *Col1a1* and *Fn1*; therefore, if the previous correlation analysis had been done using the

data derived from the mouse, the same positive correlation between mesenchymal markers and *TGFβ3* in human HGSOCs, would have been reached.

Despite the consensus achieved regarding the mesenchymal evidence across the two species, the mouse dataset showed macrophages as the main producers of *TGFβ1* in contraposition to the lymphocytes as the main *TGFβ1* secretors in the human fallopian tube. Therefore, this mouse-specific finding was further explored by executing correlation analyses between the expression of macrophages markers and *TGFβ1*. Before this, the most similar macrophage groups among the mouse and human repertoires were determined. This analysis revealed that the mouse Mac_C1qb (the one exhibiting the highest expression of *Tgfb1*) was the macrophage type that showed the highest correlation with the two classes of human macrophages, which also expressed *TGFβ1* in the fallopian tube scRNA-seq dataset. Therefore, a correlation test of conserved Mac_C1qb and human macrophages markers was run against the expression of *TGFβ1* in the HGSOC transcriptomic datasets, which showed a slightly higher positive correlation than the one that T cells and NK cells showed respect to *TGFβ1* (Figure 47, pg 135).

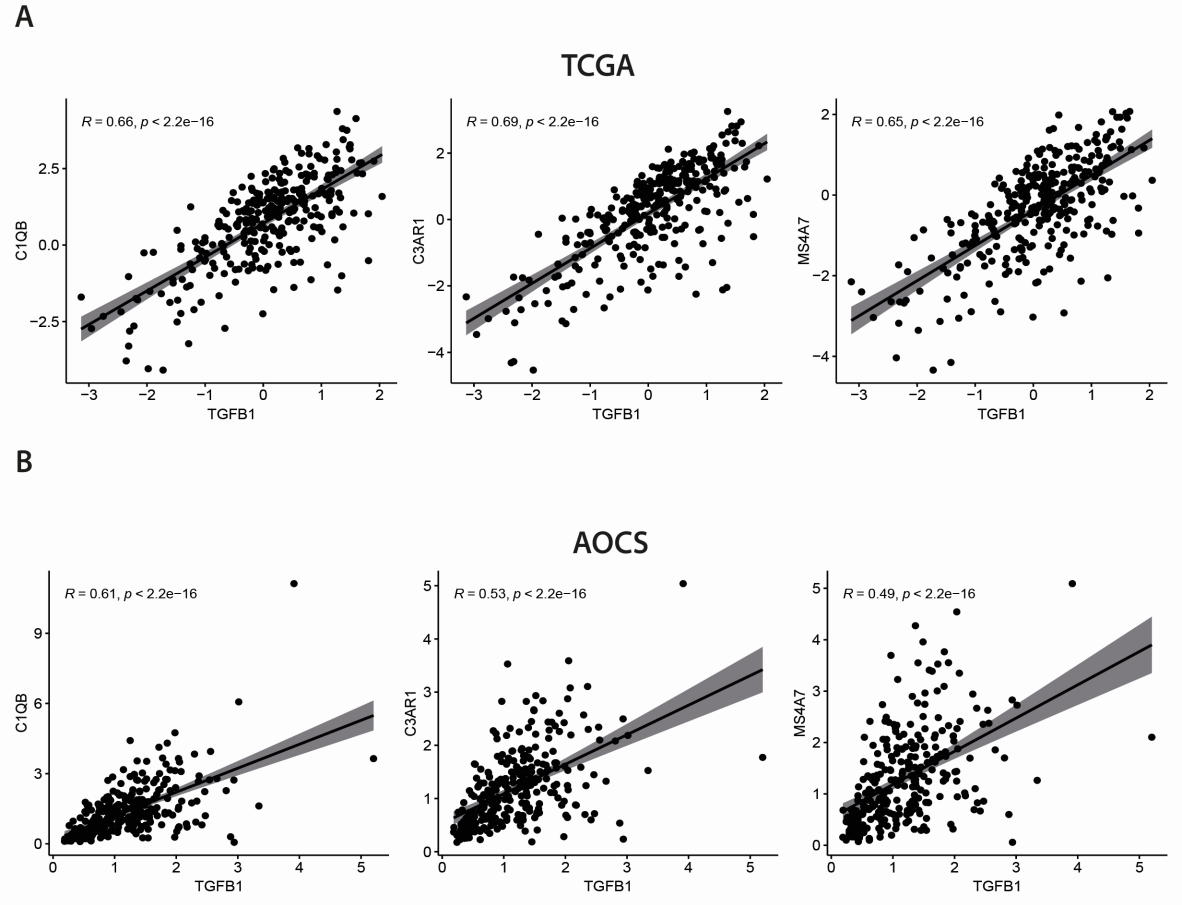


Figure 47. The positive correlation between the expression of shared murine and human macrophage gene markers and TGFβ1 in ovarian tumours
A) Correlation of the expression of *TGFβ1* with the expression of the macrophage markers *C1QB*, *C3AR1* and *MS4A7* in the TCGA dataset. B) Same as above with the AOCS data. The Spearman's correlation coefficients (R) and p-values are presented on the left corner. The black straight lines represent the linear regression result and the grey regions depict 95% confidence intervals.

Finally, prompted by the positive correlations found between the expression of cell type markers of putative TGFβ secretors and their specific TGFβ ligands, this project sought to measure their expression in HGSOCs and to correlate it with the tumour EMT status (Figure 48, pg 136). Effectively, the results showed that the cell type markers expression not only presented a good correlation with the expression of their specific TGFβ ligand, but they were also expressed at a higher level in EMT-high ovarian tumours, as their specific ligands did.

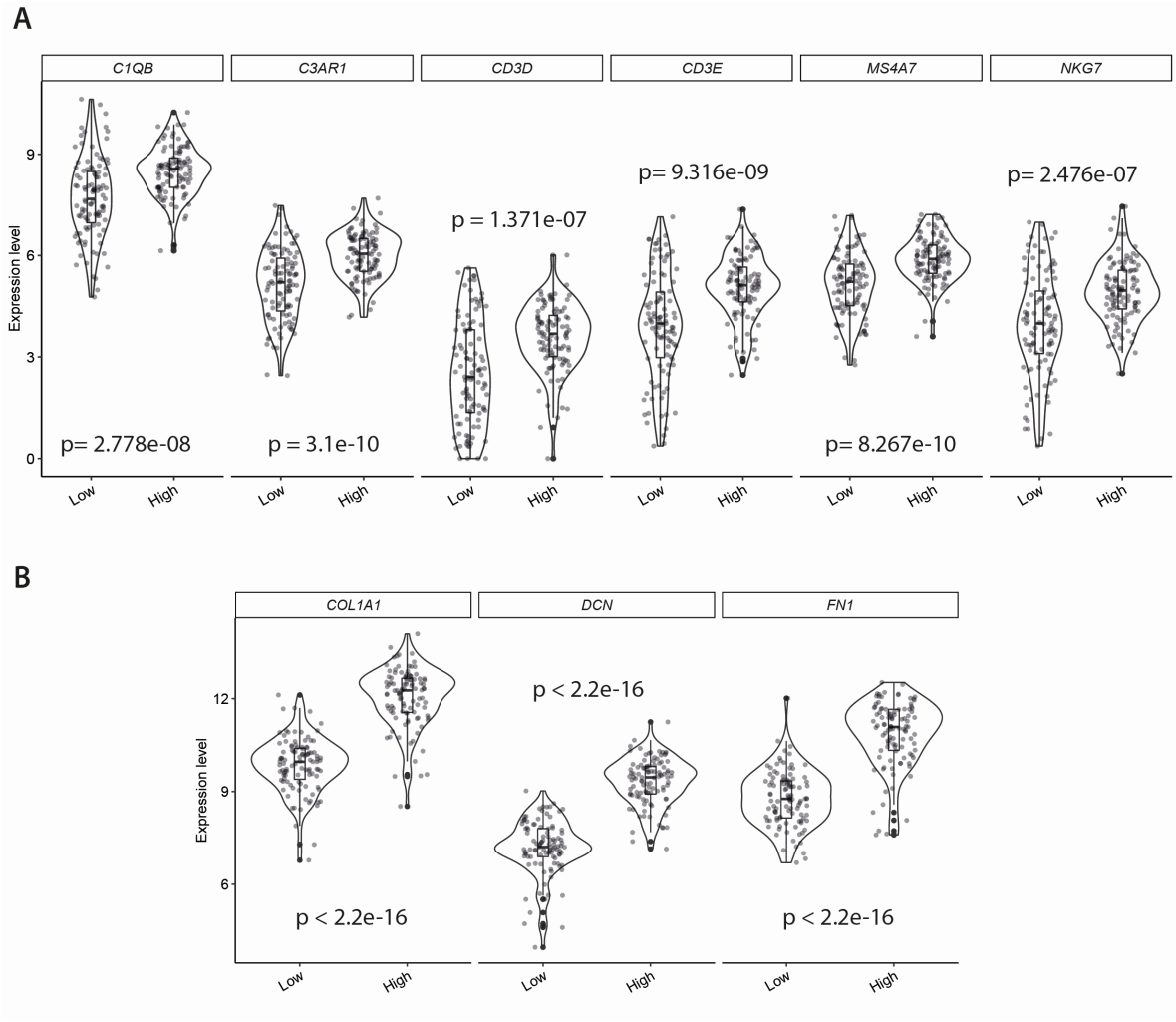


Figure 48. Differential expression of immune and mesenchymal gene markers in EMT-high versus EMT-low ovarian tumours
A) Violin plots show expression of the immune markers in low-EMT compared to high-EMT HGSOC tumours from the TCGA dataset. The immune markers are the macrophage markers *C1QB*, *C3AR1* and *MS4A7*, the T-cell markers *CD3D* and *CD3E* and the NK cell marker *NKG7*. B) Violin plots show expression of the mesenchymal markers in low-EMT compared to high-EMT HGSOC tumours from the TCGA dataset. The mesenchymal markers are *DCN*, *COL1A1* and *FN1*. The box and whiskers represent the median, two hinges (first and third quartiles) and two whiskers. The represented p-values were calculated with the Wilcoxon test.

4.5 Conclusion

In this chapter, computational and statistical analyses have unveiled new biological processes and entities underlying, not only the homeostatic oviduct and fallopian tube, but also, their pathological environment. Through the combination of interaction and expression analyses obtained from scRNA-seq datasets, and tumour-wide correlation and expression studies from averaged transcriptomic approaches, the TGFβ pathway was identified as putatively activated during homeostatic and pathological conditions of the fallopian tube. In the pathological context, two TGFβ molecules, *TGFβ1* and *TGFβ3*, were found to be highly expressed in EMT-high tumours, suggesting the possibility of a cause-effect relationship. In the normal organs, *TGFβ1* was mainly expressed by T and NK cells in fallopian tube and macrophages in the oviduct, whereas *TGFβ3* was mainly expressed by mesenchymal cells. This suggested the identity of possible cell types secreting TGFβ factors within the tumour microenvironment. To further support this hypothesis, this study not only found that the expression of those cell type markers and their specific TGFβ ligands were positively correlated in HGSOCs, but that those molecular markers also showed a differential expression in EMT-enriched tumours. This is what the data analysis shows, however, taking into account the high mesenchymal phenotype of EMT-high tumours, further analyses should be conducted to determine whether mesenchymal or cancer cells are the *TGFβ3* source in HGSOCs. Despite this open question, these results suggest a potential role for the TGFβ pathway in inducing and maintaining the EMT phenotype in EMT-high HGSOC tumours, which has previously been linked to a poor prognosis in patients.

TGFβ signalling displays a heterogenous range of activities in different environments and cell types, however, a TGFβ1 role triggering EMT in oviduct cells has previously been shown. Interestingly, this factor also acts as an immune regulator within the reproductive environment. TGFβ1 along with IL10 are known anti-inflammatory cytokines that control and supress the immune system in mouse and human alike. In the FRT, TGFβ1 was first proposed to exert this role during pregnancy, promoting immunological tolerance towards the embryo (Blois et al. 2007). This was further expanded by Yousef et al. 2016, who found that this cytokine might be linked to the oviduct immune homeostasis and immune tolerance to allogeneic sperm cells and embryos in cattle. Furthermore, the studies carried out by Jiwakanon et al. 2010 also suggested that TGFβ1 downregulates the immune response against spermatozoa. Another elegant example of this, is the *in vitro* evidence that the bovine oviduct epithelium triggers the production of interferon-tau in embryos, which modulates immune cells to produce an anti-inflammatory response where TGFβ1 is a protagonist (Talukder et al. 2018). Moreover, the present analyses have been able to highlight cell identities deserving further investigation to determine whether they are involved or affected by any of these immunosupressor, homeostatic or EMT functions potentially induced by the TGFβ pathway. The first identities to investigate would be Sec_Itga6_Dach2 and Sec_Fxyd4_Dach1, due to their capacity to receive and secrete TGFβ ligands to regulate the macrophage cluster Mac_C1qb_Pmepa1. Other interesting identities within the TGFβ framework are the intermediate cluster Cil_Mapk10 and the macrophage cluster Mac_Adgre4, given the amount of TGFβ regulation that both of them receive from their internal or external lineages.

In summary, this study has been able to shed light on the importance of this pathway in the oviduct/fallopian tube context, to identify the cell types maintaining this signalling under homeostasis and one of the concrete roles that could be carrying out within tumours derived from these organs. This set of analyses has elucidated that the TGFβ signalling is a putative communication axis in the homeostatic oviducts preferentially induced and uptaken by immune and mesenchymal cells. Furthermore, and, to the best of my knowledge, this is the first time that specific members of the TGFβ pathway have been proposed as regulators of the EMT-high phenotype of poor-prognosis HGSOC tumours. However, despite the obtained bioinformatics results and the well-known role of the TGFβ pathway as an EMT inducer, further mechanistic and molecular studies should be carried out in appropriate *in vitro* and *in vivo* models and systems to validate and expand on these findings.

V Conclusion and future directions

1 Overview and discussion

In this project, scRNA-seq was used to: (1) advance our knowledge about the cell systems, populations and underlying biology of the oviduct and the fallopian tube; (2) compare the mouse and human organs and the interactions mantaining their functionality and homeostasis; (3) discover new identities involved in the diseased state of this organ, taking HGSOC as an example.

To this end, I first accomplished the creation and optimisation of a single-cell isolation protocol for the mouse oviducts. As with any objective scientific approach, not only the advantages of this protocol were presented, but also the disadvantages were searched, identified and accounted for. For example, a technically-induced stress transcriptomic signature, very commonly found in any scRNA-seq dataset, was linked to the isolation protocol. However, attempts to bypass this stress response using alternative isolation methods proved inefficient to dissociate the organ. Despite this disadvantage, this protocol enabled the successful sequencing of 55.444 mouse oviduct cells, of which, 42.89% were selected as optimal-quality cells and used for downstream computational analysis. In these experiments, limitations with the cell capture efficiency of the 10X Genomics platform were found. When creating the transcriptomic libraries, the 10X technology displayed well below the 50% capture rate that the company claims, and the success in library preparation was highly aleatory.

Regarding the data output, all the expected oviduct lineages were detected, although the cell composition of the loaded sample and its associated data, greatly differed from each other. This remarkably affected epithelial cells, which were also lost during the FACS and scRNA-seq library preparation steps. In spite of this, secretory and ciliated cells were distinguished in the data and made up for ten clusters marked by diverse biological features. Ciliated cells presented terms related to the ciliary function, calcium, the smoothened or Hedgehog signaling pathway, GTP binding and nucleosidediphosphate kinase activity. Secretory cells enriched processes involving the endoplasmic reticulum, extracellular secretion, innate immunity and a high lipid metabolism. The first two are expected processes given the protein synthesis and secretory function of these cells. However, the discovery of the high lipid metabolic signature and the role in innate immunity, with antibacterial and antiviral molecules, were novel. Upregulation of the lipid metabolism and association with adipocytes is a hallmark of ovarian cancer cells. Therefore, the finding that normal secretory cells display a lipid metabolic signature, suggests that this is an intrinsic property of those cells and is not exclusive of the tumour ones. Furthermore, regarding the epithelial immune role, low/absent expression of the MHCI in epithelial cells, potentially ciliated cells, was documented in this project. Perhaps this could be a strategy related to the immune evasion of the allogeneic spermatozoa and zygote in the oviducts. By decreasing/dispensing the number of MHCI, the epithelium would reduce the probability of presenting sperm/zygote's material to $CD8^+$ T cells, which could elicit an immune response.

Observing deeper into the epithelial clusters, two ciliated, four secretory and two intermediate clusters were found. Among the ciliated groups, one was

identified as normal functional ciliated cells, while another resembled an immature or senescent ciliated type. The secretory clusters were:

- Sec_Itga6_Dach2, the main cluster related to innate immunity. It expressed the hormone receptors *Esr1* and *Ar* (only cluster expressing this receptor). It also established a rich communication network with almost all the cell lineages, especially, mesenchymal cells.

- Sec_Upk1a, a very inert cluster, which established a very low number of interactions with other clusters.

- Sec_Fxyd4, another low-interactive cluster whose main marker, *Fxyd4*, is expressed under progesterone regulation, during the metestrus and dioestrus stages.

- Sec_Fxyd4_Dach1, cluster validated as the epithelial cells in the uterotubal junction. It shares the same progesterone-regulated marker as stated above, it was the only cluster expressing the hormone receptors *Pgr* and *Tshr* and also displayed the highest expression of *Esr1*. This and the histological information suggest that these uterotubal epithelial cells resemble uterine rather than oviduct epithelial cells given their higher hormone-responsive phenotype. This cluster also established a large communication network with almost all members of the same and different lineages, especially, mesenchymal cells.

Futhermore, two clusters with intermediate features between the secretory and the ciliated phenotype were found, Cil_S100g and Cil_Mapk10. Cil_S100g displayed a more secretory phenotype and a low expression of mature and functional ciliated genes. Its marker gene, *S100g*, is also a progesterone-regulated gene, which along with the peptidase inhibitor activity that displayed, suggested that this cluster could be an early version of a differentiating secretory cluster into a ciliated phenotype. This differentiation could be activated by progesterone, after ovulation, to increase the number of these ciliated cells to (1) boost egg transportation and (2) promote an egg-protective function, given the peptidase inhibitor activity. This cluster was a low-interactive one, establishing almost no communication with other cells, which would fit the above theory, supposing that these cells mainly rely on progesterone signalling. Because this hormone is produced by granulosa cells, a dataset containing Cil_S100g and granulosa cells would test this theory and show whether cell-to-cell communications are established between these two subtypes. Regarding Cil_Mapk10, it showed mature and functional ciliated genes, and, interestingly, a very high lipid metabolism. It also expressed *Esr1* and Sca-1 (*Lya6*), a well known gene involved in cell proliferation, migration and stemness. This could suggest a role in proliferation/stemness, possibly related to secretory differentiation into ciliated cells. This cluster established cell-to-cell interactions with the majority of cell types, especially, mesenchymal cells. Moreover, it was the only epithelial cluster receiving growth factors from immune (*Tgfβ1*) and mesothelial cells (*Fgf1*).

After analysis of the scRNA-seq data, all the epithelial clusters were validated by snRNA-seq, which showed an excellent epithelial cell recovery and bypassed

the stress response seen by enzymatic dissociation. As a downside, it is a technique recommended for projects that do not need the characterisation of small-sized cells, such as immune/mesenchymal, as it is not able to capture them. Importantly, snRNA-seq allows the use of frozen samples, avoiding the requirement to work only with fresh samples and unlocking the opportunity to use archival material.

Additionally, this project characterised endothelial clusters, pericytes, mesothelial, mesenchymal and immune cells. The mesenchymal and immune lineages showed proliferative cells, in contrast with the epithelial ones, which indicated a low-proliferative activity of the oviduct epithelium. Mesenchymal cells displayed a homogeneous phenotype, whereas the immune repertoire, especially macrophages, was very heterogeneous. Within the macrophage heterogeneity, a developmental-origin analysis proposed the clusters M_C1qb, MC1qb_Pf4, MC1qb_Pmepa1, M_Saa3 and M_Cd300e as possible embryonic-lineage tissue-resident macrophages of the oviducts. This is further strengthened by the fact that all of them, except M_Cd300e, established a large number of interactions with the parenchymal clusters Sec_Itga6_Dach2, Sec_Fxyd4_Dach1 and Cil_Mapk10. An important pathway behind the interactions between these macrophages, concretely, Mac_Saa3 and MC1qb_Pmepa1, and these epithelial clusters, is the TGFβ pathway. Normally, a TGFβ ligand secreted by immune cells is received by these epithelial cells, however, in the case of Sec_Itga6_Dach2 and Sec_Fxyd4_Dach1, they were able to secrete *Tgfβ3* to interact with Mac_C1qb_Pmepa1.

This project also sequenced around 1000 fallopian tube cells, which were analysed together with previously generated data, accounting for a total of 40645 cells. Before, scRNA-seq had been carried out in fallopian tubes by Hu et al. 2020 and Dinh et al. 2021. However, these works had focused on studying the epithelial component, a limitation that is overcome in the current project by expanding the study to all known cell types. The fallopian tube dataset shared the cell lineages distribution seen in mouse, where immune cells are the majority, followed by mesenchymal and epithelial cells. Also, two shared features between both species are (1) that mesenchymal cells and macrophages dominate the communication landscape, suggesting that they might be key regulatory components of the organ homeostasis, and (2) that, unexpectedly, epithelial cells establish more interactions with other cell types than among themselves. An observation derived from the analysis of the aforementioned mesenchymal/immune interactions was the presence of the TGFβ pathway, predominantly, through the ligand (*Tgfβ1*). Furthermore, the presence of adaptive immunity in the fallopian tube was predominant, whereas it was little represented in the oviducts, marked by cell components of the innate immunity and the multilineage interactions of its members. Furthermore, this project executed an integrative analysis of ovarian cancer gene expression data directed by the findings extracted from the scRNA-seq analysis of the healthy organs. In this analysis, *TGFβ1* and *TGFβ3*, as well as markers of their putative secretory cells, were found to strongly correlate with an EMT-high signature in HGSOC tumours.

2 Conclusion

Overall, this project has been able to: (1) offer new insights into the cellular heterogeneity and complexity of the oviduct and fallopian tube; (2) identify possible

cell-specific biological properties; (3) draw an inter-species comparison, (4) study the diseased state of the organ; (5) initially validate some results obtained from the big-data analysis. Considering the precedents, the analysis of almost 100.000 cells from this gynaecological organ, representing two different species, is a breakthrough in the field. This advance has enabled the novel discovery and initial exploration of cell identities coexisting in the mouse oviduct (42 identities when a total of 15 was initially hypothesised) and human fallopian tube (21 identities). These results were additionally strengthened by the validation undertaken through additional diverse tools and models.

Furthermore, this thesis has deciphered and identified complex cellular relationships and interactions in these organs and their potential contribution to their biology. These results have highlighted, once more, that the concept of biology is not the product of isolated actions, but heterogeneous systemic processes. This helps us to appreciate how the appropriate orchestration and complex interactions of the functional units of systems, the cells, enable the coordinated and precise execution of the whole organ function. Moreover, this project has identified a novel research strategy to address important questions in the oncology field. Ovarian cancer and, in particular, HGSOC, is one of the most devastating and difficult-to-study diseases in women. To tackle the unsolved challenges in the field and develop more effective therapeutic strategies, a better understanding of not only the disease, but also the organ is required. For this reason, this study aimed at unravelling the normal organ cellular heterogeneity and homeostatic mechanisms that could be deregulated under pathological conditions. I believe that this thesis' work has succeeded at this, opening a new and unique perspective over the cellular and molecular properties of the healthy and diseased state of this organ.

3 Future directions

As in many time-constrained projects of this magnitude, the mechanistic validation of several observations and results generated over the course of this study would have deserved further work. Despite these limitations, this thesis has enabled the formulation of unexplored unique questions and future directions, that could not have been proposed otherwise, and that hold the potential to attract the interest of the scientific community. For example, interesting avenues for further studies in the field would be: (1) studies of the correspondence between transcriptomic identities and functionality within the organ; (2) confirmation of the developmental origin of the macrophage subtypes proposed as tissue-resident; (3) charaterisation of the possible accentuated immunological role of secretory cells; (4) the role of TGFβ signalling and the ligands TGFβ1 and TGFβ3 in aggressive EMT-high HGSOCs. Taken everything together, I believe that this project not only represents an important contribution to the gynaecological research field, but also, a valuable and promising starting point for future investigations focusing on the many evidences and hypothesis generated in this study.

VI Bibliography

Acién, P. (1993). Reproductive performance of women with uterine malformations. Human Reproduction. 8, 122-126.

Acuna, O. S., Aviles, M,, Lopez-Ubeda, R., Guillen-Martinez, A., Soriano-Ubeda, C., Torrecillas, A., Coy, P. and Izquierdo-Rico, M. J. (2017). Differential gene expression in porcine oviduct during the oestrous cycle. Reproduction, Fertility and Development. 29:2387–2399.

Adam, M., Potter, A. S. and Potter, S. S. (2017). Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Development. 144(19):3625-3632.

Adan, A., Alizada, G., Kiraz, Y., Baran, Y. and Nalbant, A. (2017). Flow cytometry: basic principles and applications. Critical Reviews in Biotechnology. 37(2):163-176.

Adil, A., Kumar, V., Jan, A. T. and Asger, M. (2021). Single-cell transcriptomics: current methods and challenges in data acquisition and analysis. Frontiers in Neuroscience. 15:591122.

Ahmed, A. A, Etemadmoghadam, D., Temple, J., Lynch, A. G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., Defazio, A., Bowtell, D. and Brenton, J. D. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. The Journal of Pathology. 221(1):49-56.

Aksoy, E., Goldman, M. and Willems, F. (2004). Protein kinase C epsilon: a new target to control inflammation and immune-mediated disorders. The International Journal of Biochemistry Cell Biology. 36(2):183-8.

Al-Dossary, A. A., Martin-Deleon, P. A. (2016). Role of exosomes in the reproductive tract Oviductosomes mediate interactions of oviductal secretion with gametes/early embryo. Frontiers in Bioscience (Landmark Ed). 21:1278-85.

Aldridge, S. and Teichmann, S. A. (2020). Single cell transcriptomics comes of age. Nature Communications.11(1):4307.

Algarra, B., Han, L., Soriano-Ubeda, C., Aviles, M., Coy, P., Jovine, L. and Jimenez-Movilla, M. (2016). The C-terminal region of OVGP1 remodels the zona pellucida and modifies fertility parameters. Scientific Reproduction. 6:32556.

Ali, A., Syed, S. M., Jamaluddin, M. F. B., Colino-Sanguino, Y., Gallego-Ortega, D. and Tanwar, P. S. (2020). Cell lineage tracing identifies hormone-regulated and Wnt-responsive vaginal epithelial stem cells. Cell Reports. 30(5):1463-1477.e7.

Alminana, C., Caballero, I., Heath, P. R., Maleki-Dizaji, S., Parrilla, I., Cuello, C., Gil, M. A., Vazquez, J. L., Vazquez, J. M., Roca, J., Martinez, E. A, Holt, W. V. and Fazeli, A. (2014). The battle of the sexes starts in the oviduct: modulation of oviductal transcriptome by X and Y-bearing spermatozoa. BMC Genomics. 15:293.

Alwosaibai, K., Abedini, A., Al-Hujaily, E. M., Tang, Y., Garson, K., Collins, O. and Vanderhyden, B. C. (2017). PAX2 maintains the differentiation of mouse oviductal epithelium and inhibits the transition to a stem cell-like state. Oncotarget. 8(44):76881-76897.

Anders, S. and Huber, W. Differential expression analysis for sequence count data. (2010). Genome Biology.11(10):R106.

Araki, Y., Nohara, M., Yoshida-Komiya, H., Kuramochi, T., Ito, M., Hoshi, H., Shinkai, Y. and Sendai, Y. (2003). Effect of a null mutation of the oviduct-specific glycoprotein gene on mouse fertilization. Biochemical Journal. 374:551–557.

Arda, I. S., Ergeneli, M., Coskun, M. and Hicsonmez, A. (2004). Tubo-ovarian abscess in a sexually inactive adolescent patient. European Journal of Pediatric Surgery. 14(1):70-2.

Ardighieri, L., Lonardi, S., Moratto, D., Facchetti, F., Shih, IeM., Vermi, W. and Kurman, R. J. (2014). Characterization of the immune cell repertoire in the normal fallopian tube. International Journal of Gynecological Pathology. 33(6):581-91.

Aronoff, D. M., Canetti, C. and Peters-Golden, M. (2004). Prostaglandin E2 inhibits alveolar macrophage phagocytosis through an E-prostanoid 2 receptor-mediated increase in intracellular cyclic AMP. Journal of Immunology. 173(1):559-65.

Artibani, M., Masuda, K., Hu, Z., Rauher, P. C., Mallett, G., Wietek, N., Morotti, M., Chong, K., KaramiNejadRanjbar, M., Zois, C. E., Dhar, S., El-Sahhar, S., Campo, L., Blagden, S. P., Damato, S., Pathiraja, P. N., Nicum, S., Gleeson, F., Laios, A., Alsaadi, A., Santana Gonzalez, L., Motohara, T., Albukhari, A., Lu, Z., Bast, R. C. Jr., Harris, A. L., Ejsing, C. S., Klemm, R. W., Yau, C., Sauka-Spengler, T., Ahmed, A. A. (2021). Adipocyte-like signature in ovarian cancer minimal residual disease identifies metabolic vulnerabilities of tumor-initiating cells. JCI Insight. 6(11):e147929.

Bahat, A., Tur-Kaspa, I., Gakamsky, A., Giojalas, L. C., Breitbart, H. and Eisenbach, M. (2003). Thermotaxis of mammalian sperm cells: a potential navigation mechanism in the female genital tract. Nature Medicine. 9(2):149-50.

Baillie, H. S., Pacey, A. A., Warren, M. A., Scudamore, I. W., Barratt and C. L. (1997) Greater numbers of human spermatozoa associate with endosalpingeal cells derived from the isthmus compared with those from the ampulla. Human Reproduction. 12(9):1985-92.

Bangs, F. and Anderson, K. V. (2017). Primary cilia and mammalian hedgehog signaling. Cold Spring Harbor Perspectives in Biology. 9(5):a028175

Baron, M., Veres, A., Wolock, S. L., Faust, A. L., Gaujoux, R., Vetere, A., Ryu, J. H., Wagner, B. K., Shen-Orr, S. S., Klein, A. M., Melton, D. A. and Yanai, I. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Systems. 3(4):346-360.e4.

Basic, M., Buettner, M., Keubler, L. M., Smoczek, A., Bruesch, I., Buchheister, S. and Bleich, A. (2018). Loss of CD14 leads to disturbed epithelial-B cell crosstalk and impairment of the intestinal barrier after E. coli Nissle monoassociation. Scientific Reports. 8(1):719.

Bathala, P., Fereshteh, Z., Li, K., Al-Dossary, A. A., Galileo, D. S. and Martin-DeLeon, P. A. (2018) Oviductal extracellular vesicles (oviductosomes, OVS) are conserved in humans: murine OVS play a pivotal role in sperm capacitation and fertility. Molecular Human Reproduction. 24(3):143-157.

Batts, T. D., Machado, H. L., Zhang, Y., Creighton, C. J., Li, Y. and Rosen, J. M. (2011). Stem cell antigen-1 (sca-1) regulates mammary tumor development and cell migration. PLoS One. 6(11):e27841.

Bauersachs, S., Blum, H., Mallok, S., Wenigerkind, H., Rief, S., Prelle, K. and Wolf, E. (2003) Regulation of ipsilateral and contralateral bovine oviduct epithelial cell function in the postovulation period: a transcriptomics approach. Biology of Reproduction. 68(4):1170-7.

Blatt, E. N., Yan, X. H., Wuerffel, M. K., Hamilos, D. L. and Brody, S. L. (1999). Forkhead transcription factor HFH-4 expression is temporally related to ciliogenesis. American Journal of Respiratory Cell and Molecular Biology. 21(2):168-76.

Blois, S. M., Kammerer, U., Alba Soto, C., Tometten, M. C., Shaikly, V., Barrientos, G., Jurd, R., Rukavina, D., Thomson, A. W., Klapp, B. F., Fernández, N. and Arck, P. C. (2007). Dendritic cells: key to fetal tolerance? Biology of Reproduction. 77(4):590-8.

Bockerstett, K. A., Wong, C. F., Koehm, S., Ford, E. L. and DiPaolo, R. J. (2018). Molecular characterization of gastric epithelial cells using flow cytometry. International Journal of Molecular Science. 19(4):1096.

Boelhauve, M., Sinowatz, F., Wolf, E. and Paula-Lopes, F. F. (2005). Maturation of bovine oocytes in the presence of leptin improves development and reduces apoptosis of in vitro-produced blastocysts. Biology of Reproduction. 73(4):737-44.

Box, A., DeLay, M., Tighe, S., Chittur, S. V., Bergeron, A., Cochran, M., Lopez, P., Meyer, E. M., Saluk, A., Thornton, S. and Brundage, K. (2020). Evaluating the Effects of Cell Sorting on Gene Expression. Journal of Biomolecular Techniques. 20-3103-004.

Buhi, W. C., Alvarez, I. M. and Kouba, A. J. (2000). Secreted proteins of the oviduct. Cells Tissues Organs. 166(2):165-79.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology. 36(5):411-420.

Calandra, T. and Roger, T. (2003). Macrophage migration inhibitory factor: a regulator of innate immunity. Nature Reviews Immunology. 3(10):791-800.

Camara Pirez, M., Steele, H., Reese, S. and Kölle, S. (2020). Bovine sperm-oviduct interactions are characterized by specific sperm behaviour, ultrastructure and tubal reactions which are impacted by sex sorting. Scientific Reports. 10(1):16522.

Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. Nature. 474(7353):609-15.

Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics. 45(10):1113-20.

Cancer Research UK. Ovarian cancer survival statistics. cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/mortalityheading-Zero.

Cardenas, H., Corvalan, L. and Imarai, M. (1998). Is there a mucosal immune system associated with the mammalian oviduct? Biological Research. 31(4):329-38.

Cao, Y., Zhu, J., Jia, P. and Zhao, Z. (2017). scRNASeqDB: A database for RNA-Seq based gene expression profiles in human single cells. Genes (Basel). 8(12):368.

Carlson, A. E., Westenbroek, R. E., Quill, T., Ren, D., Clapham, D. E., Hille, B., Garbers, D. L. and Babcock, D. F. (2003). CatSper1 required for evoked Ca2+ entry and control of flagellar function in sperm. PNAS. 100(25):14864-8.

Cavanagh, T. J., Lakey, J. R., Wright, M. J., Fetterhoff, T. and Wile, K. (1997). Crude collagenase loses islet-isolating efficacy regardless of storage conditions. Transplantation Proceedings. (4):1942-4.

Cerny, K. L., Ribeiro, R. A., Jeoung, M., Ko, C. and Bridges, P. J. (2016). Estrogen receptor alpha (ESR1)-dependent regulation of the mouse oviductal transcriptome. PLoS One. 11(1):e0147685.

Chang, H. S., Cheng, W. T., Wu, H. K. and Choo, K. B. (2000). Identification of genes expressed in the epithelium of porcine oviduct containing early embryos at various stages of development. Molecular Reproduction and Development. 56:331–335.

Chen, J., Knowles, H. J., Hebert, J. L. and Hackett, B. P. (1998). Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. Journal of Clinical Investigation. 102(6):1077-82.

Chen, C., Song, X., Wei, W., Zhong, H., Dai, J., Lan, Z., Li, F., Yu, X., Feng, Q., Wang, Z., Xie, H., Chen, X., Zeng, C., Wen, B., Zeng, L., Du, H., Tang, H., Xu, C., Xia, Y., Xia, H., Yang, H., Wang, J., Wang, J., Madsen, L., Brix, S., Kristiansen, K., Xu, X., Li, J., Wu, R. and Jia, H. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. (2017). Nature Communications. 17;8(1):875.

Chestnut, B., Casie Chetty, S., Koenig, A. L. and Sumanas, S. (2020). Single-cell transcriptomic analysis identifies the conversion of zebrafish Etv2-deficient vascular progenitors into skeletal muscle. Nature Communications. 11(1):2796.

Chmiest, D., Sharma, N., Zanin, N., Viaris de Lesegno, C., Shafaq-Zadah, M., Sibut, V., Dingli, F., Hupé, P., Wilmes, S., Piehler, J., Loew, D., Johannes, L., Schreiber, G. and Lamaze, C. (2016). Spatiotemporal control of interferon-induced JAK/STAT signalling and gene transcription by the retromer complex. Nature Communication. 7:13476.

Choi, K. C., Leung, P. C. and Jeung, E. B. (2005). Biology and physiology of Calbindin-D9k in female reproductive tissues: involvement of steroids and endocrine disruptors. Reproductive Biology and Endocrinology. 3:66.

Crowley, L., Cambuli, F., Aparicio, L., Shibata, M., Robinson, B. D., Xuan, S., Li, W., Hibshoosh, H., Loda, M., Rabadan, R. and Shen M. M. (2020) A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. Elife. 9:e59465.

Cortés, P. P., Orihuela, P. A., Zúñiga, L. M., Velásquez, L. A. and Croxatto, H. B. (2004). Sperm binding to oviductal epithelial cells in the rat: role of sialic acid

residues on the epithelial surface and sialic acid-binding sites on the sperm surface. Biology of Reproduction. 71(4):1262-9.

Coy, P., Canovas, S., Mondejar, I., Saavedra, M. D., Romar, R., Grullon, L., Matas, C. and Aviles, M. (2008). Oviduct-specific glycoprotein and heparin modulate sperm-zona pellucida interaction during fertilization and contribute to the control of polyspermy. PNAS. 105:15809–15814.

Cunha, G. R., Robboy, S. J., Kurita, T., Isaacson, D., Shen, J., Cao, M. Baskin, L. S. (2018). Development of the human female reproductive tract. Differentiation. 103, 46-65.

Curtis, J. E. (2016). The mechanics of ovulation depend on an incredibly soft and sugar-rich extracellular matrix. Biophysical Journal. 110(12):2566-2567.

Daar, A. S., Fuggle, S. V., Fabre, J. W., Ting, A. and Morris, P. J. (1984). The detailed distribution of HLA-A, B, C antigens in normal human organs. Transplantation. 38(3):287-92.

Davies, L. C., Jenkins, S. J., Allen, J. E. and Taylor, P. R. (2013). Tissue-resident macrophages. Nature Immunology. 14(10):986-95.

De Jonge, C. (2005). Biological basis for human capacitation. Human Reproduction Update. 11(3):205-14.

DeCaen, P. G., Delling, M., Vien, T. N and Clapham, D. E. (2013). Direct recording and molecular identification of the calcium channel of primary cilia. Nature. 504(7479):315-8.

Delling, M., DeCaen, P. G., Doerner, J. F., Febvay, S. and Clapham, D. E. (2013). Primary cilia are specialized calcium signalling organelles. Nature. 504(7479):311-4.

Demott, R. P. and Suarez, S. S. (1992). Hyperactivated sperm progress in the mouse oviduct. Biology of Reproduction. 46(5):779-85.

Demott, R. P., Lefebvre, R. and Suarez, S. S. (1995). Carbohydrates mediate the adherence of hamster sperm to oviductal epithelium. Biology of Reproduction. 52(6):1395-403.

Deng, W. B., Tian, Z., Liang, X. H., Wang, B. C., Yang, F. and Yang, Z. M. (2013). Progesterone regulation of Na/K-ATPase β1 subunit expression in the mouse uterus during the peri-implantation period. Theriogenology. 79(8):1196-203.

Denisenko, E., Guo, B. B., Jones ,M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R. K., Lister, R., Forrest, A. R. R. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biology. 21(1):130.

Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology. 4(5):P3.

Dieleman, S. J., Hendriksen, P. J., Viuff, D., Thomsen, P. D., Hyttel, P., Knijn, H. M., Wrenzycki, C., Kruip, T. A., Niemann, H., Gadella, B. M., Bevers, M. M. and Vos, P. L. Effects of in vivo prematuration and in vivo final maturation on developmental capacity and quality of pre-implantation embryos (2002). Theriogenology.

57(1):5-20.

Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., Hughes, T. K., Wadsworth, M. H., Burks, T., Nguyen, L. T., Kwon, J. Y. H., Barak, B., Ge, W., Kedaigle, A. J., Carroll, S., Li, S., Hacohen, N., Rozenblatt-Rosen, O., Shalek, A. K., Villani, A. C., Regev, A., Levin, J. Z. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nature Biotechnology. 38(6):737-746.

Dinh, H. Q., Lin, X., Abbasi, F., Nameki, R., Haro, M., Olingy, C. E., Chang, H., Hernandez, L., Gayther, S. A., Wright, K. N., Aspuria, P. J., Karlan, B. Y., Corona, R. I., Li, A., Rimel, B. J., Siedhoff, M. T., Medeiros, F. and Lawrenson, K. (2021) Single-cell transcriptomics identifies gene expression networks driving differentiation and tumorigenesis in the human fallopian tube. Cell Reports. 35(2):108978.

Dolmans, M. M., Michaux, N., Camboni, A., Martinez-Madrid, B., Van Langendonckt, A., Nottola, S. A. and Donnez, J. (2006). Evaluation of Liberase, a purified enzyme blend, for the isolation of human primordial and primary ovarian follicles. Human Reproduction. 21(2):413-20.

Dreismann, A. K., McClements, M. E., Barnard, A. R., Orhan, E., Hughes, J. P., Lachmann, P. J. and MacLaren, R. E. (2021). Functional expression of complement factor I following AAV-mediated gene delivery in the retina of mice and human cells. Gene Therapy. 28(5):265-276.

Drokhlyansky, E., Smillie, C. S., Van Wittenberghe, N., Ericsson, M., Griffin, G. K., Eraslan, G., Dionne, D., Cuoco, M. S., Goder-Reiser, M. N., Sharova, T., Kuksenko, O., Aguirre, A. J., Boland, G. M., Graham, D., Rozenblatt-Rosen, O., Xavier, R. J. and Regev, A. (2020). The human and mouse enteric nervous system at single-cell resolution. Cell. 182(6):1606-1622.e23.

Eddy, C. A. and Pauerstein, C. J. (1980). Anatomy and physiology of the fallopian tube. Clinical Obstetrics and Gynecology. (4):1177-93.

Efremova, M., Vento-Tormo, M., Teichmann, S. A. and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nature Protocols. 15(4):1484-1506.

Eisenbach, M. (1999). Mammalian sperm chemotaxis and its association with capacitation. Developmental Genetics. 25(2):87-94.

Ellison, R. T. 3rd., Giehl, T. J. and LaForce, F. M. (1988). Damage of the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. Infection and Immunity. 56(11):2774-81.

Elson, C. J., Salim, R., Potdar, N., Chetty, M., Ross, J. A., Kirk, E. J. on behalf of the Royal College of Obstetricians and Gynaecologists. (2016). Diagnosis and management of ectopic pregnancy. BJOG. 123:e15–e55.

Evans, W. H. and Martin, P. E. (2002). Gap junctions: structure and function (Review). Molecular Membrane Biology. 19(2):121-36.

Farquhar, C. M. (2005). Ectopic pregnancy. Lancet. 366(9485):583-91.

Fazeli, A., Affara, N. A., Hubank, M. and Holt, W. V. (2004). Sperm-induced modification of the oviductal gene expression profile after natural insemination in

mice. Biology of Reproduction. 71(1):60-5.

Feng, H., Zhang, Y. B., Gui, J. F., Lemon, S. M. and Yamane, D. (2021). Interferon regulatory factor 1 (IRF1) and anti-pathogen innate immune responses. PLoS Pathogens. 17(1):e1009220.

Fernández, J. E., Concha, A., Aránega, A., Ruiz-Cabello, F., Cabrera, T. and Garrido, F. (1991). HLA class I and II expression in rhabdomyosarcomas. Immunobiology. 182(5):440-8.

Ferron, A., Perez-Ayala, M., Concha, A., Cabrera, T., Redondo, M., Oliva, M. R., Ruiz-Cabello, F. and Garrido, F. (1989). MHC class I and II antigens on gastric carcinomas and autologous mucosa. Journal of Immunogenetics. 16(4-5):413-23.

Fischer, V., Kalbitz, M., Müller-Graf, F., Gebhard, F., Ignatius, A., Liedert, A. and Haffner-Luntzer, M. (2018). Influence of menopause on inflammatory cytokines during murine and human bone fracture healing. International Journal of Molecular Sciences. 19(7):2070.

Fléchon, J. E. and Hunter, R. H. (1981). Distribution of spermatozoa in the uterotubal junction and isthmus of pigs, and their relationship with the luminal epithelium after mating: a scanning electron microscope study. Tissue Cell. 13(1):127-39.

Ford, M. J., Harwalkar, K., Pacis, A. S., Maunsell, H., Wang, Y. C., Badescu, D., Teng, K., Yamanaka, N., Bouchard, M. Ragoussis, J. and Yamanaka, Y. Oviduct epithelial cells constitute two developmentally distinct lineages that are spatially separated along the distal-proximal axis. Under review at Cell Reports. http://dx.doi.org/10.2139/ssrn.3710743

Frantz, C., Stewart, K. M., Weaver, V. M. (2010). The extracellular matrix at a glance. Journal of Cell Science. 123(Pt 24):4195-200.

Franzén, O., Gan, L. M. and Björkegren, J. L. M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford). 2019:baz046.

Funda, D. P., Tucková, L., Farré, M. A., Iwase, T., Moro, I. and Tlaskalová-Hogenová, H. (2001). CD14 is expressed and released as soluble CD14 by human intestinal epithelial cells in vitro: lipopolysaccharide activation of epithelial cells revisited. Infection and Immunity. 69(6):3772-81.

Gabler, C., Odau, S., Muller, K., Schon, J., Bondzio, A. and Einspanier, R. (2008) Exploring cumulus-oocyte-complex-oviductal cell interactions: gene profiling in the bovine oviduct. Journal of Physiology and Pharmacology. 59 Suppl 9:29-42.

Gabor, F., Jahn, G., Sedmak, D. D. and Sinzger, C. (2020). In vivo downregulation of MHC class I molecules by HCMV occurs during all phases of viral replication but is not always complete. Frontiers in Cellular and Infection Microbiology. 10:283.

Galletti, G., De Simone, G., Mazza, E. M. C., Puccio, S., Mezzanotte, C., Bi, T. M., Davydov, A. N., Metsger, M., Scamardella, E., Alvisi, G., De Paoli, F., Zanon, V., Scarpa, A., Camisa, B., Colombo, F. S., Anselmo, A., Peano, C., Polletti, S., Mavilio, D., Gattinoni, L., Boi, S. K., Youngblood, B. A., Jones, R. E., Baird, D. M., Gostick, E., Llewellyn-Lacey, S., Ladell, K., Price, D. A., Chudakov, D. M., Newell, E. W., Casucci, M. and Lugli, E. (2020). Two subsets of stem-like CD8+ memory

T cell progenitors with distinct fate commitments in humans. Nature Immunology. 21(12):1552-1562.

Garcia-Lora, A., Algarra, I. and Garrido, F. (2003). MHC class I antigens, immune surveillance, and tumor immune escape. Journal of Cellular Physiology. 195(3):346-55.

Gautier, E. L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K. G., Gordonov, S., Mazloom, A. R., Ma'ayan, A., Chua, W. J., Hansen, T. H., Turley, S. J., Merad, M., Randolph, G. J. and Immunological Genome Consortium. (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nature Immunology. 13(11):1118-28.

Gehring, J., Hwee Park, J., Chen, S., Thomson, M. and Pachter, L. (2020). Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. Nature Biotechnology. 38(1):35-38.

Ghosh, A., Syed, S. M. and Tanwar, P. S. (2017). In vivo genetic cell lineage tracing reveals that oviductal secretory cells self-renew and give rise to ciliated cells. Development. 144(17):3031-3041.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M. F., Conway, S. J., Ng, L. G., Stanley, E. R., Samokhvalov, I. M. and Merad, M. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science. 330(6005):841-5.

Goetz, S. C., Ocbina, P. J. and Anderson, K. V. (2009). The primary cilium as a Hedgehog signal transduction machine. Methods in Cell Biology. 94:199-222.

Green, L. K. and Harris, R. E. (1976). Uterine anomalies. Frequency of diagnosis and associated obstetric complications. Obstetrics Gynecology. 47, 427-429.

Guerrier, D., Mouchel, T., Pasquier, L. and Pellerin, I. (2006). The Mayer-Rokitansky-Küster-Hauser syndrome (congenital absence of uterus and vagina) – phenotypic manifestations and genetic approaches. Journal of Negative Results in Biomedicine. 5, 1. Gumbiner, B. M. (1996). Cell adhesion: the molecular basis of tissue architecture and morphogenesis. Cell. 84(3):345-57.

Halper, J. and Kjaer, M. (2014). Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. Advances in Experimental Medicine and Biology. 802:31-47.

Hasan, M. M., Viil, J., Lättekivi, F., Ord, J., Reshi, Q. U. A., Jääger, K., Velthut-Meikas, A., Andronowska, A., Jaakma, Ü., Salumets, A. and Fazeli, A. (2020) Bovine follicular fluid and extracellular vesicles derived from follicular fluid alter the bovine oviductal epithelial cells transcriptome. International Journal of Molecular Sciences. 21(15):5365.

Heanue, T. A., Davis, R. J., Rowitch, D. H., Kispert, A., McMahon, A. P., Mardon, G. and Tabin, C. J. (2002). Dach1, a vertebrate homologue of Drosophila dachshund, is expressed in the developing eye and ear of both chick and mouse and is regulated independently of Pax and Eya genes. Mechanisms of Development. 111(1-2):75-87.

Hie, B., Peters, J., Nyquist, S. K., Shalek, A. K., Berger, B. and Bryson, B. D.

(2020). Computational Methods for Single-Cell RNA Sequencing. Annual Review of Biomedical Data Science. 3:1, 339-364.

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., Leiserson, M. D. M., Niu, B., McLellan, M. D., Uzunangelov, V., Zhang, J., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A. A., Van't Veer, L. J., Lopez-Bigas, N., Laird, P. W., Raphael, B. J., Ding, L., Robertson, A. G., Byers, L. A., Mills, G. B., Weinstein, J. N., Van Waes, C., Chen, Z., Collisson, E. A.; Cancer Genome Atlas Research Network, Benz, C. C., Perou, C. M. and Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 158(4):929-944.

Holmes, C. and Stanford, W. L. (2007). Concise review: stem cell antigen-1: expression, function, and enigma. Stem Cells. 25(6):1339-47.

Hourcade, J. D, Pérez-Crespo, M., Fernández-González, R., Pintado, B. and Gutiérrez-Adán, A. (2010). Selection against spermatozoa with fragmented DNA after postovulatory mating depends on the type of damage. Reproductive Biology and Endocrinology. 8:9.

Hoving, J. C., Keeton, R., Höft, M. A., Ozturk, M., Otieno-Odhiambo, P. and Brombacher, F. (2020). IL-4 receptor-alpha signalling of intestinal epithelial cells, smooth muscle cells, and macrophages plays a redundant role in oxazolone colitis. Mediators of Inflammation. 2020:4361043.

Hoyler, T., Klose, C. S., Souabni, A., Turqueti-Neves, A., Pfeifer, D., Rawlins, E. L., Voehringer, D., Busslinger, M. and Diefenbach, A. (2012). The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. Immunity. 37(4):634-48.

Hsu, Y. C., Mildenstein, K., Hunter, K., Tkachenko, O. and Mullen, C. A. (2014). Acute lymphoid leukemia cells with greater stem cell antigen-1 (Ly6a/Sca-1) expression exhibit higher levels of metalloproteinase activity and are more aggressive in vivo. PLoS One. 9(2):e88966.

Hu, Z., Artibani, M., Alsaadi, A., Wietek, N., Morotti, M., Shi, T., Zhong, Z., Santana Gonzalez, L., El-Sahhar, S., KaramiNejadRanjbar, M., Mallett, G., Feng, Y., Masuda, K., Zheng, Y., Chong, K., Damato, S., Dhar, S., Campo, L., Garruto Campanile, R., Soleymani Majd, H., Rai, V., Maldonado-Perez, D., Jones, S., Cerundolo, V., Sauka-Spengler, T., Yau, C. and Ahmed, A. A. (2020). The repertoire of serous ovarian cancer non-genetic heterogeneity revealed by single-cell sequencing of normal fallopian tube epithelial cells. Cancer Cell. 37(2):226-242.

Huang, H. L., Hsing, H. W., Lai, T. C., Chen, Y. W., Lee, T. R., Chan, H. T., Lyu, P. C., Wu, C. L., Lu, Y. C., Lin, S. T., Lin, C. W., Lai, C. H., Chang, H. T., Chou, H. C. and Chan, H. L. (2010). Trypsin-induced proteome alteration during cell subculture in mammalian cells. Journal of Biomedical Science. 17(1):36.

Huang, V. W., Lee, C. L., Lee, Y. L., Lam, K. K., Ko, J. K., Yeung, W. S., Ho, P. C. and Chiu, P. C. (2015). Sperm fucosyltransferase-5 mediates spermatozoa-oviductal epithelial cell interaction to protect human spermatozoa from oxidative damage. Molecular Human Reproduction. 21(6):516-26.

Hugentobler, S. A., Diskin, M. G., Leese, H. J., Humpherson, P. G., Watson, T.,

Sreenan, J. M. and Morris, D. G. (2007). Amino acids in oviduct and uterine fluid and blood plasma during the estrous cycle in the bovine. Molecular Reproduction and Development. 74:445–454.

Humblot, P., Holm, P., Lonergan, P., Wrenzycki, C., Lequarré, A. S., Joly, C. G., Herrmann, D., Lopes, A., Rizos, D., Niemann, H. and Callesen, H. (2005). Effect of stage of follicular growth during superovulation on developmental competence of bovine oocytes. Theriogenology. 63(4):1149-66.

Hunter, R. H. and Nichol, R. (1986). A preovulatory temperature gradient between the isthmus and ampulla of pig oviducts during the phase of sperm storage. Journal of Reproduction and Fertility. 77(2):599-606.

Hunter, R. H., Fléchon, B. and Fléchon, J. E. (1991) Distribution, morphology and epithelial interactions of bovine spermatozoa in the oviduct before and after ovulation: a scanning electron microscope study. Tissue Cell. 23(5):641-56.

Hwang, B., Lee, J. H. and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental and Molecular Medicine. 50(8):1-14.

Ibrahim, S., Salilew-Wondim, D., Rings, F., Hoelker, M., Neuhoff, C., Tholen, E., Looft, C., Schellander, K. and Tesfaye, D. (2015). Expression pattern of inflammatory response genes and their regulatory micrornas in bovine oviductal cells in response to lipopolysaccharide: implication for early embryonic development. PLoS One. 10(3):e0119388.

Ignotz, G. G., Lo, M., Perez, C., Gwathmey, T. M. and Suarez, S. S. (2001). Characterization of a fucose-binding protein from bull sperm and seminal plasma responsible for formation of the oviductal sperm reservoir. Biology of Reproduction. 64, 1806–1811.

Imtiaz, M. T., Schripsema, J. H., Sigar, I. M. and Ramsey, K. H. (2006). Outcome of urogenital infection with Chlamydia muridarum in CD14 gene knockout mice. BMC Infectious Diseases. 6:144.

Ishido, S., Wang, C., Lee, B. S., Cohen, G. B. and Jung, J. U. (2000). Downregulation of major histocompatibility complex class I molecules by Kaposi's sarcoma-associated herpesvirus K3 and K5 proteins. Journal of Virology. 74(11):5300-9.

Jenkins, T. G., Aston, K. I., Trost, C., Farley, J., Hotaling, J. M. and Carrell, D. T. (2015). Intra-sample heterogeneity of sperm DNA methylation. Molecular Human Reproduction. 21(4):313-9.

Jindal, A., Gupta, P., Jayadeva and Sengupta D. (2018). Discovery of rare cells from voluminous single cell expression data. Nature Communications. 9(1):4719.

Jiwakanon, J., Berg, M., Persson, E., Fossum, C. and Dalin, A. M. (2010). Cytokine expression in the gilt oviduct: effects of seminal plasma, spermatozoa and extender after insemination. Animal Reproduction Science. 119(3-4):244-57.

Justice, M. J. and Dhillon, P. (2016). Using the mouse to model human disease: increasing validity and reproducibility. Disease Models and Mechanisms. 9(2):101-3.

Kalucka, J., de Rooij, L. P. M. H., Goveia, J., Rohlenova, K., Dumas, S. J., Meta, E., Conchinha, N. V., Taverna, F., Teuwen, L. A., Veys, K., García-Caballero, M., Khan, S., Geldhof, V., Sokol, L., Chen, R., Treps, L., Borri, M., de Zeeuw, P.,

Dubois, C., Karakach, T. K., Falkenberg, K. D., Parys, M., Yin, X., Vinckier, S., Du, Y., Fenton, R. A., Schoonjans, L., Dewerchin, M., Eelen, G., Thienpont, B., Lin, L., Bolund, L., Li, X., Luo, Y. and Carmeliet, P. (2020). Single-Cell transcriptome atlas of murine endothelial cells. Cell. 180(4):764-779.e20.

Kessler, M., Hoffmann, K., Brinkmann, V., Thieck, O., Jackisch, S., Toelle, B., Berger, H., Mollenkopf, H. J., Mangler, M., Sehouli, J., Fotopoulou, C. and Meyer, T. F. (2015). The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. Nature Communications. 6:8989.

Khan, M. R., Chandrashekran, A., Smith, R. K. and Dudhia, J. (2016). Immunophenotypic characterization of ovine mesenchymal stem cells. Cytometry A. 89(5):443-50.

Kim, J. M., Park, J. E., Yoo, I., Han, J., Kim, N., Lim, W. J., Cho, E. S., Choi, B., Choi, S., Kim, T. H., Te Pas, M. F. W., Ka, H. and Lee, K. T. (2018). Integrated transcriptomes throughout swine oestrous cycle reveal dynamic changes in reproductive tissues interacting networks. Scientific Reports. 8(1):5436.

Kim, J., Coffey, D. M., Creighton, C. J., Yu, Z., Hawkins, S. M. and Matzuk, M. M. (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. PNAS. 109(10):3921-6.

Kodithuwakku, S. P., Miyamoto, A. and Wijayagunawardane, M. P. (2007). Spermatozoa stimulate prostaglandin synthesis and secretion in bovine oviductal epithelial cells. Reproduction. 133(6):1087-94.

Kölle, S., Dubielzig, S., Reese, S., Wehrend, A., König, P. and Kummer, W. (2009). Ciliary transport, gamete interaction, and effects of the early embryo in the oviduct: ex vivo analyses using a new digital videomicroscopic system in the cow. Biology of Reproduction. (2):267-74.

Koller, F. L., Hwang, D. G., Dozier, E. A. and Fingleton, B. (2010). Epithelial interleukin-4 receptor expression promotes colon tumor growth. Carcinogenesis. 31(6):1010-7.

Komatsu, M. and Fujita, H. (1978). Electron-microscopic studies on the development and aging of the oviduct epithelium of mice. Anatomy and Embryology. 152(3), 243– 259.

Kuhn, E., Kurman, R. J., Sehdev, A. S. and Shih, IeM. (2012). Ki-67 labeling index as an adjunct in the diagnosis of serous tubal intraepithelial carcinoma. International Journal of Gynecological Pathology. 31(5):416-22.

Kulkarni, A., Anderson, A. G., Merullo, D. P. and Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. Current Opinion in Biotechnology. 58:129-136.

L'Horset, F., Perret, C., Brehier, A. and Thomasset, M. (1990). 17 beta-estradiol stimulates the calbindin-D9k (CaBP9k) gene expression at the transcriptional and posttranscriptional levels in the rat uterus. Endocrinology. 127(6):2891-7.

Lam, X., Gieseke, C., Knoll, M. and Talbot, P. (2000). Assay and importance of adhesive interaction between hamster (Mesocricetus auratus) oocyte-cumulus complexes and the oviductal epithelium. Biology of Reproduction. 62(3):579-88.

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., Weynand, B., Verbeken, E., De Leyn, P., Liston, A., Vansteenkiste, J., Carmeliet, P., Aerts, S. and Thienpont B. (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine. 24(8):1277-1289.

Lee, K. F., Yao, Y. Q., Kwok, K. L., Xu, J. S. and Yeung, W. S. (2002) Early developing embryos affect the gene expression patterns in the mouse oviduct. Biochemical and Biophysical Research Communications. 292(2):564-70.

Lee, S. H., Oh, H. J., Kim, M. J., Setyawan, E. M. N. and Lee, B. C. (2019) Interaction of the EGFR signaling pathway with porcine cumulus oocyte complexes and oviduct cells in a coculture system. Journal of Cellular Physiology. 234(4):4030-4043.

Lee, S. H., Oh, H. J., Kim, M. J. and Lee, B. C. (2020) Exosomes derived from oviduct cells mediate the EGFR/MAPK signaling pathway in cumulus cells. Journal of Cellular Physiology. 235(2):1386-1404.

Leese, H. J. (1988). The formation and function of oviduct fluid. Journal of Reproduction and Fertility. 82(2):843-56.

Leese, H. J., Hugentobler, S. A., Gray, S. M., Morris, D. G., Sturmey, R. G., Whitear, S. L., Sreenan, J. M. (2008). Female reproductive tract fluids: composition, mechanism of formation and potential role in the developmental origins of health and disease. Reproduction Fertility and Development. 20(1):1-8.

Lefebvre, R., DeMott, R. R., Suarez, S. S. and Samper, J. C. (1995). Specific inhibition of equine sperm binding to oviductal epithelium. Biology of Reproduction. 52(monograph_series1):689-696.

Leone, P., Shin, E. C., Perosa, F., Vacca, A., Dammacco, F. and Racanelli, V. (2013). MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. Journal of the National Cancer Institute. 105(16):1172-87.

Li, J., Xing, S., Zhao, G., Zheng, M., Yang, X., Sun, J., Wen, J. and Liu, R. (2020). Identification of diverse cell populations in skeletal muscles and biomarkers for intramuscular fat of chicken by single-cell RNA sequencing. BMC Genomics. 21(1):752.

Lin, P., Bhatnagar, K. P., Nettleton, S. and Nakajima, S. T. (2002). Female genital anomalies affecting reproduction. Fertility and Sterility. 78, 899-915.

Litviňuková, M., Talavera-López, C., Maatz. H, Reichart, D., Worth, C. L., Lindberg, E. L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E. R., Roberts, K., Tuck, L., Fasouli, E. S., DeLaughter, D. M., McDonough, B., Wakimoto, H., Gorham, J. M., Samari, S., Mahbubani, K. T., Saeb-Parsy, K., Patone, G., Boyle, J. J., Zhang, H., Zhang, H., Viveiros, A., Oudit, G. Y., Bayraktar, O. A., Seidman, J. G., Seidman, C. E, Noseda, M., Hubner, N. and Teichmann, S. A. (2020) Cells of the adult human heart. Nature. 588(7838):466-472.

Lonergan, P., Rizos, D., Ward, F. and Boland, M. P. (2001). Factors influencing oocyte and embryo quality in cattle. Reproduction Nutrition Development. 41(5):427-37.

Machado, L., Geara, P., Camps, J., Dos Santos, M., Teixeira-Clerc, F., Van Herck, J., Varet, H., Legendre, R., Pawlotsky, J. M., Sampaolesi, M., Voet, T., Maire, P., Relaix, F. and Mourikis, P. (2021). Tissue damage induces a conserved stress response that initiates quiescent muscle stem cell activation. Cell Stem Cell. S1934-5909(21)00017-5.

Maillo, V., Gaora, P. Ó., Forde, N., Besenfelder, U., Havlicek, V., Burns, G. W., Spencer, T. E., Gutierrez-Adan, A., Lonergan, P. and Rizos, D. (2015). Oviduct-embryo interactions in cattle: two-way traffic or a one-way street? Biology of Reproduction. 92(6):144.

Malarkannan, S. (2020). NKG7 makes a better killer. Nature Immunology. 21(10):1139-1140. Maneschi, F., Zupi, E., Marconi, D., Valli, E., Romanini, C. and Mancuso, S. (1995). Hysteroscopically detected asymptomatic Müllerian anomalies. Prevalence and reproductive implications. Journal of Reproductive Medicine. 40, 684-688.

Mao, S., Shah, A. S., Moninger, T. O., Ostedgaard, L. S., Lu, L., Tang, X. X., Thornell, I. M., Reznikov, L. R., Ernst, S. E., Karp, P. H., Tan, P., Keshavjee, S., Abou Alaiwa, M. H. and Welsh, M. J. (2018). Motile cilia of human airway epithelia contain hedgehog signaling components that mediate noncanonical hedgehog signaling. PNAS. 115(6):1370-1375.

Marey, M. A., Liu, J., Kowsar, R., Haneda, S., Matsui, M., Sasaki, M., Takashi, S., Hayakawa, H., Wijayagunawardane, M. P., Hussein, F. M. and Miyamoto, A. (2013). Bovine oviduct epithelial cells downregulate phagocytosis of sperm by neutrophils: prostaglandin E2 as a major physiological regulator. Reproduction. 147(2):211-9.

Marey, M. A., Matsukawa, H., Sasaki, M., Ezz, M. A., Yousef, M. S., Takahashi, K. I. and Miyamoto, A. (2020). Bovine oviduct epithelial cells suppress the phagocytic activity of neutrophils towards sperm but not for bacteria in vitro: Immunofluorescence and electron microscopic observations. Histology and Histopathology. 35(6):589-597.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research. 18(9):1509-17.

Marquardt, N., Kekäläinen, E., Chen, P., Lourda, M., Wilson, J. N., Scharenberg, M., Bergman, P., Al-Ameri, M., Hård, J., Mold, J. E., Ljunggren, H. G. and Michaëlsson, J. (2019). Unique transcriptional and protein-expression signature in human lung tissue-resident NK cells. Nature Communications. 10(1):3841.

Martínez-Pastor, F., Tizado, E. J., Garde, J. J., Anel, L. and de Paz, P. (2011). Statistical Series: Opportunities and challenges of sperm motility subpopulation analysis. Theriogenology. 75(5):783-95.

Massagué, J. (2012). TGFβ signalling in context. Nature Reviews Molecular Cell Biology. 13(10):616-30.

McComb, P., Langley, L., Villalon, M. and Verdugo, P. (1986) The oviductal cilia and Kartagener's syndrome. Fertility and Sterility. 46(3):412-6.

McGinnis, C. S., Murrow, L. M. and Gartner, Z. J. (2019a). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell

Systems. 8(4):329-337.e4. McGinnis, C. S., Patterson, D. M., Winkler, J., Conrad, D. N., Hein, M. Y., Srivastava, V., Hu, J. L., Murrow, L. M., Weissman, J. S., Werb, Z., Chow, E. D. and Gartner, Z. J. (2019b). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nature Methods. 16(7):619-626.

Merle, N. S., Church, S. E., Fremeaux-Bacchi, V. and Roumenina, L. T. (2015). Complement System Part I - Molecular Mechanisms of Activation and Regulation. Frontiers in Immunology. 6:262.

Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. and Roumenina, L. T. (2015). Complement System Part II: Role in Immunity. Frontiers in Immunology. 6:257.

Michalska, A., Blaszczyk, K., Wesoly, J. and Bluyssen, H. A. R. (2018). A positive feedback amplifier circuit that regulates interferon (IFN)-stimulated gene expression and controls type I and type II IFN responses. Frontiers in Immunology. 9:1135.

Miki, K. and Clapham, D. E. (2013). Rheotaxis guides mammalian sperm. Current Biology. 23(6):443-52.

Mitoyan, L., Chevrier, V., Hernandez-Vargas, H., Ollivier, A., Homayed, Z., Pannequin, J., Poizat, F., De Biasi-Cador, C., Charafe-Jauffret, E., Ginestier, C. and Guasch, G. (2021) A stem cell population at the anorectal junction maintains homeostasis and participates in tissue regeneration. Nature Communications. 12(1):2761.

Mkaddem, S. B., Murua, A., Flament, H., Titeca-Beauport, D., Bounaix, C., Danelli, L., Launay, P., Benhamou, M., Blank, U., Daugas, E., Charles, N. and Monteiro, R. C. (2017). Lyn and Fyn function as molecular switches that control immunoreceptors to direct homeostasis or inflammation. Nature Communications. 8(1):246.

Mogilenko, D. A., Shpynov, O., Andhey, P. S., Arthur, L., Swain, A., Esaulova, E., Brioschi, S., Shchukina, I., Kerndl, M., Bambouskova, M., Yao, Z., Laha, A., Zaitsev, K., Burdess, S., Gillfilan, S., Stewart, S. A., Colonna, M., Artyomov, M. N. (2021). Comprehensive profiling of an aging immune system reveals clonal GZMK+ CD8+ T cells as conserved hallmark of inflammaging. Immunity. 54(1):99-115.e12.

Morcel, K., Camborieux, L. and Guerrier, D. (2007). Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome. Orphanet Journal of Rare Diseases. 2, 13.

Moro, K., Yamada, T., Tanabe, M., Takeuchi, T., Ikawa, T., Kawamoto, H., Furusawa, J., Ohtani, M., Fujii, H. and Koyasu, S. (2010). Innate production of T(H)2 cytokines by adipose tissue-associated c-Kit(+)Sca-1(+) lymphoid cells. Nature. 463(7280):540-4.

Mullen, R. D. and Behringer, R. R. (2014). Molecular genetics of Müllerian duct formation, regression and differentiation. Sexual Development. 8(5), 281-296.

Muramatsu, H., Zou, P., Kurosawa, N., Ichihara-Tanaka, K., Maruyama, K., Inoh, K., Sakai, T., Chen, L., Sato, M. and Muramatsu, T. (2006). Female infertility in mice deficient in midkine and pleiotrophin, which form a distinct family of growth factors. Genes Cells. 11(12):1405-17.

Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. and Desai, T. J.

(2018). Single-cell Wnt signaling niches maintain stemness of alveolar type2 cells. Science. 359(6380):1118-1123.

Narayana, S. K., Helbig, K. J., McCartney, E. M., Eyre, N. S., Bull, R. A., Eltahla, A., Lloyd, A. R. and Beard, M. R. (2015). The interferon-induced transmembrane proteins, IFITM1, IFITM2, and IFITM3 inhibit hepatitis C virus entry. The Journal of Biological Chemistry. 290(43):25946-59.

Nathan, A., Reinhardt, P., Kruspe, D., Jörß, T., Groth, M., Nolte, H., Habenicht, A., Herrmann, J., Holschbach, V., Toth, B., Krüger, M., Wang, Z. Q., Platzer, M. and Englert, C. (2017). The Wilms tumor protein Wt1 contributes to female fertility by regulating oviductal proteostasis. Human Molecular Genetics. 26(9):1694-1705.

Neefjes, J., Jongsma, M. L., Paul, P. and Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology. 11(12):823-36.

Neill, D. R., Wong, S. H., Bellosi, A., Flynn, R. J., Daly, M., Langford, T. K., Bucks, C., Kane, C. M., Fallon, P. G., Pannell, R., Jolin, H. E. and McKenzie, A. N. (2010). Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature. 464(7293):1367-70.

Ner-Gaon, H., Melchior, A., Golan, N., Ben-Haim, Y. and Shay, T. (2017). JingleBells: A repository of immune-related single-cell RNA-sequencing datasets. Journal of Immunology. 198(9):3375-3379

Neumann, H., Schmidt, H., Cavalié, A., Jenne, D. and Wekerle, H. (1997). Major histocompatibility complex (MHC) class I gene expression in single neurons of the central nervous system: differential regulation by interferon (IFN)-gamma and tumor necrosis factor (TNF)-alpha. The Journal of Experimental Medicine. 185(2):305-16.

Ng, A. and Barker N. (2015). Ovary and fimbrial stem cells: biology, niche and cancer origins. Nature Reviews Molecular Cell Biology. 16(10):625-38.

Nguyen, Q. H., Pervolarakis, N., Nee, K. and Kessenbrock, K. (2018). Experimental considerations for single-cell RNA sequencing approaches. Frontiers in Cell and Developmental Biology. 6:108.

Nie, G. Y., Li, Y., Wang, J., Minoura, H., Findlay, J. K. and Salamonsen, L. A. (2000). Complex regulation of calcium-binding protein D9k (calbindin-D(9k)) in the mouse uterus during early pregnancy and at the site of embryo implantation. Biology of Reproduction. 62(1):27-36.

Norwood, J. T., Hein, C. E., Halbert, S. A. and Anderson, R. G. (1978) Polycationic macromolecules inhibit cilia-mediated ovum transport in the rabbit oviduct. PNAS. 75(9):4413-6.

O'Doherty, A. M., Di Fenza, M. and Kölle, S. (2016). Lipopolysaccharide (LPS) disrupts particle transport, cilia function and sperm motility in an ex vivo oviduct model. Scientific Reports. 6:24583.

O'Flanagan, C. H., Campbell, K. R., Zhang, A. W., Kabeer, F., Lim, J. L. P., Biele, J., Eirew, P., Lai, D., McPherson, A., Kong, E., Bates, C., Borkowski, K., Wiens, M., Hewitson, B., Hopkins, J., Pham, J., Ceglia, N., Moore, R., Mungall, A. J.,

McAlpine, J. N., CRUK IMAXT Grand Challenge Team, Shah, S. P. and Aparicio, S. (2019). Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. Genome Biology. 20(1):210.

Ohtsuka, M., Inoko, H., Kulski, J. K. and Yoshimura, S. (2008). Major histocompatibility complex (Mhc) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. BMC Genomics. 9:178.

Orozco, R. C., Marquardt, K., Mowen, K. A., Pratumchai, I., Teijaro, J. R. and Sherman. L. A. (2020). The Pro-autoimmune allele of tyrosine phosphatase PTPN22 enhances anti-viral and anti-tumor immunity. The Journal of Immunology. 204 (1 Supplement) 247.3.

Ortiz, M. E., Llados, C. and Croxatto, H. B. (1989). Embryos of different ages transferred to the rat oviduct enter the uterus at different times. Biology of Reproduction. 41(3):381-4.

Owhor, L. E., Reese, S. and Kölle, S. (2019). Salpingitis impairs bovine tubal function and sperm-oviduct interaction. Scientific Reports. 9(1):10893.

Pacey, A. A., Hill, C. J., Scudamore, I. W., Warren, M. A., Barratt, C. L. and Cooke, I. D. (1995). The interaction in vitro of human spermatozoa with epithelial cells from the human uterine (fallopian) tube. Human Reproduction. 10(2):360-6.

Paik, D. Y., Janzen, D. M., Schafenacker, A. M., Velasco, V. S., Shung, M. S., Cheng, D., Huang, J., Witte, O. N. and Memarzadeh, S. (2012). Stem-like epithelial cells are concentrated in the distal end of the fallopian tube: a site for injury and serous cancer initiation. Stem Cells. 30(11):2487-97.

Panchision, D. M., Chen, H. L., Pistollato, F., Papini, D., Ni, H. T. and Hawley, T. S. (2007). Optimized flow cytometric analysis of central nervous system tissue reveals novel functional relationships among cells expressing CD133, CD15, and CD24. Stem Cells. 25(6):1560-70.

Panda, D., Gjinaj, E., Bachu, M., Squire, E., Novatt, H., Ozato, K. and Rabin, R. L. (2019). IRF1 maintains optimal constitutive expression of antiviral genes and regulates the early antiviral response. Frontiers in Immunology. 10:1019.

Paris, E. A., Bahr, J. M., Bitterman, P., Basu, S., Abramowicz, J. S. and Barua, A. (2021). Incidence of malignant transformation in the oviductal fimbria in laying hens, a preclinical model of spontaneous ovarian cancer. PLoS One. 16(7):e0255007.

Pastva, A. M., Wright, J. R. and Williams, K. L. (2007). Immunomodulatory roles of surfactant proteins A and D: implications in lung disease. Proceedings of the American Thoracic Society. 4(3):252-7.

Paula-Lopes, F. F., Boelhauve, M., Habermann, F. A., Sinowatz, F. and Wolf, E. (2007) Leptin promotes meiotic progression and developmental capacity of bovine oocytes via cumulus cell-independent and -dependent mechanisms. Biology of Reproduction. 76(3):532-41.

Perez-Cerezales, S., Boryshpolets, S. and Eisenbach, M. (2015). Behavioral mechanisms of mammalian sperm guidance. Asian Journal of Andrology. 17(4):628-32.

Perets, R., Wyant, G. A., Muto, K. W., Bijron, J. G., Poole, B. B., Chin, K. T.,

Chen, J. Y., Ohman, A. W., Stepule, C. D., Kwak, S., Karst, A. M., Hirsch, M. S., Setlur, S. R., Crum, C. P., Dinulescu, D. M. and Drapkin, R. (2013). Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. Cancer Cell. 24(6):751-65.

Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S. and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols. 9(1):171-81. Platanias, L. C. (2005). Mechanisms of type-I- and type-II-interferon-mediated signalling. Nature Reviews Immunology. 5(5):375-86.

Puga Molina, L. del C., Luque, G. M., Balestrini, P. A., Marín-Briggiler, C. I., Romarowski, A. and Buffone, M. G. (2018) Molecular basis of human sperm capacitation. Frontiers in Cell and Developmental Biology. 6:72.

Qiao, B., Wang, Q., Zhao, Y. and Wu, J. (2020). miR-205-3p Functions as a Tumor Suppressor in Ovarian Carcinoma. Reproductive Sciences. 27(1):380-388.

Quill, T. A., Sugden, S. A., Rossi, K. L., Doolittle, L. K., Hammer, R. E. and Garbers, D. L. (2003). Hyperactivated sperm motility driven by CatSper2 is required for fertilization. PNAS. 100(25):14869-74.

Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. and Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. Trends in Genetics. 36(4):298-311.

Rana, P., Kazmi, I., Singh, R., Afzal, M., Al-Abbasi, F. A., Aseeri, A., Singh, R., Khan, R. and Anwar, F. (2013). Ectopic pregnancy: a review. Archives of Gynecology and Obstetrics. 288(4):747-57.

Reeve, L., Lashen, H. and Pacey, A. A. (2005) Endometriosis affects sperm-endosalpingeal interactions. Human Reproduction. 20(2):448-51.

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., Hacohen, N., Haniffa, M., Hemberg, M., Kim, S., Klenerman, P., Kriegstein, A., Lein, E., Linnarsson, S., Lundberg, E., Lundeberg, J., Majumder, P., Marioni, J. C., Merad, M., Mhlanga, M., Nawijn, M., Netea, M., Nolan, G., Pe'er, D., Phillipakis, A., Ponting, C. P., Quake, S., Reik, W., Rozenblatt-Rosen, O., Sanes, J., Satija, R., Schumacher, T. N. , Shalek, A., Shapiro, E., Sharma, P., Shin, J. W., Stegle, O., Stratton, M., Stubbington, M. J. T., Theis, F. J., Uhlen, M., van Oudenaarden, A., Wagner, A., Watt, F., Weissman, J., Wold, B., Xavier, R., Yosef, N.; Human Cell Atlas Meeting Participants. (2017). The Human Cell Atlas. Elife. 6:e27041.

Reichard, A. and Asosingh, K. (2019). Best Practices for Preparing a Single Cell Suspension from Solid Tissues for Flow Cytometry. Cytometry A. 95(2):219-226.

Reshi, Q. U. A., Viil, J., Ord, J., Lättekivi, F., Godakumara, K., Hasan, M. M., Nõmm, M., Jääger, K., Velthut-Meikas, A., Jaakma, Ü., Salumets, A. and Fazeli, A. (2020) Spermatozoa induce transcriptomic alterations in bovine oviductal epithelial cells prior to initial contact. Journal of Cell Communication and Signaling. 14(4):439-451.

Ren, D., Navarro, B., Perez, G., Jackson, A. C., Hsu, S., Shi, Q., Tilly, J. L. and Clapham, D. E. (2001). A sperm ion channel required for sperm motility and male

fertility. Nature. 413(6856):603-9.

Ribas-Maynou, J., García-Peiró, A., Abad, C., Amengual, M. J., Navarro, J. and Benet, J. (2012). Alkaline and neutral Comet assay profiles of sperm DNA damage in clinical groups. Human Reproduction. 27(3):652-8.

Riggs, M. J., Lin, N., Wang, C., Piecoro, D. W., Miller, R. W., Hampton, O. A., Rao, M., Ueland, F. R. and Kolesar, J. M. (2020). DACH1 mutation frequency in endometrial cancer is associated with high tumor mutation burden. PLoS One. 15(12):e0244558.

Ríos, M., Parada-Bustamante, A., Velásquez, L. A., Croxatto, H. B. and Orihuela, P. A. (2011). Participation of the oviductal s100 calcium binding protein G in the genomic effect of estradiol that accelerates oviductal embryo transport in mated rats. Reproductive Biology and Endocrinology. 9:69.

Rock, J. A. and Schlaff, W. D. (1985). The obstetric consequences of uterovaginal anomalies. Fertility and Sterility. 43, 681-692.

Rodriguez-Martinez, H., Nicander, L., Viring, S., Einarsson, S. and Larsson, K. (1990). Ultrastructure of the uterotubal junction in preovulatory pigs. Anatomy, Histology and Embryology. 19: 16-3.

Roly, Z. Y., Backhouse, B., Cutting, A., Tan, T. Y., Sinclair, A. H., Ayers, K. L., Major, A. T. Smith, C. A. (2018). The cell biology and molecular genetics of Müllerian duct development. Wiley Interdisciplinary Reviews Developmental Biology. 7(3):e310.

Roly, Z. Y., Godini, R., Estermann, M. A., Major, A. T., Pocock, R.,and Smith, C. A. (2020a). Transcriptional landscape of the embryonic chicken Müllerian duct. BMC Genomics. 21(1):688.

Roly, Z. Y., Major, A. T., Fulcher, A., Estermann, M. A., Hirst, C. E., and Smith, C. A. (2020b). Adhesion G-protein-coupled receptor, GPR56, is required for Müllerian duct development in the chick. The Journal of Endocrinology. 244(2):395-413.

Rosa, A., Chande, A., Ziglio, S., De Sanctis, V., Bertorelli, R., Goh, S. L., McCauley, S. M., Nowosielska, A., Antonarakis, S. E., Luban, J., Santoni, F. A. and Pizzato, M. (2015). HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. Nature. 526(7572):212-7.

Rydell-Törmänen, K. and Johnson, J. R. (2019). The applicability of mouse models to the study of human disease. Methods in Molecular Biology.1940:3-22.

Saatcioglu, H. D., Kano, M., Horn, H., Zhang, L., Samore, W., Nagykery, N., Meinsohn, M. C., Hyun, M., Suliman, R., Poulo, J., Hsu, J., Sacha, C., Wang, D., Gao, G., Lage, K., Oliva, E., Morris Sabatini, M. E., Donahoe, P. K. and Pépin, D. (2019). Single-cell sequencing of neonatal uterus reveals an Misr2+ endometrial progenitor indispensable for fertility. Elife. 8:e46349.

Saccary, L., She, Y. M., Oko, R. and Kan, F. W. (2013). Hamster oviductin regulates tyrosine phosphorylation of sperm proteins during in vitro capacitation. Biology of Reproduction. 89:38.

Salhab, M., Tosca, L., Cabau, C., Papillier, P., Perreau, C., Dupont, J., Mermillod, P. and Uzbekova, S. (2011). Kinetics of gene expression and signaling in bovine

cumulus cells throughout IVM in different mediums in relation to oocyte developmental competence, cumulus apoptosis and progesterone secretion. Theriogenology. 75(1):90-104.

Salhab, M., Dhorne-Pollet, S., Auclair, S., Guyader-Joly, C., Brisard, D., Dalbies-Tran, R., Dupont, J., Ponsart, C., Mermillod, P. and Uzbekova, S. (2013). In vitro maturation of oocytes alters gene expression and signaling pathways in bovine cumulus cells. Molecular Reproduction and Development. 80(2):166-82.

Santana Gonzalez, L., Rota, I. A., Artibani, M., Morotti, M., Hu, Z., Wietek, N., Alsaadi, A., Albukhari, A., Sauka-Spengler, T. and Ahmed, A. A. (2021) Mechanistic drivers of Müllerian duct development and differentiation into the oviduct. Frontiers in Cell and Developmental Biology. 9:605301.

Santana González, L., Artibani, M. and Ahmed, A. A. Studying Müllerian duct anomalies - from cataloguing phenotypes to discovering causation. Diseases Models Mechanisms. 1;14(6):dmm047977.

Saravelos, S. H., Cocksedge, K. A. and Li, T. C. (2008). Prevalence and diagnosis of congenital uterine anomalies in women with reproductive failure: a critical appraisal. Human Reproduction Update. 14, 415-429.

Schmaltz-Panneau, B., Cordova, A., Dhorne-Pollet, S., Hennequet-Antier, C., Uzbekova, S., Martinot, E., Doret, S., Martin, P., Mermillod, P. and Locatelli, Y. (2014). Early bovine embryos regulate oviduct epithelial cell gene expression during in vitro co-culture. Animal Reproduction Science. 149:103–116.

Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S. E., Pollard, J. W., Frampton, J., Liu, K. J. and Geissmann, F. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science. 336(6077):86-90.

Sehdev, A. S., Kurman, R. J., Kuhn, E. and Shih, IeM. (2010). Serous tubal intraepithelial carcinoma upregulates markers associated with high-grade serous carcinomas including Rsf-1 (HBXAP), cyclin E and fatty acid synthase. Modern Pathology. 23(6):844-55.

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P., Gertner, R. S., Gaublomme, J. T., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X., Ding, R., Raychowdhury, R., Friedman, N., Hacohen, N., Park, H., May, A. P. and Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature. 510(7505):363-9.

Shi, G., Kenney, A. D., Kudryashova, E., Zani, A., Zhang, L., Lai, K. K., Hall-Stoodley, L., Robinson, R. T., Kudryashov, D. S., Compton, A. A. and Yount, J. S. (2021). Opposing activities of IFITM proteins in SARS-CoV-2 infection. The EMBO Journal. 40(3):e106501.

Sinha, M. and Lowell, C. A. Immune defense protein expression in highly purified mouse lung epithelial cells. (2016). American Journal of Respiratory Cell and Molecular Biology. 54(6):802-13.

Smith, T. T. and Yanagimachi, R. (1991) Attachment and release of spermatozoa from the caudal isthmus of the hamster oviduct. Journal of Reproduction and Fertility. 91(2):567-73.

Smits, K., De Coninck, D. I., Van Nieuwerburgh, F., Govaere, J., Van Poucke, M., Peelman, L., Deforce, D., Van Soom, A. (2016). The equine embryo influences immune-related gene expression in the oviduct. Biology of Reproduction. 94(2):36.

Song, P. I., Abraham, T. A., Park, Y., Zivony, A. S., Harten, B., Edelhauser, H. F., Ward, S. L., Armstrong, C. A. and Ansel, J. C. (2001). The expression of functional LPS receptor proteins CD14 and toll-like receptor 4 in human corneal cells. Investigative Ophthalmology Visual Science. 42(12):2867-77.

Song, Z. H., Li, Z. Y., Li, D. D., Fang, W. N., Liu, H. Y., Yang, D. D., Meng, C. Y., Yang, Y. and Peng, J. P. (2016). Seminal plasma induces inflammation in the uterus through the γδ T/IL-17 pathway. Scientific Reports. 6:25118.

Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J. A., Wetzel, C. H., Zimmer, R. K. and Hatt, H. (2003). Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science. 299(5615):2054-8.

Spilman, C. H. and Harper, M. J. (1975). Effects of prostaglandins on oviductal motility and egg transport. Gynecologic and Obstetric Investigation. 6(3-4):186-205.

Stillman, R. J., Rosenberg, M. J. and Sachs, B. P. (1986). Smoking and reproduction. Fertility and Sterility. 46(4):545-66.

Stival, C., Puga Molina, L. del C., Paudel, B., Buffone, M. G., Visconti, P. E. and Krapf, D. (2016). Sperm capacitation and acrosome reaction in mammalian sperm. Advances in Anatomy, Embryology and Cell Biology. 220:93-106.

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B. Z., Mauck, W. M. 3rd, Smibert, P. and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biology. 19(1):224.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. 3rd, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019). Comprehensive integration of single-cell data. Cell. 177(7):1888-1902.e21.

Suarez, S. S. (1987). Sperm transport and motility in the mouse oviduct: observations in situ. Biology of Reproduction. 36(1):203-10.

Suarez, S., Redfern, K., Raynor, P., Martin, F., Phillips, D. M. (1991). Attachment of boar sperm to mucosal explants of oviduct in vitro: possible role in formation of a sperm reservoir. Biology of Reproduction. 44(6):998-1004.

Sun, F., Bahat, A., Gakamsky, A., Girsh, E., Katz, N., Giojalas, L. C., Tur-Kaspa, I. and Eisenbach, M. (2005). Human sperm chemotaxis: both the oocyte and its surrounding cumulus cells secrete sperm chemoattractants. Human Reproduction. 20(3):761-7.

Svensson, V., da Veiga Beltrame, E. and Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. Database (Oxford). 2020:baaa073.

Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators. (2018). Single-cell transcriptomics of 20 mouse organs cre-

ates a Tabula Muris. Nature. 562(7727):367-372.

Talbot, P., Geiske, C. and Knoll, M. (1999). Oocyte pickup by the mammalian oviduct. Molecular Biology of the Cell. 10(1):5-8.

Talukder, A. K., Rashid, M. B., Yousef, M. S., Kusama, K., Shimizu, T., Shimada, M., Suarez, S. S., Imakawa, K. and Miyamoto, A. (2018). Oviduct epithelium induces interferon-tau in bovine day-4 embryos, which generates an anti-inflammatory response in immune cells. Scientific Reports. 8(1):7850.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K. and Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods. 6(5):377-82.

Tang, X., Huang, Y., Lei, J., Luo, H. and Zhu, X. (2019). The single-cell sequencing: new developments and medical applications. Cell & Bioscience. 9:53.

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., Hawrylycz, M., Koch, C. and Zeng, H. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nature Neuroscience. 19(2):335-46.

Tachtsidis, A., McInnes, L. M., Jacobsen, N., Thompson, E. W. and Saunders, C. M. (2016). Minimal residual disease in breast cancer: an overview of circulating and disseminated tumour cells. Clinical Experimental Metastasis. 33(6):521-50.

Tatsumi, K., Higuchi, T., Fujiwara, H., Nakayama, T., Itoh, K., Mori, T., Fujii, S. and Fujita, J. (1999). Expression of calcium binding protein D-9k messenger RNA in the mouse uterine endometrium during implantation. Molecular Human Reproduction. 5(2):153-61.

Teodosio, C., Mayado, A., Sánchez-Muñoz, L., Morgado, J. M., Jara-Acevedo, M., Álvarez-Twose, I., García-Montero, A. C., Matito, A., Caldas, C., Escribano, L. and Orfao, A. (2015). The immunophenotype of mast cells and its utility in the diagnostic work-up of systemic mastocytosis. Journal of Leukocyte Biology. 97(1):49-59.

Teves, M. E., Barbano, F., Guidobaldi, H. A., Sanchez, R., Miska, W., Giojalas, L. C. (2006). Progesterone at the picomolar range is a chemoattractant for mammalian spermatozoa. Fertility and Sterility. 86:745–749.

Theocharis, A. D., Skandalis, S. S., Gialeli, C. and Karamanos, N. K. (2016). Extracellular matrix structure. Advanced Drug Delivery Reviews. 97:4-27.

Tiwari, R., de la Torre, J. C., McGavern, D. B. and Nayak, D. (2019). Beyond tethering the viral particles: immunomodulatory functions of tetherin (BST-2). DNA and Cell Biology. 38(11):1170-1177.

Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology. 19(1A):A68-77.

Torres-Poveda, K., Bahena-Román, M., Madrid-González, C., Burguete-García, A. I., Bermúdez-Morales, V. H., Peralta-Zaragoza, O. and Madrid-Marina, V. (2014). Role of IL-10 and TGF-β1 in local immunosuppression in HPV-associated cervical neoplasia. World Journal of Clinical Oncology. 5: 753-763.

Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., Johnson, D. S., Trivett, M. K., Etemadmoghadam, D., Locandro, B., Traficante, N., Fereday, S., Hung, J. A., Chiew, Y. E., Haviv, I.; Australian Ovarian Cancer Study Group, Gertig, D., DeFazio, A. and Bowtell, D. D. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clinical Cancer Research. 14(16):5198-208.

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A. and Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature. 509(7500):371-5.

Tuffrey, M., Alexander, F., Inman, C., Ward and M. E. (1990). Correlation of infertility with altered tubal morphology and function in mice with salpingitis induced by a human genital-tract isolate of Chlamydia trachomatis. Journal of Reproduction and Fertility. 88(1):295-305.

van de Leemput, E. E., Vos, P. L., Zeinstra, E. C., Bevers, M. M., van der Weijden, G. C. and Dieleman, S. J. (1999). Improved in vitro embryo development using in vivo matured oocytes from heifers superovulated with a controlled preovulatory LH surge. Theriogenology. 15;52(2):335-49.

van den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., Robin, C. and van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nature Methods. 14(10):935-936.

van der Putten, M. E., Engel, M. and van Well, G. T. (2008). Salpingitis. A rare cause of acute abdomen in a sexually inactive girl: a case report. Cases Journal. 1(1):326.

Vanacker, J., Camboni, A., Dath, C., Van Langendonckt, A., Dolmans, M. M., Donnez, J., Amorim, C. A. (2011). Enzymatic isolation of human primordial and primary ovarian follicles with Liberase DH: protocol for application in a clinical setting. Fertility and Sterility. 96(2):379-383.e3.

Velliyagounder, K., Kaplan, J. B., Furgang, D., Legarda, D., Diamond, G., Parkin, R. E. and Fine, D. H. (2003). One of two human lactoferrin variants exhibits increased antibacterial and transcriptional activation activities and is associated with localized juvenile periodontitis. Infection and Immunity. 71(11):6141-7.

Vembadi, A., Menachery, A. and Qasaimeh, M. A. (2019). Cell Cytometry: Review and Perspective on Biotechnological Advances. Frontiers in Bioengineering and Biotechnology. 7:147.

Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J. E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., Payne, R. P., Ivarsson, M. A., Lisgo, S., Filby, A., Rowitch, D. H., Bulmer, J. N., Wright, G. J., Stubbington, M. J. T., Haniffa, M., Moffett, A. and Teichmann, S. A. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature. 563(7731):347-353.

Vermeren, S., Miles, K., Chu, J. Y., Salter, D., Zamoyska, R. and Gray, M. (2016).

PTPN22 is a critical regulator of Fcγ receptor-mediated neutrophil activation. Journal of Immunology. 197(12):4771-4779.

Vrtačnik, P., Kos, Š., Bustin, S. A., Marc, J. and Ostanek, B. (2014). Influence of trypsinization and alternative procedures for cell preparation before RNA extraction on RNA integrity. Analytical Biochemistry. 463:38-44.

Wagner, A., Ekhlasi-Hundrieser, M., Hettel, C., Petrunkina, A., Waberski, D., Nimtz, M. and Töpfer-Petersen, E. (2002). Carbohydrate-based interactions of oviductal sperm reservoir formation-studies in the pig. Molecular Reproduction and Development. 61(2):249-57.

Wagner, A, Regev, A and Yosef N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nature Biotechnology. 34(11):1145-1160.

Walko, G., Castañón, M. J., Wiche, G. (2015). Molecular architecture and function of the hemidesmosome. Cell and Tissue Research. 360(2):363-78.

Walter, J., Monthoux, C., Fortes, C., Grossmann, J., Roschitzki, B., Meili, T., Riond, B., Hofmann-Lehmann, R., Naegeli, H. and Bleul, U. (2020) The bovine cumulus proteome is influenced by maturation condition and maturational competence of the oocyte. Scientific Reports. 10(1):9880.

Wang, H., Guo, Y., Wang, D., Kingsley, P. J., Marnett, L. J., Das, S. K., DuBois, R. N. and Dey, S. K. (2004). Aberrant cannabinoid signaling impairs oviductal transport of embryos. Nature Medicine. 10(10):1074-80.

Wang, L., Feng, Z., Wang, X., Wang, X. and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 26(1):136-8.

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C. and Macosko, E. Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell. 177(7):1873-1887.e17.

Wijayagunawardane, M. P., Miyamoto, A., Taquahashi, Y., Acosta, T. J., Nishimura, M. and Sato, K. (2001). Angiotensin II and atrial natriuretic peptide in the cow oviductal contraction in vitro: direct effect and local secretion of prostaglandins, endothelin-1, and angiotensin II. Biology of Reproduction. 65(3):799-804.

Winuthayanon, W., Bernhardt, M. L., Padilla-Banks, E., Myers, P. H., Edin, M. L., Lih, F. B., Hewitt, S. C., Korach, K. S. and Williams, C. J. (2015). Oviductal estrogen receptor α signaling prevents protease-mediated embryo death. Elife. 4:e10453.

Wira, C. R., Fahey, J. V., Sentman, C. L., Pioli, P. A. and Shen, L. (2005). Innate and adaptive immunity in female genital tract: cellular responses and interactions. Immunological Reviews. 206:306-35.

Wong, S. H., Walker, J. A., Jolin, H. E., Drynan, L. F., Hams, E., Camelo, A., Barlow, J. L., Neill, D. R., Panova, V., Koch, U., Radtke, F., Hardman, C. S., Hwang, Y. Y., Fallon, P. G. and McKenzie, A. N. (2012). Transcription factor RORα is critical for nuocyte development. Nature Immunology. 13(3):229-36.

Wright, C. A., Kozik, P., Zacharias, M. and Springer, S. (2004). Tapasin and other

chaperones: models of the MHC class I loading complex. Biological Chemistry. 385(9):763-78.

Ximerakis, M., Lipnick, S. L., Innes, B. T., Simmons, S. K., Adiconis, X., Dionne, D., Mayweather, B. A., Nguyen, L., Niziolek, Z., Ozek, C., Butty, V. L., Isserlin, R., Buchanan, S. M., Levine, S. S., Regev, A., Bader, G. D., Levin, J. Z. and Rubin, L. L. (2019). Single-cell transcriptomic profiling of the aging mouse brain. Nature Neuroscience. 22(10):1696-1708.

Xu, H., Yu, S., Yuan, X., Xiong, J., Kuang, D., Pestell, R. G. and Wu, K. (2017). DACH1 suppresses breast cancer as a negative regulator of CD44. Scientific Reports. 7(1):4361.

Yagi, R., Zhong, C., Northrup, D. L., Yu, F., Bouladoux, N., Spencer, S., Hu, G., Barron, L., Sharma, S., Nakayama, T., Belkaid, Y., Zhao, K. and Zhu, J. (2014). The transcription factor GATA3 is critical for the development of all IL-7Rα-expressing innate lymphoid cells. Immunity. 40(3):378-88.

Yamane, D., Feng, H., Rivera-Serrano, E. E., Selitsky, S. R., Hirai-Yuki, A., Das, A., McKnight, K. L., Misumi, I., Hensley, L., Lovell, W., González-López, O., Suzuki, R., Matsuda, M., Nakanishi, H., Ohto-Nakanishi, T., Hishiki, T., Wauthier, E., Oikawa, T., Morita, K., Reid, L. M., Sethupathy, P., Kohara, M., Whitmire, J. K. and Lemon, S. M. (2019). Basal expression of interferon regulatory factor 1 drives intrinsic hepatocyte resistance to multiple RNA viruses. Nature Microbiology. 4(7):1096-1104.

Yamanouchi, H., Umezu, T., and Tomooka, Y. (2010). Reconstruction of oviduct and demonstration of epithelial fate determination in mice. Biology of Reproduction. 82(3):528-33.

Yeste, M., Jones, C., Amdani, S. N. and Coward, K. (2017). Oocyte activation and fertilisation: crucial contributors from the sperm and oocyte. Results and Problems in Cell Differentiation. 59:213-239.

Yona, S., Kim, K. W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A., Hume, D. A., Perlman, H., Malissen, B., Zelzer, E. and Jung, S. (2013). Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. Immunity. 38(1):79-91,

Yousef, M. S., Marey, M. A., Hambruch, N., Hayakawa, H., Shimizu, T., Hussien, H. A., Abdel-Razek, A. K., Pfarrer, C. and Miyamoto, A. (2016). Sperm binding to oviduct epithelial cells enhances TGFβ1 and IL10 expressions in epithelial cells as well as neutrophils in vitro: prostaglandin E2 as a main regulator of anti-inflammatory response in the bovine oviduct. PLoS One. 11(9):e0162309.

Zanoni, I. and Granucci, F. (2013). Role of CD14 in host protection against infections and in metabolism regulation. Frontiers in Cellular and Infection Microbiology. 3:32.

Zhao, Y., Yang, X., Jia, Z., Reid, R. L., Leclerc, P. and Kan, F. W. (2016). Recombinant human oviductin regulates protein tyrosine phosphorylation and acrosome reaction. Reproduction. 152:561–573.

Zhao, F. and Yao, H. H. (2019). A tale of two tracts: history, current advances, and future directions of research on sexual differentiation of reproductive tracts. Biology

of Reproduction. 101(3), 602-616.

Zhu, J., Ghosh, A. and Sarkar, S. N. (2015). OASL-a new player in controlling antiviral innate immunity. Current Opinion in Virology (COVIRO). 12:15-9.

Zhu, M., Iwano, T. and Takeda, S. (2019). Estrogen and EGFR pathways regulate Notch signaling in opposing directions for multi-ciliogenesis in the fallopian tube. Cells. 8(8):933.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Molecular Cell. 65(4):631-643.e4.

Zihni, C., Mills, C., Matter, K., Balda, M. S. (2016). Tight junctions: from simple barriers to multifunctional molecular gates. Nature Reviews Molecular Cell Biology. 17(9):564-80.