UNIVERSITY OF BIRMINGHAM

Single-cell transcriptome analysis: Computational study of Fallopian Tube Epithelium from Ovarian and Endometrial Cancers

1643630

Year 3 Biomedical Science Research Project

March 2019

Project Supervisor:

Dr Christopher Yau

Acknowledgements

Acknowledgements for my supervisor Dr Christopher Yau for providing

Help and guidance with Rstudio coding and with the whole project

TABLE OF CONTENTS

ABSTRACT

Background

Ovarian cancers (OCs) is one of the leading causes of death from gynaecological malignancies. The mortality of this disease is due to the late diagnosis when OCs are highly advanced. The evidence suggest that secretory cells from the fallopian tube are the site of origin for Epithelial Ovarian Cancers (EOCs).

Aim

Analyse the transcriptome profile of Fallopian Tube Epithelial (FTE) cells from ovarian cancers and endometrial cancers, whilst observing the changes in transcriptome cause by cryopreservation and cell culturing.

Methods

An observational study of the data acquire from a single cell RNA sequencing assay of FTE cells from OC and Endometrial cancer patients. Using RStudio software, the data was visualised through the generation of gene expression plots, gene correlation plots and heatmaps via coding. Wilcoxon signed rank test, welch t-tests and spearman correlation tests were implemented to determine the significance of findings.

Results

Out of all preservation methods, cryopreserved cells presented the highest level of heterogeneity when compared to fresh cells. Cell markers for ciliated FTE cells (CCDC78 and CCDC17) are scarce in 2-day and in 6-day cultured cells. Expression for cell cycle associated genes were stable in all preservation methods. Expression of TP53 gene was

widely expressed across patients, whereas SOX2 overlapping transcript was found scarce in FTE cells.

Conclusion

Results indicate that cell preservation techniques alter the transcriptome, by changing the expression of specific genes, such as FTE cell marker genes. However, fresh cells were detecting genes associated to DNA damage stress, questioning the sample integrity of fallopian tubes.

ABBREVIATIONS

OC	Ovarian Cancer
EOC	Epithelial ovarian cancer
HGSOC	High-grade Serous ovarian Carcinoma
TSG	Tumour suppressor gene
OCP	Oral Contraceptive Pill
PCOS	Polycystic Ovarian Syndrome
FTE	Fallopian Tube Epithelium
STIC	Serous Tubal Intra-Epithelial Carcinoma
SNS	Sympathetic Nervous System
scRNA-seq	Single Cell RNA Sequencing
cDNA	Complementary DNA
RT-qPCR	Reverse Transcriptase quantitative PCR
SCI	Single Cell Isolation
MBMT	Microdroplet-based microfluid technology
Smart-Seq2	Switch Mechanism at 5' end of RNA Template sequencing 2
TSO	Template Switching Olygos
UMI	Unique Molecular Identifier
DE	Differentially Expressed

CHAPTER 1 – Introduction

1.1. Ovarian cancer

1.1.1. Background information

Ovarian cancers (OCs) are the sixth most common type of cancers amongst women in UK, however the impact of OC does not only arise from its prevalence amongst women, but in its mortality, where OCs are one of the leading causes of death from gynaecological cancers. This high rate of mortality is due to the late diagnosis of these cancers, where OCs initially asymptomatic, are identified at an advanced stage, in which OC patients present metastasis onto organs such as lungs and liver or present tumours spread to the peritoneum outside the pelvis with close proximity to lymph nodes at most (stage IV and stage III respectively) [1-3]. Epithelial Ovarian Cancer (EOC) is the most common classification of OC, where 90% of diagnosed ovarian cancers are classified as EOC in the UK [1,2].

Current tests for EOC diagnosis require the concentration analysis of mucin 16 protein (CA125) in serum along with ultrasound imaging. Additionally, CA125 tests have been used for screening other malignancies such as prostate, lung and colorectal cancer. Clinical trials from UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) have shown that annual CA125 biomarker screening coupled with imaging had induced a limited decrease of EOC mortality [4]. An increase of serum CA125 concentration can been attributed to the initial development of EOC, however annual screenings become inefficient in detecting increases of CA125 when patients have developed the cancer before screening, demonstrating the restricted efficiency of this protocol to detect early stages of EOC development [4,5]. As a consequence, these findings emphasize the present demand in identifying reliable biomarkers for early EOC detection. Several risk factors which increase

the likelihood of developing EOC has been considered with sufficient evidence, whilst the mechanism which underlie the initial formation of EOC has not been fully established.

1.1.2. Risk factors and proposed mechanisms

The development and progression of EOCs rely on somatic mutations which will determine the aggressiveness of the disease. For instance, High-grade serous ovarian carcinoma (HGSOC) is an aggressive common type of EOC which is mostly found presenting mutations in tumour suppressor genes (TSGs), such as TP53 and to a lesser extent in BRCA1, BRCA2 and PTEN [3,6]. Furthermore, patients with hereditary breast cancer ovarian syndrome associated to BRCA1/2 inherited mutations account at least 5% of EOC discovered cases. As a consequence, the susceptibly to EOC will depend on hereditary factors [7,8]. As observed in figure 1.1, OC incidences have a strong correlation to age, where the average number of diagnosed OCs per year peak in the 65 to 69 age group patients. Therefore, EOC can also be developed through aging [1].

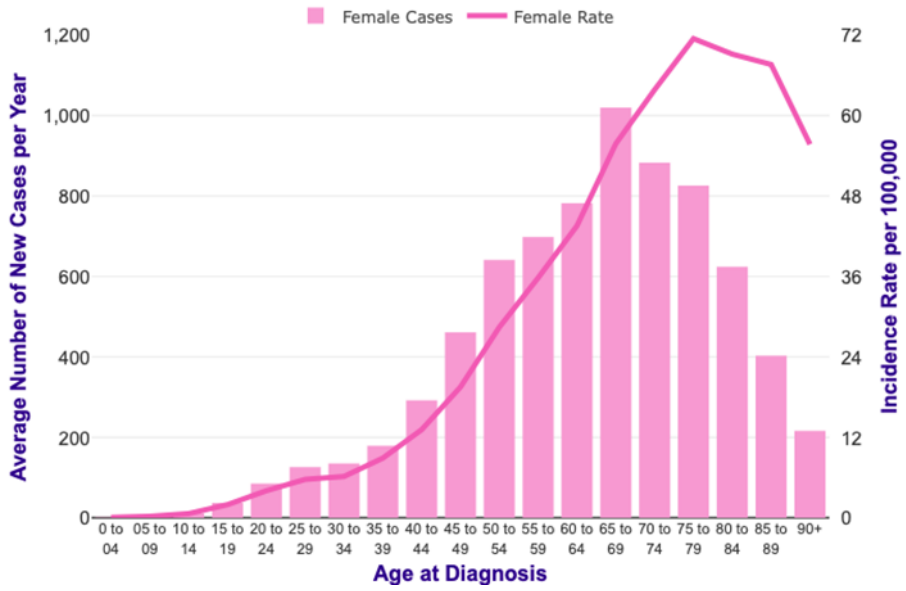


Figure 2.1: A graph representation depicting the correlation between age of diagnosis, incidence rate and average number of discovered cases per year. The incidence rate rises sharply from patients around the age of 45 to 49, peaking in 75 to 79 age-group-patients. On the other hand, the average number of diagnosed cases peak around the age of 65 to 69. [1]

Sustained continuation of ovulation cycles was considered a potential risk factor for OC development, theorising that the ovarian mesothelium is damaged during ovulation and the lack of ovulation breaks would prevent its replenishment [9]. This hypothesis was later supported by epidemiological studies, where nulliparous women were at higher risk of acquiring OC compared to patients who had ovulation breaks induced by pregnancy and breast-feeding [10,11]. Moreover, the use of oral contraceptive pills (OCP), which also induce ovulation breaks, has shown a negative correlation with OC risks [12]. On the other hand, the continuous ovulation hypothesis has been refuted, since patients with polycystic ovarian syndrome (PCOS) have irregular ovulation cycles and are also at high risk or acquiring EOC [13]. Additionally, the preventative effect of OCPs for EOC has been observed during the use of progestogens such as levonorgestrel, which is used as contraception without preventing ovulation [14,15].

As an alternative for the continuous ovulation hypothesis, the excessive exposure of gonadotropins was theorised to be the responsible of increasing the risks for OC, where overstimulation of gonadotropin hormone receptors found in OSE would lead to cancer-inducing proliferation [16]. This theory could explain the reduced risk for OC in women who have had ovulation breaks through pregnancy or OCPs, since these women maintain low levels of gonadotropins during these ovulation pauses. Furthermore, the gonadotropin theory would also explain the increased risks for OC in PCOS patients, since PCOS causes gonadotropin imbalances by enhancing the presence of luteinizing hormone in plasma [17]. Animal models have demonstrated the effects of gonadotropin hormones exposure in enhancing angiogenesis and growth in OC [18,19]. Although animal studies have demonstrated the effect of gonadotropins in increasing OC progression, these findings do not prove that gonadotropins have a promoting effect in OC. Moreover, there is no strong

evidence suggesting that excessive gonadotropin exposure induces malignant transformation [20].

However, the high heterogeneity presented amongst the different EOC types suggest that the cell origin of this disease is highly diverse. For instance, endometriosis induced by retrograde menstruation has been associated as precursors for endometroid and clear-cell EOCs, suggesting that these EOC subtypes arise from endometrial tissue accumulated close to the ovaries [21,22]. The fallopian tube epithelium (FTE) has also shown to have a possible role in the origin of specific EOC types. This claim was first considered when examining the fallopian tube histology in patients who were BRCA1 mutation carriers or had a noticeable family history of OC [23]. These patients received a bilateral salpingo-oophorectomy for OC risk reduction, however dysplastic lesions found in the distal side of fallopian tube epithelium were identified in most of the patients after the risk reduction treatment [23,24]. These lesions known as serous tubal intra-epithelial carcinomas (STICs) expressed abnormal levels of p53, suggesting that TP53 mutations could be precursors for STIC development [25,26]. To further consolidate these findings, Przybycin et al [27] had identified the presence of STICs in 59% of patients presenting serous EOC, whereas Kuhn et al [26] demonstrated the commonality of p53 overexpression between STICs and its metastasized counterparts located in OSE. Therefore, cells from FTE could be the early precursors of serous EOCs, in which the underlying mechanism could be summarised in figure 1.2.

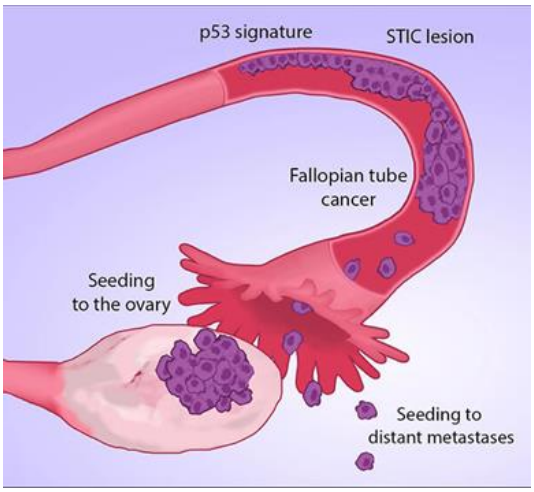


Figure 1.2: Initial steps of OC development. Secretory FTE cells develop TP53 mutations, causing an excessive production of p53. Cells with overexpressed p53 develop into dysplastic lesions known as serous tubal intra-epithelial carcinomas (STICS). After the development of STICS, cells from these lesions migrate towards the OSE and to other organs for metastasis. [28]

1.1.3. Fallopian tube and ovarian cancer associated discoveries

After establishing a relation between the presences of STIC lesions in the distal side of fallopian tubes with the origin of serous EOC, the biology of these lesions was studied to establish the original cell type responsible for the dysplasia. STICs arise from the malignant transformation of FTE secretory cells, presenting mutations in TSGs such as in TP53 and in PTEN [3,6,29]. Furthermore, a recent study from Hellner et al [30] has shown that SOX2 is overexpressed in non-dysplastic secretory cells from patients with HGSOC. This study also suggests that SOX2 expression in secretory FTE cells could be a precursor for STICs and could be used as a prognostic marker to determine whether HGSOC will reoccur after chemotherapy.

In relation to the progression of OC, chronic stress has been suggested to enhance ovarian tumour growth, particularly stress induced by the constant activation of the sympathetic nervous system (SNS) or by the exposure to noradrenaline [31-33]. As depicted in figure 1.3, the stimulation of β2-adrenoreceptor (ADRB2) in OC, will lead to the synthesis of interleukin-8 (IL8), a chemokine which will facilitate the OC progression by promoting

angiogenesis [32]. A murine invitro study has revealed the possible crosstalk between ovaries and FTE secretory cells with silenced PTEN, where these secretory cells induced the secretion of noradrenaline from the ovary. Although the mechanism which underlies this crosstalk is has not been established, this finding reveals the capability of ovarian tissue to stimulate the stress pathway in the absence of SNS [33].

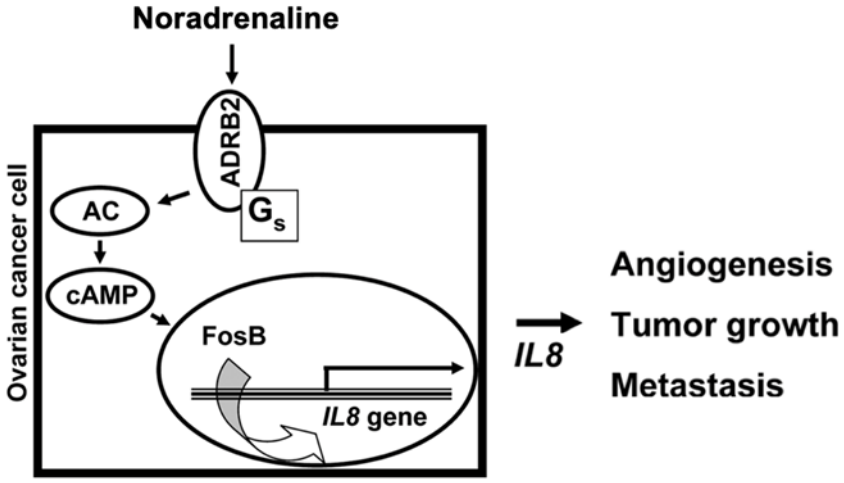


Figure 1.3: Depiction of a neural stress pathway associated with ovarian cancer progression. Noradrenaline activates the β2-adrenoreceptor (ADRB2), which triggers the activity of Adenylyn Cyclase (AC), increasing the concentration of cAMP. The increase of cAMP leads to the activation of transcriptor factor FosB, inducing the expression of IL8 responsible for angiogenesis and tumour development in OC [32].

Single cell RNA sequencing (scRNA-seq), a novel approach for transcriptome analysis has been implemented to analyse the in vivo effects of oestradiol exposure in promoting dysplastic lesions in FTE and OSE. The in vivo study was used to simulate the detrimental effects of hormone replacement therapy in promoting EOC, whilst scRNA-seq had been used to unravel the cancer associated pathways induced by oestradiol. Genes such as the Growth Regulating Oestrogen Receptor Binding 1 (GREB1) was found upregulated in FTE and OSE after oestradiol exposure [34]. ScRNA-seq analyses have more advantages than conventional RNA microarray techniques, however scRNA-seq is relatively new and has scarcely been applied for OC studies. Therefore, scRNA-seq is one of the principal subjects of this project.

1.2. Single cell RNA sequencing

1.2.1. Background information

Since the late 1970s, earliest studies involved in transcriptome analysis consisted in collecting mRNA which would be converted to complementary DNA (cDNA) for quantification analysis using techniques such as northern blotting or reverse transcriptase quantitative PCR (RT-qPCR) [35, 36]. These approaches enabled the recording of a small number of expressed genes, however the biology of recorded transcripts remained unknown until the later development of high throughput technologies i.e.: DNA sequencing [37]. Northern blotting and RT-qPCR became obsolete during the introduction of RNA microarray analysis in the early 2000s [38]. Microarrays significantly reduced the cost and labour for transcriptomic studies, however these experiments still relied upon the pre-existing knowledge of the analysed genes. Furthermore, comparing microarray results across multiple experiments demanded intricate methods of normalisation [39].

However, the following cost reductions in sequencing technologies has led to the development of scRNA-seq, a novel instrument which reveals the transcriptome profile in individual cells [40,41]. ScRNA-seq provides a significant upgrade in the analysis of RNA transcripts, since previous studies have suggested that gene expression is physiologically heterogeneous across individual cells [42]. These claims were afterwards supported by scRNA-seq experiments [41], where this procedure was first applied in the whole transcriptome analysis of single mouse blastomeres [40]. Indeed, the first scRNA-seq study was able to reveal the expression of more genes than microarray techniques whilst also sequencing mRNAs presenting unknown splice junctions [40]. From this study, a variety of alternative scRNA-seq methodologies have been introduced, nevertheless, its process

consists on the following three steps: (1) the isolation of single cells, (2) preparation of scRNA-seq libraries and (3) pre-processing of the data.

1.2.2. Cell Isolation

To implement scRNA-seq analysis on individual cells, these cells must be separated away from its extracellular matrix and whilst cleaving cellular adhesions, figure 1.4 depicts some of the techniques used for single cell isolation (SCI).

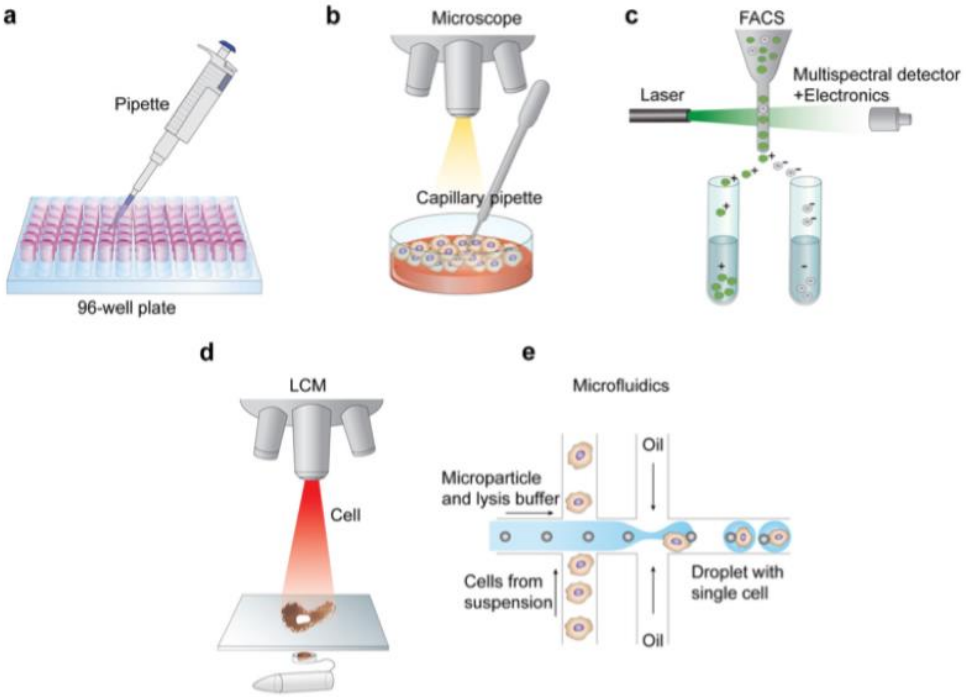


Figure 1.4: Illustrations of single cell isolation techniques. (a) Limiting dilution methods achieve volumes with low cell concentrations, with the purpose of obtaining a singular cell per well. (b) Micromanipulation utilize capillary pipettes for cell isolation aided with microscopic visualization. (c) Fluorescent active cell sorting (FACS) is implemented to achieve the purification of fluorescent-tagged cells from a heterogeneous population. (d) Laser capture microdissection (LCM) makes use of laser technologies to isolate cells from solid samples. (e) Microfluid technology generate nanolitre-sized droplets which capture the cells for scRNA-seq, droplets with single cells contain beads used for Drop-seq procedures [43].

Limiting dilution (fig. 1.4a) and micromanipulation (fig. 1.4b) are traditional methods for SCI which present unique characteristics and limitations, where limiting dilution involves separating cell found in medium by dilution, with the aim of achieving low cell concentrations of 0.5 cells per well. However, results from this method are inefficient, since only a third of

the prepared wells using this process will achieve successful SCI [44]. On the other hand, micromanipulation requires the use of capillary pipettes which can extract individual cells such as single blastomeres or single cell from early embryos [40,45]. Although micromanipulation could be automated and provides visual confirmation for SCI, this process presents low throughput when preparing a large population of isolated cells [45]. Enzymatic treatment is another traditional SCI method used for isolating cells which present strong cell adhesions or are highly attached to its extracellular matrix. This method could be applied for single cell genomic analyses, however the use of enzymes such as collagenases could act as a cellular stressor, altering cellular surface proteins and subsequently changing the transcriptome profile [46,47].

To overcome the limitations given from traditional SCI techniques, advanced methods such as fluorescent active cell sorting (FACS) (fig. 1.4c), laser capture microdissection (LCM) (fig. 3d) or microdroplet-based microfluid technology (MBMT) (fig. 1.4e) could be considered. FACS is used to separate cell populations presenting specific cell markers which are tagged with fluorescent monoclonal antibodies. This technique presents high throughput, which is desired by studies requiring a large population of cells and can perform either positive or negative selection of antibody-tagged cells. However, FACS can present technical challenges, since this method might induce the inclusion of cells caused by sorting pressure during the procedure, which could alter the read quality of scRNA-seq [48]. LCM techniques facilitate the dissection of individual cells from thin-sectioned tissues which could be either fresh, frozen or formalin-fixed. Although it is important to note that most formalin-fixed cells happen to be degraded for scRNA-seq [49], LCM should be considered an alternative for isolating cells which are highly fixed to its microenvironment [47,49,50]. MBMTs is a SCI method for organising a suspension of cells, where these are trapped in buffer droplets

trapped in a hydrophobic fluid, these droplets are subsequently used as chambers where the needed reactions for scRNA library preparation take place [51].

1.2.3. scRNA library preparation

Transforming RNAs into cDNA via reverse transcription is essential step before sequencing the transcriptome of single cells. During the first scRNA-seq procedure, polyd(T) primers used to capture mRNAs for reverse transcription, however mRNAs can form secondary structures such as hairpins and loops which makes reverse transcriptase prone to stalling, thus the sequence generated for mRNAs longer than 3 kbp were incomplete and more biased to sequencing closer to the mRNA poly(A) tails (3' biased) [40,52]. Alternatives methods for preparing single cell mRNA libraries have been designed to overcome these limitations.

Switch Mechanism at the 5' end of RNA Templates sequencing 2 (Smart-seq2) is a protocol for scRNA-seq which is used to sequence mRNA transcripts at full length with the aim of detecting the presence of different mRNA sequence isoforms, this protocol is outlined in figure 1.5. Along with polyd(T)primers, smart-seq2 utilizes betaine and high concentrations of magnesium chloride (9-12mM) to overcome mRNA secondary structures and enhance cDNA yields.

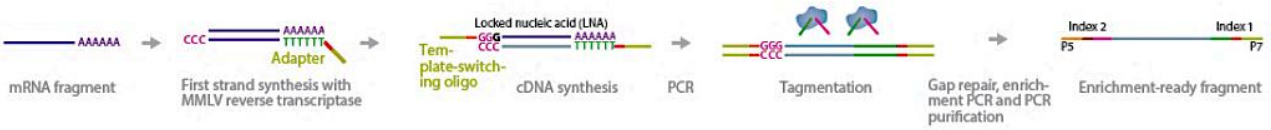


Figure 1.4: Summarized flowchart for Smart-seq2 library preparation. Tagmentation by Tn5 enzyme cleaves cDNA sequences while incorporating a recognized sequenced on the cleaved site, these cleave sites with the Tn5 implemented sequences, are used to create the Index 2 and Index 1 segment of the cDNA fragment which is optimised for cDNA sequencing [54].

Furthermore, smart-seq2 depends on template-switching oligos (TSOs) which contain three riboguanosines at the 3’ end. The last riboguanosine of the 3’ end is modified producing a locked nucleic acid which will be bound next to the 5’ end of the mRNAs which will be reverse transcribed. This provides thermostability ensures that the generation of a complete cDNA sequence from the moloney murine leukaemia reverse transcriptase, since this enzyme will add extra bps after mRNA reverse transcription, which are complementary to the TSOs [53]. After reverse transcription, cDNAs undergo tagmentation catalysed by Tn5 tansposases which will prepare the cDNA libraries for sequencing. [53, 54]

Libraries generated from Smart-seq2 require that each test tube contains the transcriptome of one single cell, which can increase the cost for sequencing or limit the number or reads per cell when trying to analyse a large number of single cell transcriptomes. To reduce the cost, Cel-seq and Drop-seq protocols can be applied instead of Smart-seq 2. Cel-seq and Drop-seq rely on unique molecular identifiers (UMIs), which are 4bp-long sequences that are assigned onto each cDNA read. Each cell will be presenting a unique UMI implemented onto its cDNA generated sequences, therefore UMIs will be used as cellular barcodes to determine the cell origin of each reads during sequencing. This enables multiplex sequencing, where transcriptome libraries from all cells can be sequenced together within a sample [53-56]. However, like the first scRNA-seq protocol, Cel-seq and Drop-seq are heavily 3’ biased [40,55,56], which is the major drawback when overcoming the disadvantages of Smart-seq2.

1.2.4. Data processing

Before analysing the data from scRNA-seq, quality control (QC) and data cleansing is assessed after single cell RNA reads are sequenced. QC can be performed using FastQC, which is an online-accesible software which statistically analyse the quality of reads based on aspects such as length distribution and guanine-cystosine content [57]. Data cleansing will consist on eliminating low quality reads following a desired criterion and eliminating UMIs from the reads if implemented, Trim Galore is a software use to sequences added onto the reads during scRNA library preparations (e.g. UMIs), whilst selection of quality reads will depend on made on the specific scRNA-seq study [58, 59]. To align reads with known sequences found in the genome, Spliced Transcripts Alignment to a Reference (STAR) software or Burrow-Wheeler Aligner (BWA) is used to map the sequences with exome regions found in the human genome atlas [59-61]. Kalisto on the other hand, is an alternative tool for mapping sequences, used to map scRNA-seq reads via pseudoalignment with a reference transcriptome profile and it is normally used to compare reads with known mRNA isoforms [59,62]. After mapping the sequenced reads with a reference genome or transcriptome, statistical analyses are subsequently implemented to evaluate the data.

1.2.5 Hypotheses and Aims

This research project aimed to analyse the scRNA-seq results obtained from FTE cells which had isolated from endometrial cancer patients and OC patients. Additionally, since information related to transcriptome changes caused by different cell preservation techniques is scarce in the literature, this project firstly intended to observe the transcriptomal changes induced by cell culturing and cell cryopreservation.

Hypothesis

- Our hypothesis is that cell preservation techniques can induce transcriptomal alteration, these alterations subsequently provoke changes in the composition of FTE cells and in the expression of genes which are known to be cancer associated.

Research Aims

- Observe the level of heterogeneity between fresh cells and cells treated with either cryopreservation or different durations of cell culturing.

- Identify FTE cell population changes induced by cryopreservation and different cell culturing periods.

- Recognise the presence of cell cycle alterations induced by different types of cell preservation.

- Acquire evidence of genetic expression related to OC pathways in FTE cells.

2. Materials and Methods

The undertaken research was an observational, quantitative study exploring the scRNA-seq data acquired from the Gynaecological Oncology Targeted Therapy Study 01. The data, which was provided in a .rds format, presented the expression measurements for 14,176 genes in 3,806 cells across 6 OC patients and 4 Endometrial Cancer (EC) patients. Additionally, methodological information was supplied, documenting the conducted experiments and procedures which were responsible for generating the data. These methodologies are summarised the flow chart from figure 2.1, however this chapter will (1) outline the details of the undertaken experiments and procedures, as well as (2) describe the coding implemented for data interpretation and statistical analyses.

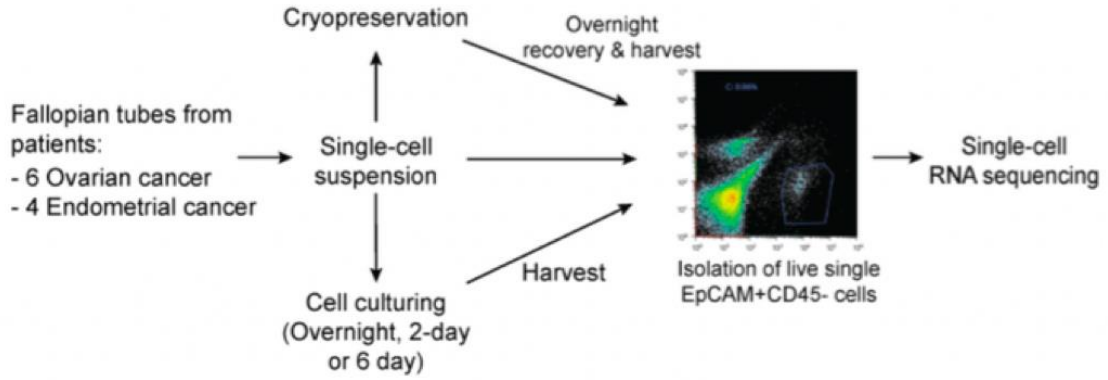


Figure 2.1.: Experimental procedure outline for data acquisition. Samples of fallopian tubes were obtained from ovarian cancer patients and endometrial cancer patients, cells from the fallopian tube were suspended and subsequently cryopreserved or cultured for different time periods. Overnight recovery was provided to cryopreserved cells before harvesting. All harvested cells which were EpCAM+ but CD45- were purified for scRNA-seq (figure provided by supervisor).

2.1 scRNA-seq methodology details

2.1.1. Tissue dissociation, cell culturing and cryopreservation

Fallopian tube samples were processed within an hour after collecting these from the hospital, where relevant information details of the patient can be found in table S1. Tissue disassociation was achieved using Human Tumour Dissociation kit (Miltenyi Biotec). After obtaining single-cell

suspensions, cells were either cultured or cryopreserved. To obtain exclusive culturing of FTE cells from patients, the medium was adopted for FTE culturing using the same protocol from Kessler et al. 2015 [63]. Cultured cells were harvested overnight, after 2 days or after 6 days. Cryopreservation was executed by freezing suspended cell at -80ºC for 24 hours and subsequently preserved in liquid nitrogen. cDNA generation was not possible from cells thawed from liquid nitrogen, due to the harshness of cryopreservation (results not provided), therefore cryopreserved cells were cultured in Basal Medium 2 overnight for recovery.

2.1.2. Cell sorting, library construction and cDNA sequencing

To isolate FTE epithelial cells from either freshly-dissociated cells or from cell culture harvests, cells were stained with EpCAM-APC and CD45-FITC antibodies and sorted through FACS, isolating EpCAM+CD45- cells individually into wells containing lysis buffer and RNase inhibitor (Clontech). After cell sorting, cDNA synthesis and library construction were performed following the Smart-seq2 protocol from Picelli et al. 2014 [53]. After the selection of wells containing good-quality single cell cDNA, the cDNA content was normalised to 2µg/ml before sequencing. Miniaturized Nextera XT was used for multiplex sequencing of 384 cells per batch whilst following the RNA-seq protocol from Mora-Castillo et al. 2016 [64].

2.1.3. scRNA-seq data pre-processing and quality control

Before STAR mapping the sequenced reads to UCSC hg19 human genome assembly, index1 and index2 sequences were trimmed from the reads using Trim Galore. After genome mapping, the number of reads assigned to genes were counted by FeatureCount [65]. The expression quality was visualised through R package scater [66] and low-quality cells were filtered out if the cell reads followed any these three criteria:

1. Number of detected genes per cell is less than 1,200 or more than 7,500.
2. The total number of counted reads is less than 0.1 million.
3. The percentage count of the top 200 detected genes is more than 85%.

These criteria ensured enough complexity of the sequenced transcriptome, whilst the upper cut-off from the first criterion eliminated putative doublets from the analysis. The number of detected genes and the number of total reads averaged 3,665 and 763,233 respectively. The experiment ended up with 3,806 sequenced cells: in which 2,113 were fresh cells, 1,135 cells were subjected to cryopreservation, 404 cells were O.N. cultured, 51 cells were 2-day cultured and, 103 cells were 6-day cultured.

2.2. RStudio coding and data analysis

2.2.1. Installing packages and loading the data

RStudio has been the software utilised for generating the desired plots for data visualisation and for implementing statistical analyses onto the data. The latest version of libraries “scran” and “scater” were installed and loaded in RStudio with the use of the following code, green text followed with a hashtag are brief annotations explaining the segments of the full code:

```
# Code for "scater" and "scan" instalation
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("scater", version = "3.8")

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("scran", version = "3.8")

# Load libraries onto Rstudio
library(scater)
library(scran)
```

After installation, we associate the .rds file with the name “sce” as a simplified method to refer to the datafile:

```
sce = readRDS("ft_data.rds")
```

Summary information of the cell related to the experiment and information related to the analysed genes were obtained using the functions provided by scater, these functions being “colData()” and “rowData()” respectively. The following code provided the gene information and cell information in a .txt document, which has been transferred into a excel document for better visualisation:

```
# Generating cell information data
cellInfo = colData(sce)
write.table(cellInfo, file="cellInfo.txt", sep="\t", quote=FALSE)

# Generating gene informaiton data
geneInfo = rowData(sce)
write.table(geneInfo, file="geneInfo.txt", sep="\t", quote=FALSE)
```

2.2.2. Identifying Differentially Expressed genes

The genetic expression of cryopreserved, overnight (O.N.) cultured, 2-day cultured and 6-day cultured cells was compared with the genetic expression of fresh cell which were scRNA sequenced subsequently after tissue disassociation and FACS purification. Welch t-test between fresh cells and each preservation type was performed comparing the average read counts for each of the 14,176 gene. This was possible using the “findMarkers()” function, provided by scran. Results from the welch t-test were generated from the following code and were subsequently converted to excel, “preservation A” denotes the desired preservation method which will be compared to fresh cells:

```
# comparing Fresh cells to "preservation A" cells
# find entries (cells) which are either Fresh OR "preservation A"
subset_idx = which( sce$source == "Fresh" | sce$source == "preservation A" )

# do findMarkers only on these cells
out = findMarkers(sce[, subset_idx], sce$source[subset_idx])
out[["Fresh"]]

#generate a text
write.table(out[["Fresh"]], file="Fresh_to_preservation_A.txt", sep="\t", quote=FALSE)
```

2.2.3. T-SNE plot generation

To visualise the variability across the recorded cells in a two-dimensional plot, two T-SNE plots were generated from the sce data. The first plot was made to distinguish cells based on the preservation type, whilst the second plot distinguishes cells based on the batch origin during library sequencing. Since the distribution of cells in T-SNE is stochastic and randomised, the “set.seed()” command was implemented to make the generation of T-SNE plots more reproducible. After testing different settings, “set.seed(100)” showed most optimal for T-SNE plot visualisation. T-SNE plots were generated through these codes:

```
# 1) T-SNE plot distinguishing cells by preservation
set.seed(100)
sce = runTSNE(sce)
plotTSNE(sce, colour_by = "source")

# 2) T-SNE plot distinguishing cells by batch origin
set.seed(100)
sce = runTSNE(sce)
plotTSNE(sce, colour_by = "batch")
```

2.2.4 Heatmap Generation

The library “pheatmap” was installed onto Rstudio, this code was designed to visualise the 50 genes presenting the highest variability or read counts across cells from all types of preservation:

```
# Installing and loading the pheatmap library onto RStudio
install.packages("pheatmap")
library("pheatmap")

# Extracting the logarithm of gene expression counts from our variable sce
# "logcounts()" function provided by scater.
logCount = logcounts(sce)

# Using the function rowVars contained in the library matrixStats to compute
# the variance in expression values for each gene across all cells:
logCount_var = matrixStats::rowVars(logcounts(sce))

# Finding the top 50 most variable genes.
gene_idx = order(logCount_var, decreasing=TRUE)
gene_idx = gene_idx[1:50]
```

```
# Sorting the cells by preservation method
col_idx = order(colData(sce)$source)

# Creating a data frame containing the preservation method information by
# copying a reordered version of the source cell information column
my_sample_col <- data.frame(method = colData(sce)$source[col_idx])
row.names(my_sample_col) <- colData(sce)$Sample[col_idx]

# Scaling of all the measurements to ensure that different genes are
# visually comparable:
logCount_scaled = t(scale( t(logCount)))

# Final code for heatmap plotting
plotHeatmap(sce, features = gene_idx,
            colour_columns_by=("source"),
            columns = col_idx,
            cluster_cols=FALSE, center=TRUE, symmetric=TRUE, zlim=c(-3, 3))
```

2.2.5. Expression plots and Wilcoxon signed rank test

Plots representing the logarithmic distribution of read counts for specific genes across the different presentation types were generated using the “plotExpression()” function provided after loading scater. As seen in the code, the function was used to generate plots for more than 1 gene at a time:

```
## expression plots for desired genes represented as GeneA to GeneD
plotExpression(sce, features=c("GeneA","GeneB","GeneC", "GeneD"), x="source", colour_by = "source")
```

This code was used to observe the genetic expression of secretory and ciliated cell markers [67,68] and the expression of OC associated genes [30-33]. The presence of cell cycle associated genes was also visualised using the “plotExpression()” function.

Most of the expression plots, presented a significant number of outliers, which could impact the reliability of the Welch t-tests. Thus, to determine whether the difference between fresh cells and each preservation type was significant, Wilcoxon signed rank test with continuity correction was implemented using the following code:

```
# Setting logCount as logarithmic raw read counts
logCount = logcounts(sce)

# Wilcoxon test code
wilcox.test(logCount["GeneA", which(sce$source=="Fresh")], logCount["GeneA", which(sce$source=="preservation A")])
```

2.2.6. Correlation plots and Spearman correlation test

Correlation plots were made to compare the logarithmic read counts of two different genes across all cells. These plots were generated using the following code:

```
geneA = "GeneA" # gene A of interest
geneB = "GeneB" # gene B of interest

logCounts = logcounts(sce) # logaritmic raw read counts

# creating a data frame to hold the expression values for the two genes of interest
df = data.frame( geneA = logCounts[geneA, ], geneB = logCounts[geneB, ], source = sce$source )
# generating the correlation graph from the created data frame
ggplot(df, aes(x=geneA, y=geneB, colour=source)) + geom_point() + xlab("GeneA") + ylab("GeneB")
```

Spearman correlation tests were executed for each plot, cells which presented no reads in any of the two genes from the analysed correlation plot were excluded from the spearman correlation test. The code for spearman correlation was developed using the functions provided by the “Hmisc” library:

```
geneA = "GeneA" # gene A of interest
geneB = "GeneB" # gene B of interest

# creating the dataframe with the logcounts for each gene in each cell, excluding cells which
# have undetected expression for one of each genes
df = data.frame( geneA = logCounts[geneA, ], geneB = logCounts[geneB, ], source = sce$source )
# excluding cells which have 0 readcounts for any of the genes of interest
idx = which(df$geneA > 0 & df$geneB > 0)

# calculate the correlation using rcorr() function
rcorr(df$geneA[idx], df$geneB[idx], type = "spearman")
```

3. Results

3.1 Introduction

This chapter will consist on the presentation and interpretation of RStudio coded plots, along with inferential statistical analyses. Extraction of the experimental details related to the cells and detected genes can be found in the excel supplementary material. However, an extract of these documents can be observed in figure S3.1 and S3.2. Raw data obtained from the welch T-tests comparing the number of read counts across the different cell sources are also found in the supplementary material.

3.2 Observation of cell preservation heterogeneity

The results from the welch t-tests are outlined in figure 3.1, which quantify the number of detected differently expressed (DE) genes compared to fresh cells. These DE genes exhibited strong evidence of variability ($p<0.001$). Our analysis show that cryopreservation produced the greatest change in the cellular transcriptome, whereas the genetic expression induced by cell culturing is fluctuated over time. 2-day cultured cells presented the closest transcriptomal resemblance to fresh cells, where 26.2% of all detected genes where observed as DE, however the percentage of DE genes is increased in 6-day cultured cells up to 43.26%.

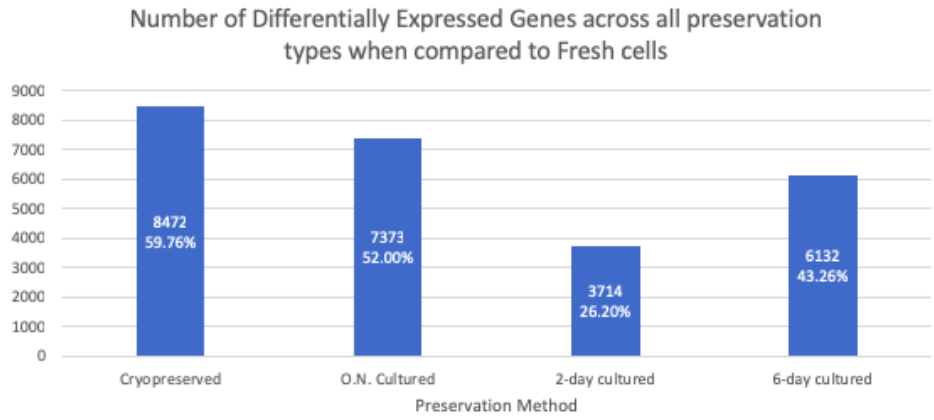
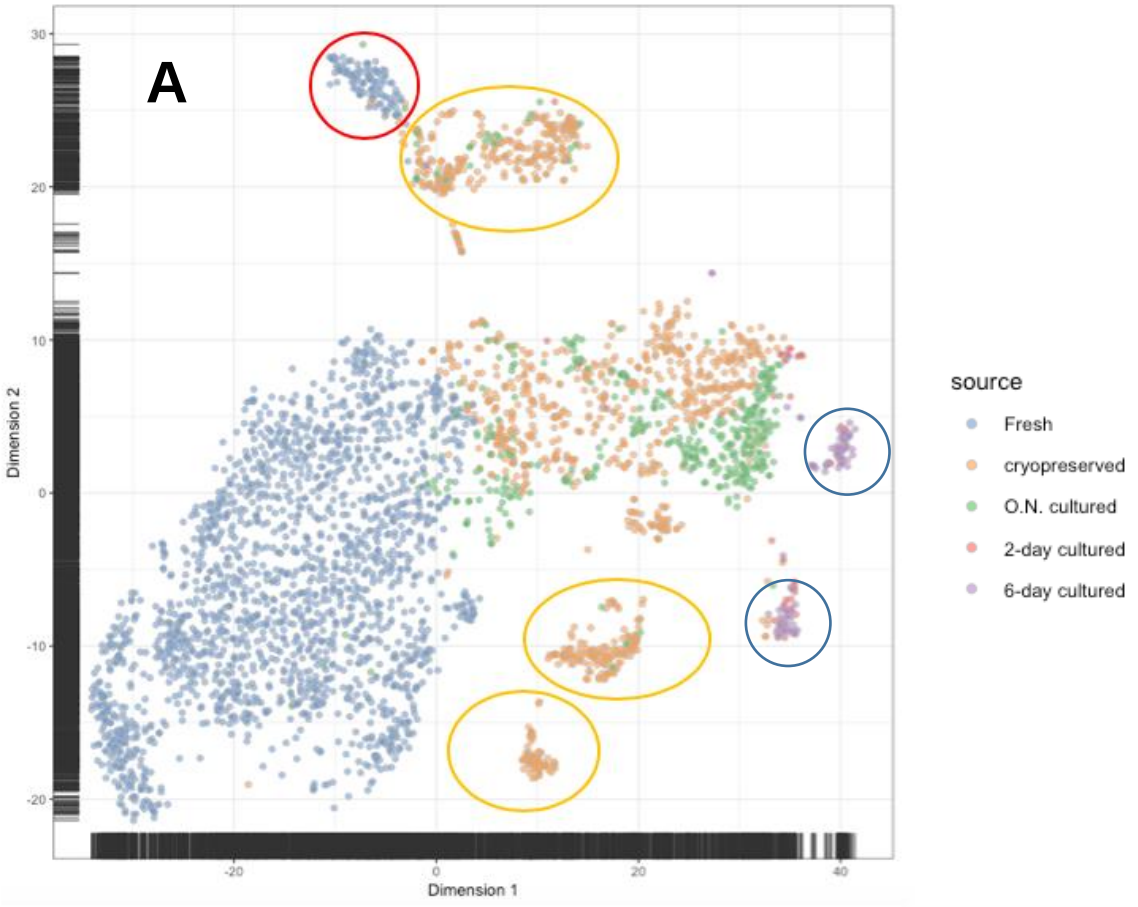


Figure 3.1: Differential Expression across all cell preservation types: Cryopreserved cells present the highest level of dissimilarity, where 59.76% of observed genes are expressed differently to fresh cells. O.N cultured cells express 52% of observed genes differently to fresh cells. However, this level of dissimilarity is reduced in 2-day cultured cells.

Generated T-SNE plots are shown in figure 3.2A and 3.2B, where cell-representing dots are coloured by source origin in figure 3.2A whilst dots from figure 3.2B are coloured by the batch origin during cDNA sequencing. The proximity between dots is directly proportional to the similarity across cells, the dot distribution presented in these plots show that the detected transcriptome of fresh cells is constant with minor levels of variability among the fresh cell population. Whilst cell which have been cryopreserved, or 6-day cultured present a higher degree of variability, as they are found disperse forming independent clusters in T-SNE plot 4.2A. However, as seen in plot 4.2B, most of these cells presented in clusters share a common batch origin, which suggest that cells from different batches will have an inevitable effect in gene sequencing and transcriptome analysis. This batch effect is strongest in cells which have undergone cryopreservation or cell culturing. 2-day cultured cells are hardly visible in these plots, this caused by the low number of sequenced cells which were 2-day cultured (51 out of 3,806 cells).



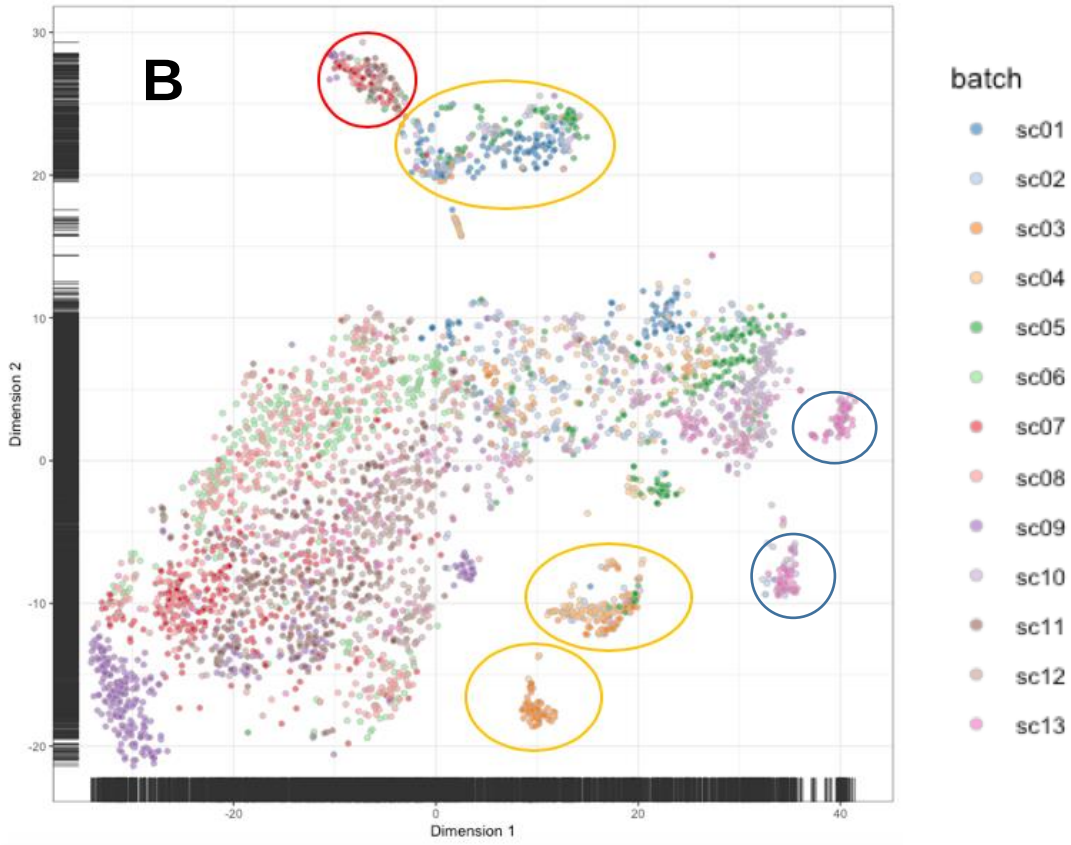


Figure 3.2: T-SNE plots presentation. These are two different T-SNE plots which have been generated from the same data, each dot from both T-SNE plots represents an individual cell. The red circled cells represent the fresh cells with the highest heterogeneity among the population of fresh cells. These highly heterogenic fresh cells arise from batches sc05 to sc11. Yellow-circled clusters found closest to the Dimension 1 axis are mainly cryopreserved cells from batch sc03 and sc04. Blue-circled clusters are mostly 6-day cultured cells from batch sc07.

For further analysis, the relative expression of the 50 most variable genes across all the cell population were visualised in a heatmap depicted in figure 3.3. From the heatmap, the expression of KRT7 was observed to be expressed highest in 2-day and 6-day cultured cells. KRT7 is a widely known marker expressed in secretory FTE cells, therefore the analysis of this gene was then considered relevant for further analysis. The expression of genes such as FOS and EGR1 was found highest in fresh cells, therefore further analysis was also implemented on these genes. From the heatmap visualisation, it is complicated to determine any gene which is expressed predominantly in cryopreserved cells.

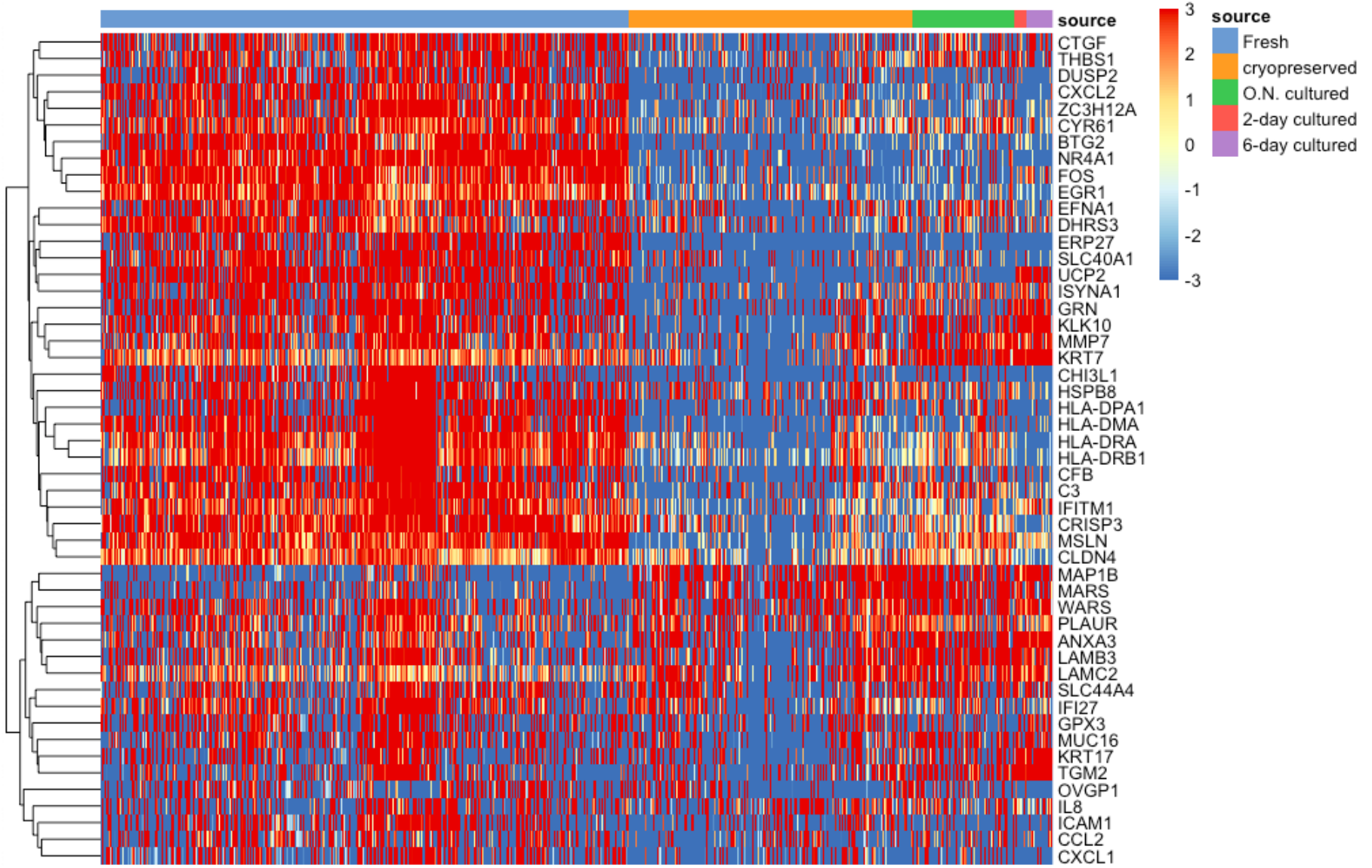


Figure 3.3: Heatmap presentation of top 50 variable genes: Each cell is depicted as a column in the heatmap; the columns have been ordered by their preservation method. The hierarchical clustering seen on the left side of the figure indicate the expression similarity among the variable genes. Relative expression variability has been scaled to parameters between 3 to -3. The expression of the top 32 variable genes (from CTGF to CLDN4) are shown to be highly predominant in fresh cells.

3.3. Prevalence examination FTE cell markers

For the analysis of FTE markers across the cell population, an expression plot was generated for the EPCAM gene along with GAPDH housekeeping gene, which can be seen in figure 3.4. As the housekeeping gene, EPCAM was predicted to be highly expressed across cells from all sources. EPCAM expression was shown to be stable across the different preservation types, however the average expression of GAPDH in cultured cells (O.N. 2-day and 6-day cultured) was slightly higher than the GAPDH expression observed in fresh cells.

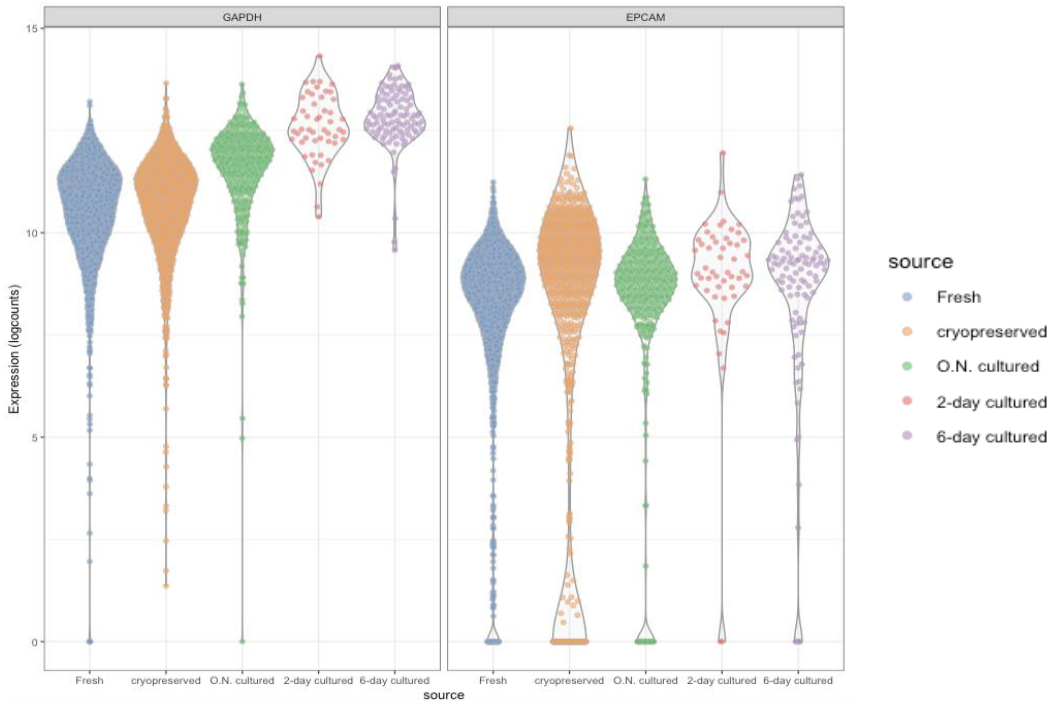


Figure 3.4: Expression plots for FTE housekeeping genes. The detected expression can be interpreted as the logarithmic number counter reads for a particular gene (logcounts). The Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1.

From the heatmap, an elevated expression of KRT7 secretory cell marker in cultured cells was observed. Therefore, expression plots were made comparing KRT7 expression with other known cell markers found in FTE. KRT7 was compared along with PAX8, CCDC78 and CCDC17, this is showed in figure 3.5. As observed in the figure, KRT7 expression is increased in cells subjected to cell culturing. PAX8 expression is detected lowest in cryopreserved cells, but stable across the rest

of the other cell sources. CCDC17 and CCDC78 have been poorly detected across the population. From the plot, expression of CCDC17 and CCDC78 could be visually interpreted as highest in fresh, cryopreserved and O.N cells. The evidence claiming that CCDC17 and CCDC78 expression is different between fresh cells and 2-day or 6-day cultured cells has moderate significance ($p < 0.05$), as observed in the Wilcoxon results recorded in Table S2 and S3.

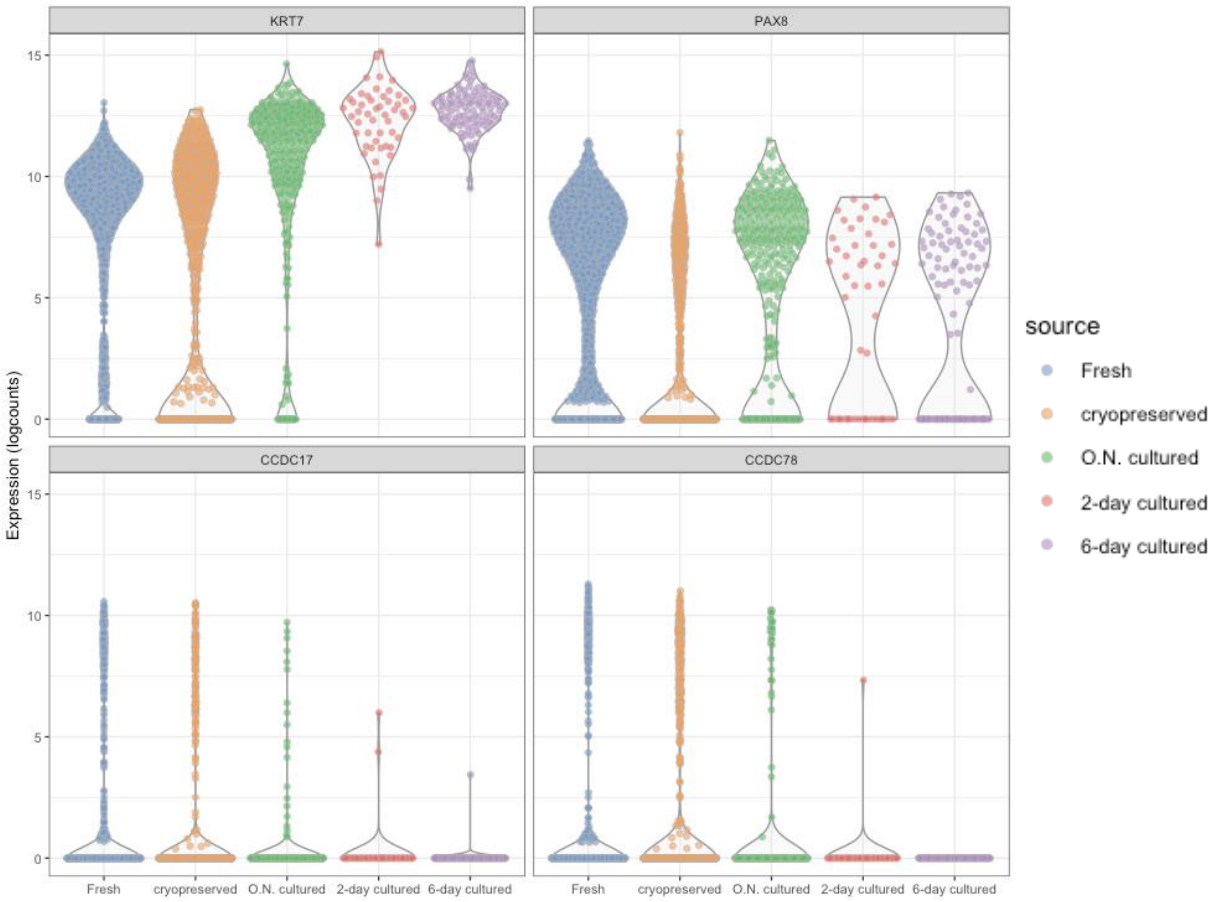


Figure 3.5: Expression plots for FTE cell marker genes. The expression of either CCDC17 or CCDC78 was detected in approximately 10% of all sequenced cells. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1.

After observing the expression trends for these FTE cell markers, correlation plots were generated to determine the co-expression of KRT7 with PAX8 and of CCDC17 with CCDC78. These correlation plots, depicted in figure 3.6, demonstrate that PAX8 is frequently co-expressed with KRT7. Not all cells present this co-expression of PAX8 and KRT7, as a consequence, the evidence of correlation is not estimated to be significant ($p > 0.05$). This result does not apply to CCD17 and CCDC78, since

the expression of these genes are correlated and co-expressed with high statistical significance (r_s = +0.44, p < 0.001).

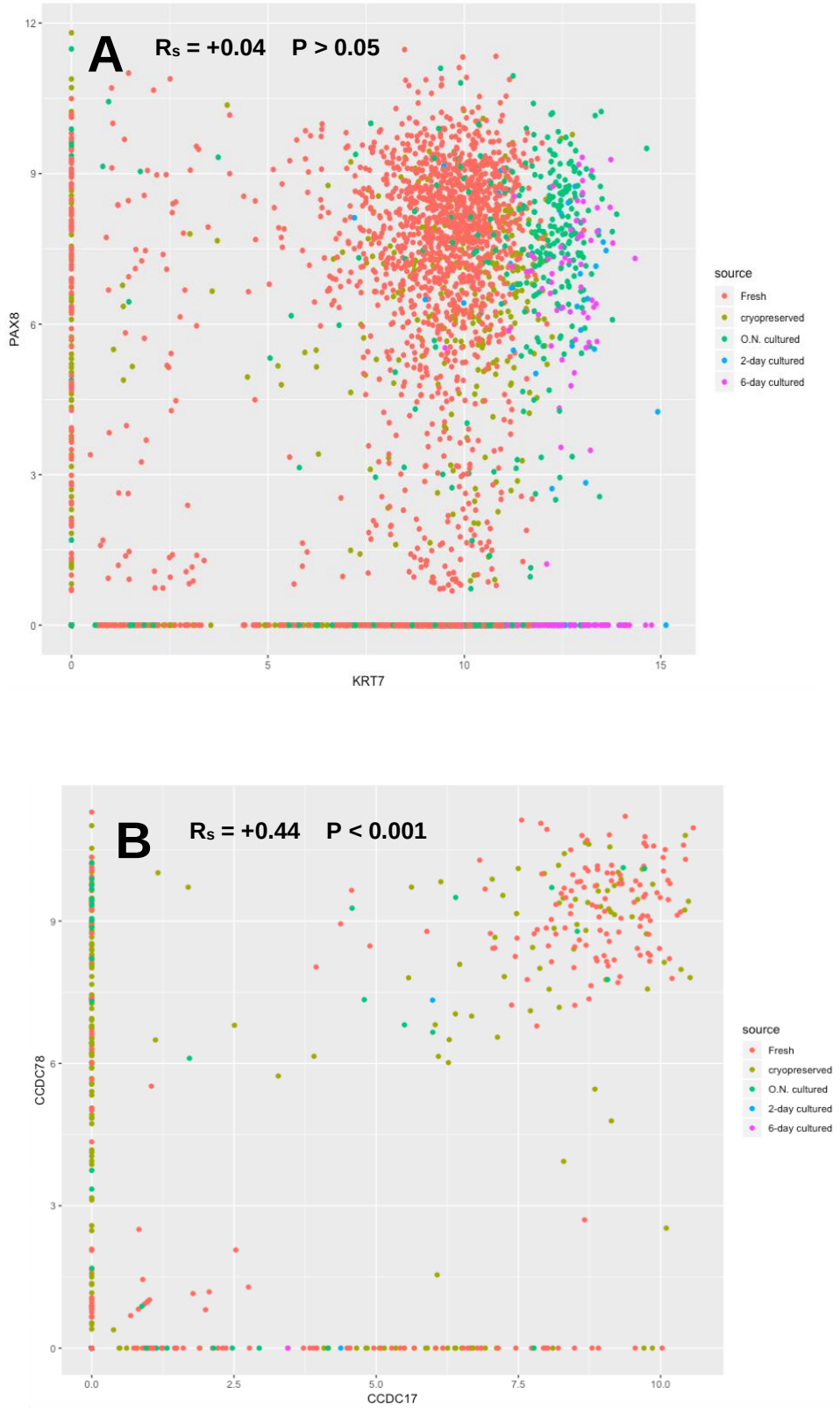


Figure 3.6: Correlation plots for FTE cell marker genes. Plot A depicts a correlation plot for KRT7 and PAX8 expression (logcounts), whereas plot B depicts a correlation plot for CCDC17 and CCDC78 expression. A cluster of cells found at the top right of the plot is a visual indicative for co-expression, but not an indicative for positive correation. Results from spearman correlation test are presented in for each plot. Dots which are either on the X or Y axis are cells that have at least one of the genes from the corresponding correlation plot undetected.

3.4. Evaluation of cell cycle and upregulated fresh-cell genes

To evaluate the presence of cell cycle changes caused by either cryopreservation of culturing. G1 phase-associated genes were analysed. As depicted in figure 3.7, the expression of CCNE1, NXF1, TRMT2A and DNAJB6 genes were observed. Despite the statistical significance, the overall expression of these genes is stable across the different cell sources with minor variations. CCNE1 is expressed lowest in 2-day cultured cells, however the difference between fresh and 2-day cultured cells in CCNE1 expression has statistically little significance ($p > 0.05$).

The average expression for NXF1 is highest in fresh cells, however this gene is considerably expressed in all cell preservation methods. TRMT2A is detected highest in fresh and cryopreserved cells and DNAJB6 is detected more in cryopreservation treated cells.

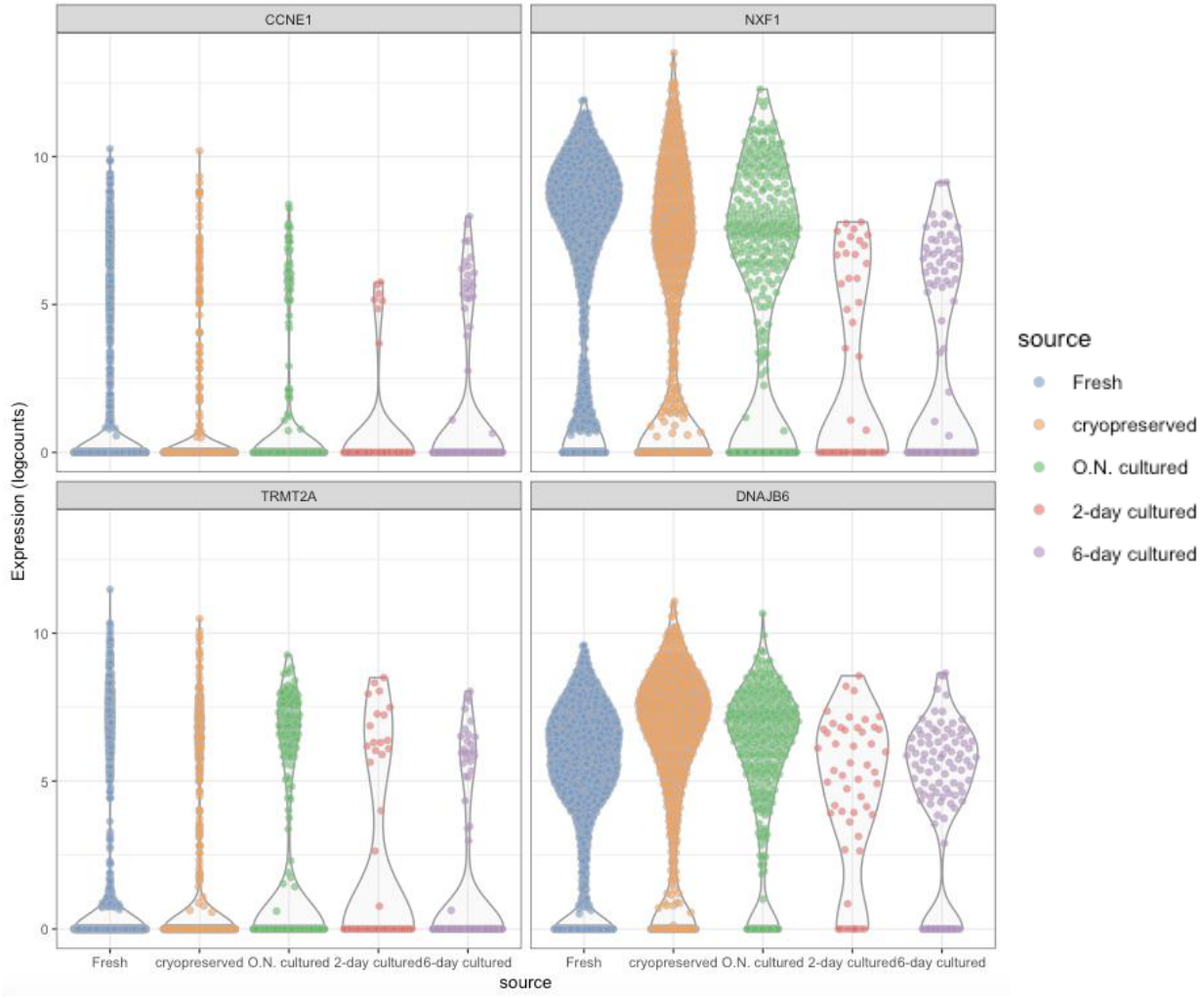


Figure 3.7: Expression plots for G1/S phase genes. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1

G2/M phase-associated genes were analysed similarly as seen in figure 3.8. Where also these genes are equally expressed across the different cell sources, however with negligible dissimilarities. In a similar trend, CCNF and CDCA8 is expressed most in cryopreserved cells and least in O.N. and 2-day cultured cells. TOP2A is slightly more predominant in fresh cells and less predominant in 2-day cultured cells. CDK1 expression is lowest in 2-day and O.N. cultured cells but is stable among other cell sources.

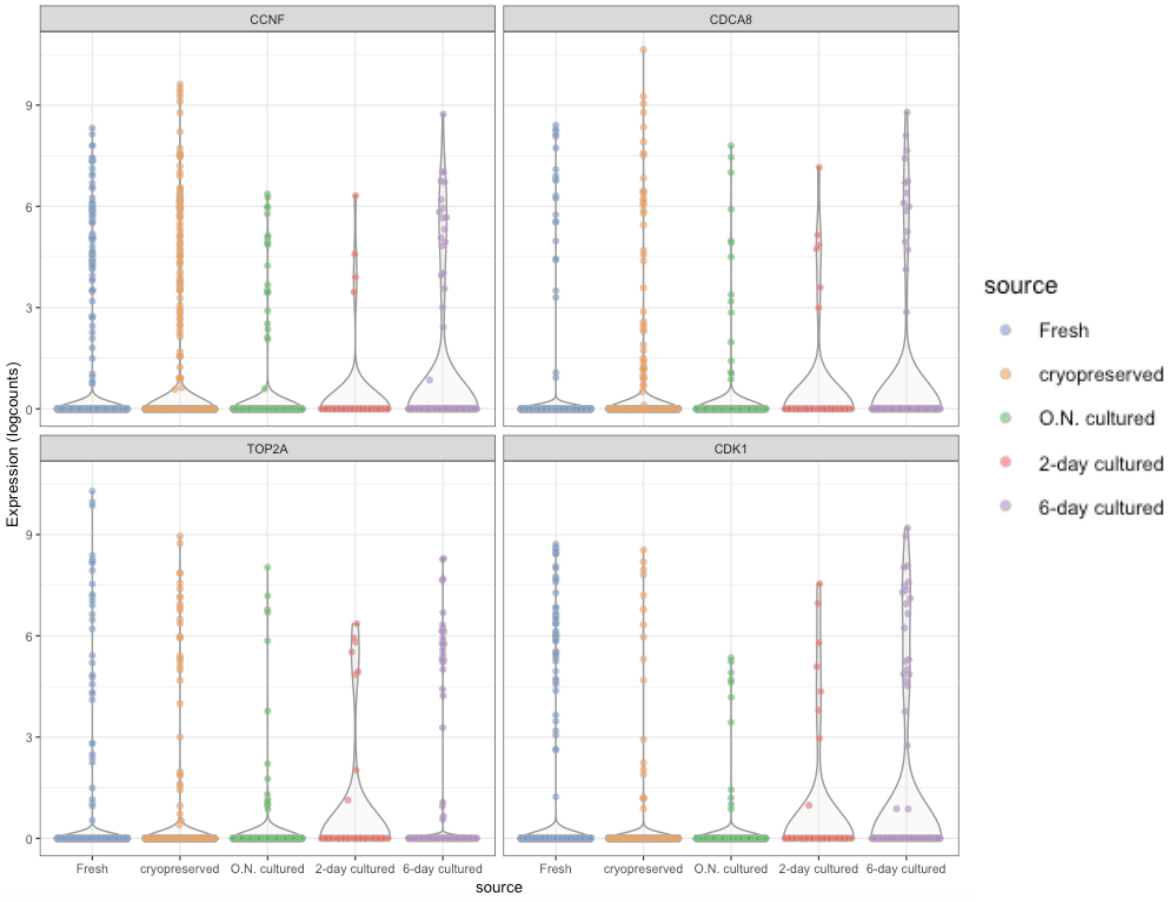


Figure 3.8: Expression plots for G2/M phase genes. Overall these genes are scarcely detected within the cell populationl. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1

Selected genes which were detected as upregulated in fresh cells were analysed through an expression plot, illustrated in figure 3.9. These plots were used to compare the expression distribution across the different cell preservation methods. Where FOS and JUN and EGR1 are expressed in a similar pattern for all preservation technique, which might suggest the possibility of a common functional pathway between these genes.

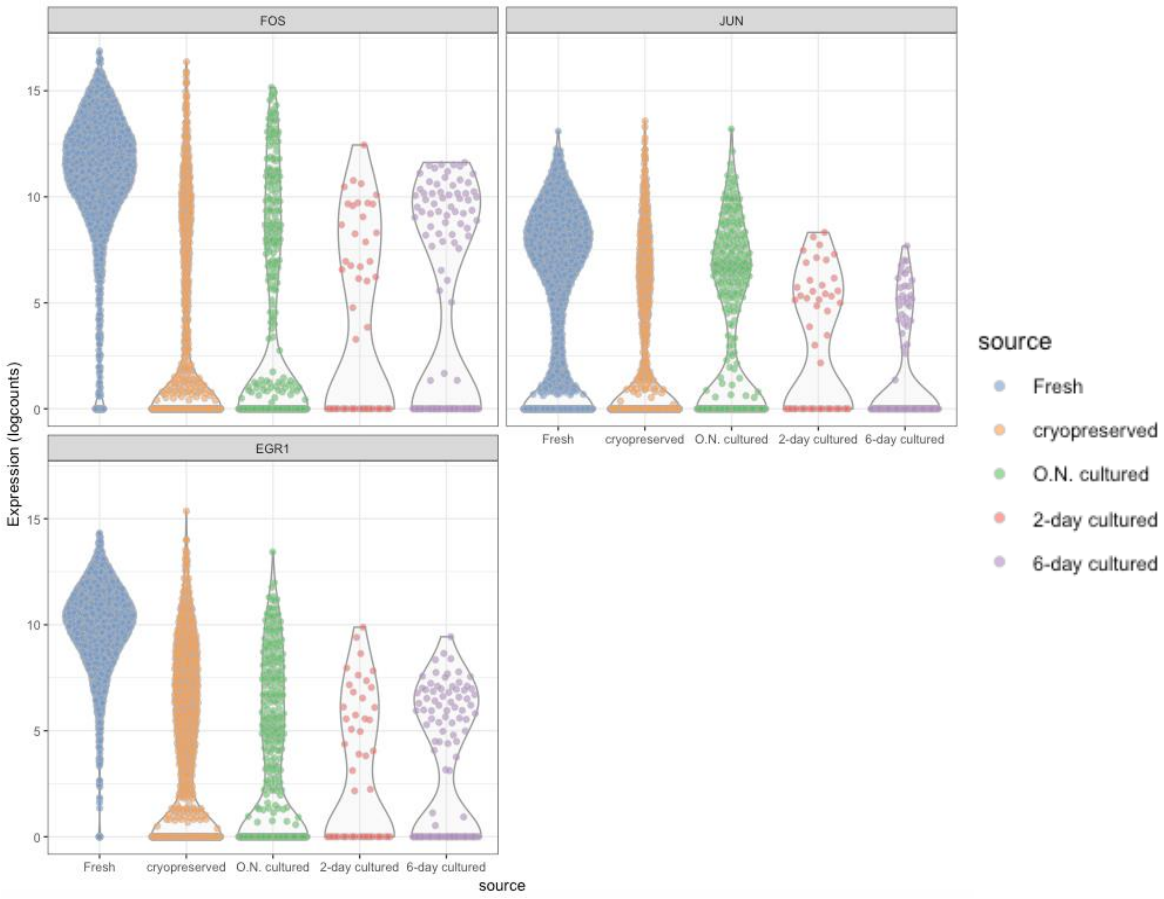


Figure 3.9: Expression plots for fresh cell upregulated genes. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1

3.5. Examination of cancer associated pathways in FTE

All cells have been obtained from OC patients and EC patients, therefore it was hypothesized that these cells might be expressing specific gene which could be cancer associated. Expression of TP53 gene and SOX2-OT were observed for each patient and for each source origin, as illustrated in figure 3.10 and 3.11. Minor expression changes of SOX2-OT were detected across each patient and with no statistical significance ($p>0.05$). On the other hand, TP53 expression was detected highly variable

across each patient. Which might suggest that some patients can be presenting overaccumulation of p53 as well. TP53 expression is more abundant in fresh, 2-day cultured and 6-day cultured cells, however this diverse expression across cell sources is most likely due to the transcriptome heterogeneity between patients.

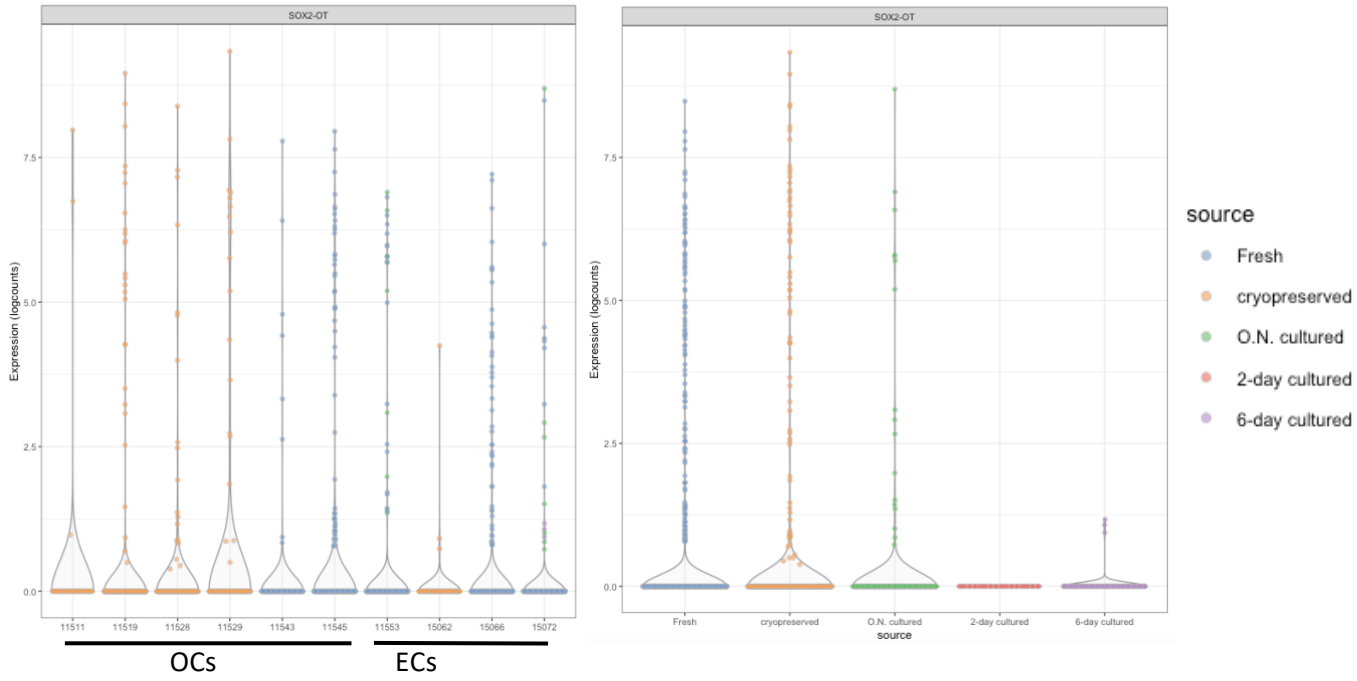


Figure 3.10: SOX2-OT expression plots across cell source and patient. Left plot depicts the expression across patient, where each column represents a patient. The Patients ID assigned to each column are the following: 11511, 11519, 11528, 11529, 11543, 11545, 11553, 15062, 15066, 15072, detailed information of patients is found in figure S3.3 Lines below the left plot indicate the column patients that either present Ovarian Cancer (OC) or Endometrial Cancer (EC). Right plot illustrates the expression across each type of cell source. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1

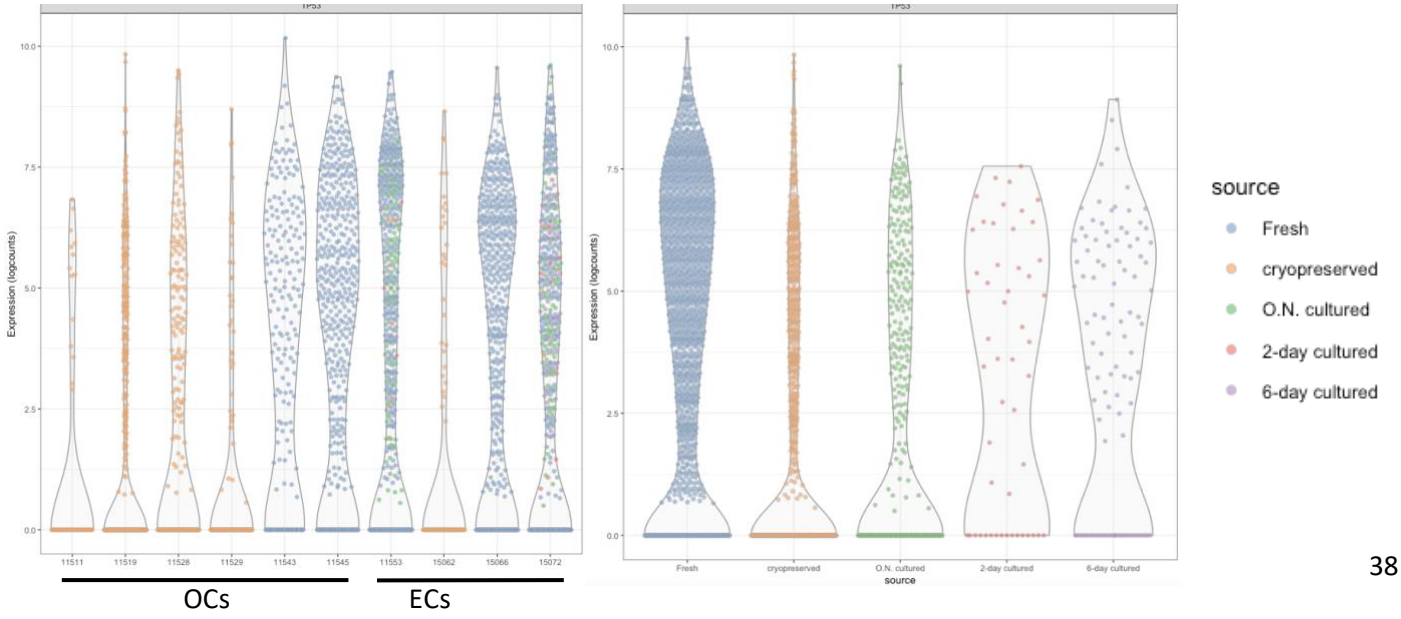


Figure 3.11: TP53 expression plots across cell source and patient. Left plot depicts the expression across patient, where each column represents a patient. As in figure 4.9, the Patients ID assigned to each column are the following: 11511, 11519, 11528, 11529, 11543, 11545, 11553, 15062, 15066, 15072. Lines below the left plot indicate the column patients that either present Ovarian Cancer (OC) or Endometrial Cancer (EC). Right plot illustrates the expression across each type of cell source. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1.

Furthermore, the presence of the SNS-related OC pathway was examined in the FTE cells as seen in figure 3.12. Expression plots for FOSB, IL8 and ADRB2 were generated and analysed to investigate its expression variability across the preservation methods. FOSB was highly predominant if fresh cells. Whereas IL8 expression was more abundant in cryopreserved cells, but constant across all cell populations. ADRB2 was detected in approximately 11% of the total sequenced cells, where its expression was is similarly for each source type.

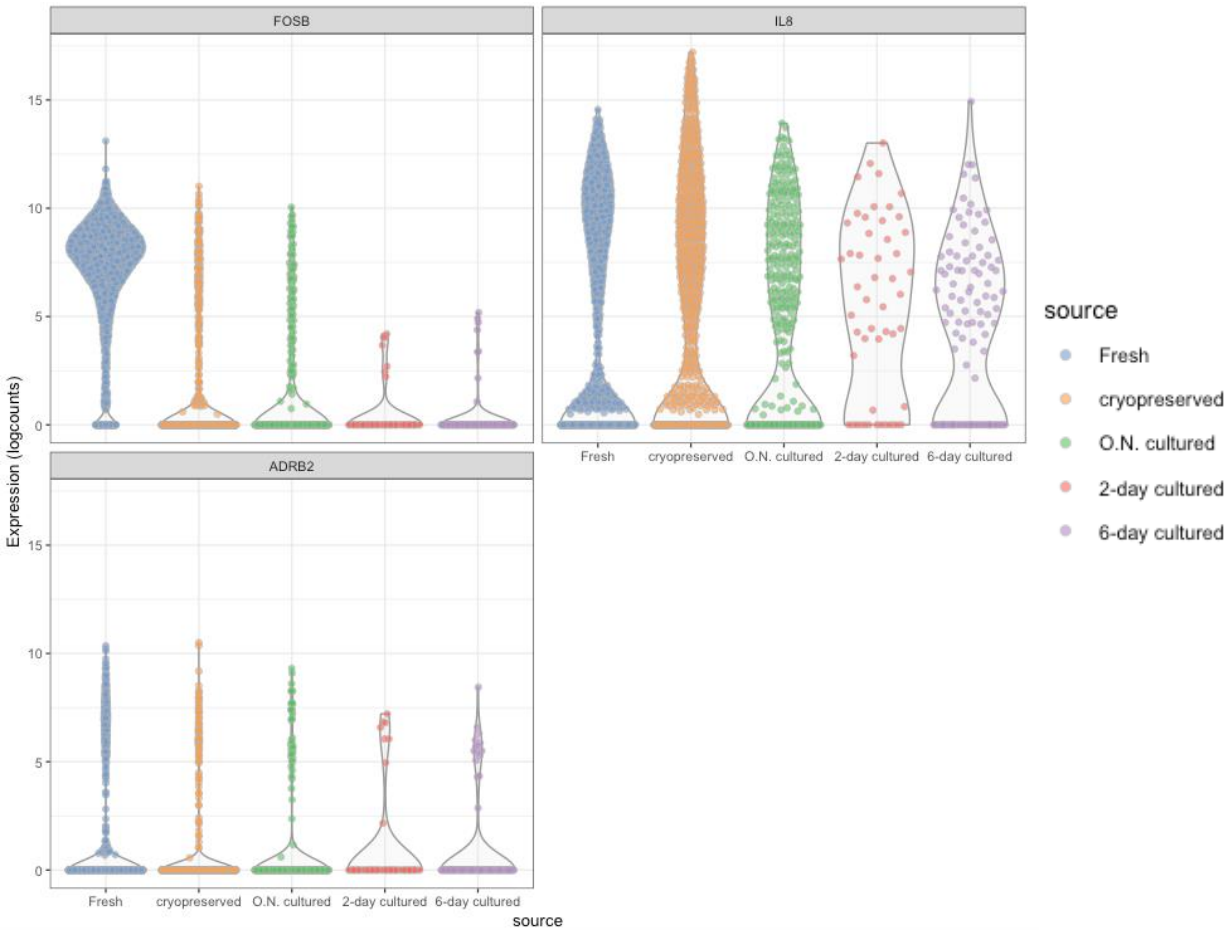
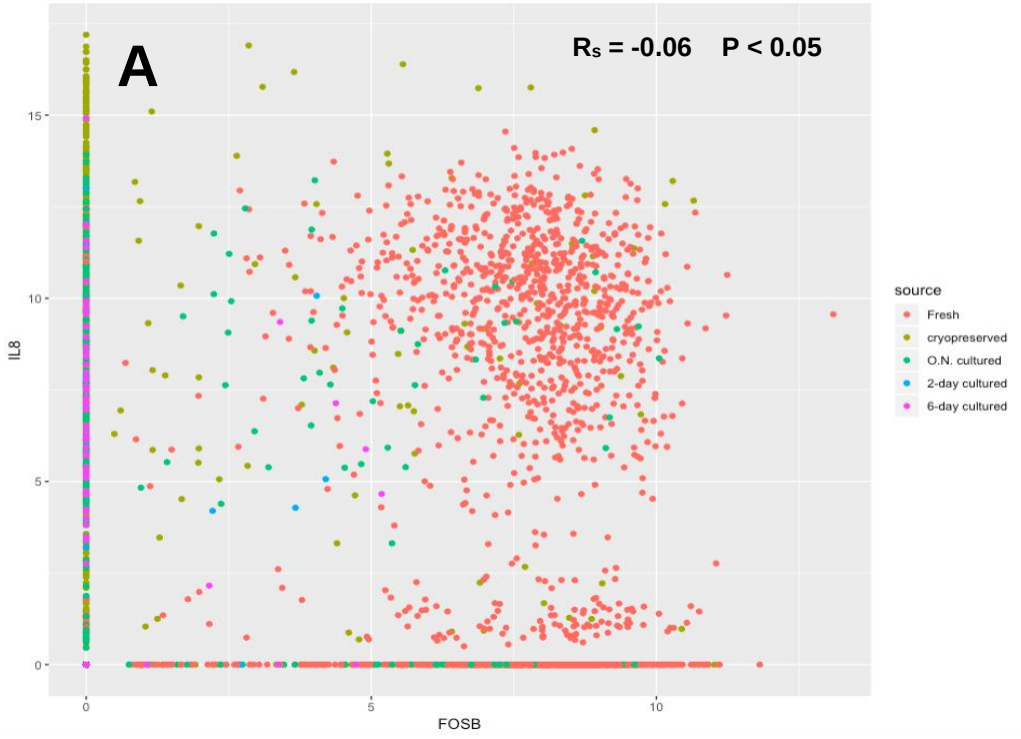


Figure 4.12: Expression plots SNS-related OC pathway genes. Wilcoxon test results comparing the difference between fresh cells to preservation types is presented in tables S3.1

After detecting the expression of these genes, the correlation and co-expression between these genes were examined. Two correlation plots were produced to analyse the correlation of FOSB with IL8 and ADRB2. As a result, IL8 and ADRB2 was identified to have moderate co-expression. However, as seen in figure 3.13, the estimated correlation coefficient for FOSB and IL8 was weak (r_s = -0.06, p<0.05) and the correlation for FOSB and ADRB2 was not significant (r_s = -0.01, p>0.05).



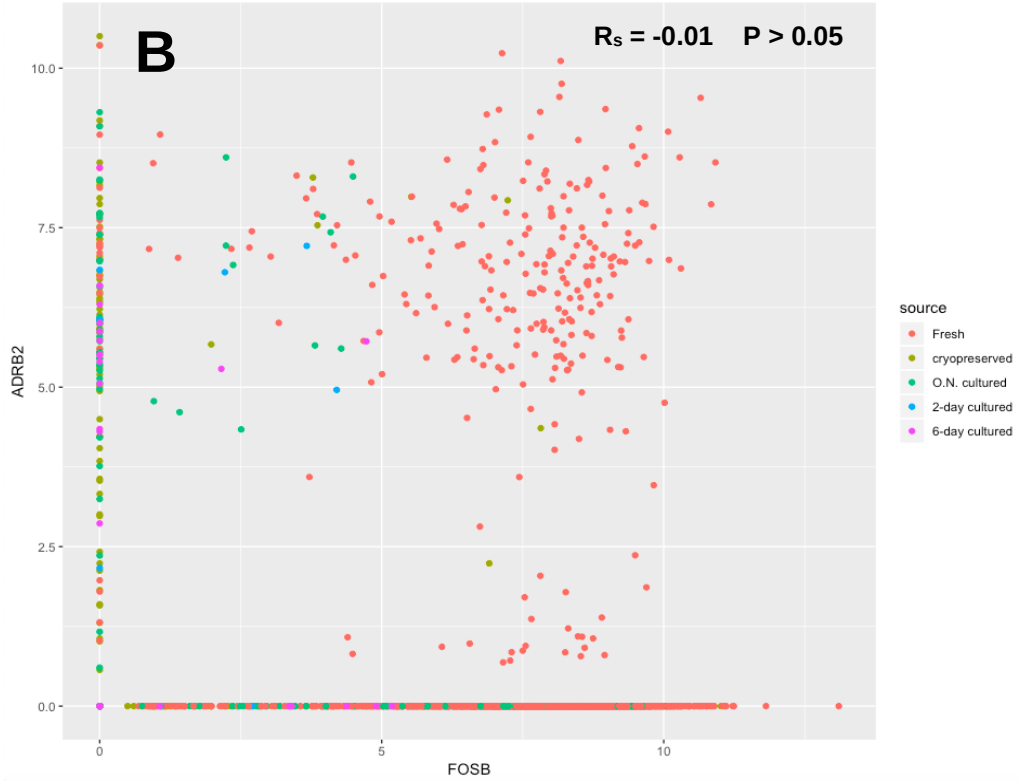


Figure 4.13: Correlation plots for SNS-related OC pathway genes. Plot A depicts a correlation plot for FOSB and IL8 expression (logcounts), whereas plot B depicts a correlation plot for ADRB2 and FOSB expression. A cluster of cells found at the top right of the plot is a visual indicative for co-expression, but not an indicative for positive correlation. Results from spearman correlation test are presented in for each plot. Dots which are either on the X or Y axis are cells that have at least one of the genes from the corresponding correlation plot undetected

4. Discussion

4.1 Summary of results

Overall, results show that cell preservation methods do induce some changes in gene expression. Some of these changes include alterations of FTE cell markers, and other specific genes such as FOS and EGR1. However, none of the implemented cell preservation processes drastically affected the expression levels of cell cycle associated genes or altered the expression of either TP53 or SOX2-OT oncogene. Regardless, the RStudio libraries simplified the process of interpreting the dense data generated from scRNA-seq.

4.2 Results interpretation

4.2.1 Heterogeneity across cell preservation types.

From the generated results, it was detected that 60% of the analysed genes were expressed differently between fresh cells and cryopreserved cells. Suggesting that cryopreservation should have an impactful effect in reprogramming FTE cells. The effects of cryopreservation have been studied using different cell types. For instance, the cryopreservation of B-cells has been documented, where these cells present transcriptomal changes during preservation [69]. However, this same paper had reported an upregulation of the FOSB gene during cryopreservation, which is contrary to our results, since FOSB was detected highest in fresh cells. This could be due to procedure differences of cryopreservation between experiments, where differences in thawing periods or cryopreservation processes will vary depending on the experimental condition.

Since cryopreservation is shown to alter genetic expression, Single-cell transcriptome analyses rely on processing disassociated cells from freshly extracted tissues, for better data acquisition and reliability [40, 53]. However, this adds extra difficulty in studies that have complex experimental setups in which tissue disassociation is not easily achievable, such as neuronal and bone marrow tissue [70,71]. As a consequence, different methods of cryopreservation have been proposed for scRNA-seq to minimise the reprogramming effects of cryopreservation [72,73].

Expression from cultured cells presented a lower number of DE genes compared to cryopreserved cells. A steady but linear increase or decrease of DE genes across the different periods of culturing was expected, however a fluctuation was recorded instead, where the number of DE genes was lowest in 2-day cultured cells. This could be due to the experimental procedure when single cell transcriptome reads were acquired. Where only 51 2-day cultured cells were sequenced presenting good quality reads, since the average number of detected genes per cell was at 3,665 genes, the low number of DE genes found in 2-day cultured cells could be due to the low gene detection induced by the small number of sequenced cells from this experimental condition. There is enough evidence in the literature that report transcriptome alterations caused by cell culturing [74,75]. Which is expected, since changes in the cell microenvironment are highly likely to induce epigenetic modifications [76].

From the T-SNE plots, the presence of high heterogeneity was observed within cryopreserved cell population. This level of heterogeneity was due to the formation of cell clusters depicted in the plot, where most of these clusters also shared one or two batches in common. This suggest that the batch origin for sequencing might have an impact in causing the perceived heterogeneity across the cryopreserved population. Technical batch effects are a common issue found in high throughput scRNA-seq [43,77]. Batch effect happens when two same experiments are conducted in different batches or plates, leading to two variable results. The reason behind why batch effect is greater in

cryopreserved cells, is due to the fact that these cells were overnight cultured for recovery, which is an additional experimental step compared to the fresh cells which have been sequenced right after tissue disassociation. The addition of this experimental step could make the procedure of sequencing cryopreserved cell more prone to technical batch effects.

4.2.2 Expression changes in FTE cell markers

It is known in the literature that secretory FTE cells are $KRT7^+$ and $PAX8^+$, whilst ciliated FTE cells are $CCDC78^+$ and $CCDC17^+$ [67,68]. Furthermore, from the correlation plots depicting the FTE cell markers, secretory cell markers are highly co-expressed, whereas ciliated cell markers are co-expressed in a positive correlation. However, PAX8 itself is a reliable marker for secretory FTE cells, whereas CCDC78 alone can be used to detect ciliated FTE cells [78,79]. From the recorded results, it is known that KRT7 expression is slightly increased in 2-day and 6-day cultured cells, PAX8 expression is diminished in during cryopreservation and CCDC78 expression along with CCDC17 expression is scarce in 2-day and 6-day cultured cells. Therefore, there is a possibility that secretory cells are less detected in cryopreserved cells but more detected in 2-day and 6-day cultured cells. Whereas ciliated cells are almost absent in 2-day and 6-day cultured cells. These findings show that preservation methods impact the constitution of FTE, which is a significant fact to consider for any study which is interested in analysing the cellular population of FTE.

4.2.3 Cell cycle genes and upregulated fresh-cell genes

No significant change of expression in G1/S associated genes was presented among all cell sources, G1 phase related genes were expected to be highly expressed in cryopreserved cells, since these cells might be undergoing cell growth during cell recovery after the induced stressed by cryopreservation. However, DNAJB6 is a G1 phase related gene in which presents is known to reduce caspase-3 activity, inhibiting the cell from apoptosis [80].

Regardless, expression of G2/M associated genes are transiently expressed and less significantly affected by different preservation methods, which it is expected, since preservation methods are not known to inhibit DNA and cell replication. Information in the literature related to the effect of cell cycle is significantly scarce. Where the effects of cryopreserving early-mouse embryos in different cell cycle stages have been observed in years 1991 and 1992 [81,82].

On the other hand, FOS, JUN and EGR1 gene were found highly predominant in fresh cells and were expressed similarly across all presentation types. From the observed expression similarity, it can be theorised that these three genes are involved in a common pathway. In fact, FOS JUN and EGR1 are upregulated during early stress responses. FOS and JUN genes synthetize the c-Fos and JunD transcription factors respectively, which work in conjunction to prevent stress-induced apoptosis [83]. When EGR1 gene is upregulated, JNK1 recruited for DNA damage repairs [84]. From these findings, it can be postulated that fresh cells have experienced some form of DNA damaging stress before sequencing. The origin of this stress could be from UV light or from inadequate manipulation of sample. However, the cause of the detected cellular stress remains unknown.

4.2.4 Cancer associated pathways in FTE

SOX2 overexpression is known to be detected in secretory FTE cells on patients with HGSOC [30]. Furthermore, SOX2 overexpression in normal FTE can be an indictive for HGSOC recurrence. However, in the database, there was no detected SOX2 transcript, but instead the data had quantified the number of detected SOX2 overlapping transcript (SOX2-OT) reads. Knowing that expression of SOX2 and SOX2-OT has shown to be concordant in breast cancer [85]. It was expected that high expression of SOX2-OT across most OC patients would be detected in FTE. However, SOX2-OT was detected in only 5% from all the cells of fallopian tube. It is known that SOX2-OT expression play an important role in the regulation of SOX2 gene in breast cancer [85].

Furthermore, SOX2-OT is upregulated in ovarian cancer tumours and plays a role in cell migration and tumour invasion [86].

Expression of TP53 was presented widely across most patients, it is known that TP53 overexpression is a known phenotype in STIC found in FTE [25,26]. However, the expression of TP53 could not be interpreted as either normal or pathogenic across the patients, since no control was present which could be used to compare the detected TP53 expression in OC and EC patients with the normal physiological expression level of TP53.

The expression analyses of FOSB, IL8 and ADRB2 was detected in FTE cells, however no significant correlation was found between these three genes. The detection of these genes in FTE signifies there is evidence showing that HGSOC cancers which arise from the FTE cells can upregulate the expression of FOSB, IL8 and ADRB2 to take advantage of the neural stress pathway associated with OC progression [32]. However, the independent expression of FOSB, IL8, and ADRB2 genes demonstrate that these genes present separate functions in FTE cells.

4.3 Limitations

Besides collecting some significant data, limitations were presented during the study. Firstly, the high expression of FOS, JUN and EGR1 in fresh cells was indicative that the FTE cells were subjected to DNA damaging stress. This can also be an indicative of issues experienced when attempting to maintain the sample integrity of Fallopian tube epithelium. However, it is unknow how the samples were treated after the extraction of these in the hospital.

Smart-seq2 is an optimised scRNA-seq protocol for the analysis of isoform transcripts. However, it possesses some limitations, due to its process, smart-seq2 protocol has preferential amplification and sequencing for transcripts which are highly abundant, which limit the detection rare and

transients' transcripts. Furthermore, Smart-seq2 presents a particular limitation for quantifying the number of detected transcripts. After the PCR of all cDNA generated libraries for each cell, not all cDNA fragments are sequenced, potentially leading to gene dropouts. However, these dropouts could also happen by the limited RNA capture efficiency, which can vary randomly [87].

Most importantly, regardless of the detection of batch effect, no batch correction libraries were implemented onto the data frame. Which is important to consider, since the results obtained would radically differ if batch corrected values were implemented onto the database and potentially different conclusions would be made from the data [88].

4.4 Future Directions

Short term future directions of this project will consist in comparing the findings of this paper when the database is batch corrected, this batch correction would be achieved by R packages such as limma. However, for the longer term, this research will focus on the advancement of scRNA-seq methodologies. The closest advancement is the use of isolated cellular nuclei for scRNA-seq [89]. Sequencing isolated nuclei would aid in the scRNA-seq analyses of cell with complex morphologies, such as the analysis of axonal neurones. Furthermore, single nuclear RNA sequencing includes diminished dissociation bias and better compatibility with frozen samples. This is possible due to the ability of nuclear RNA sequencing of removing cytoplasmic transcriptional stress responses [90].

4.5 Conclusion

In conclusion, our results show that the preservation methods do induce changes in the transcriptome, where the level of induced heterogeneity is highest when cells are cryopreserved. These changes in the transcriptome vary from the alteration of epithelial cell markers presented in FTE, to alteration of specific genes such as DNAJB6. However, cell preservation methods do not alter the expression cell cycle genes or the expression of tumour suppressor genes and oncogenes,

such as TP53 and SOX2-OT respectively. The process of cryopreservation has shown to alter the cellular transcriptome; therefore, cryopreservation methods are being optimised to minimise the alterations on the transcriptome. Induction of heterogeneity is not only achieved by cell preservation, since technical noise such as batch effect, gene drop outs and RNA capture efficiency can alter the detection of heterogeneity. Future directions related to single cell transcriptome analyses, will be focused on optimising the current methods of scRNA-seq.

References

1. Cancer Research UK, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Zero, [March] [2019]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: a cancer journal for clinicians. 2015 Jan;65(1):5-29.
3. Labidi-Galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M, Bhattacharya R, Novak M, Jones S, Phallen J, Hruban CA. High grade serous ovarian carcinomas originate in the fallopian tube. Nature communications. 2017 Oct 23;8(1):1093.
4. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, Reding DJ, Greenlee RT, Yokochi LA, Kessel B, Crawford ED. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. Jama. 2011 Jun 8;305(22):2295-303.
5. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, Amso NN, Apostolidou S, Benjamin E, Cruickshank D, Crump DN. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. The Lancet. 2016 Mar 5;387(10022):945-56.
6. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011 Jun;474(7353):609.
7. Lynch HT, Casey MJ, Snyder CL, Bewtra C, Lynch JF, Butts M, Godwin AK. Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management. Molecular oncology. 2009 Apr 1;3(2):97-137.
8. Lynch HT, Snyder C, Casey MJ. Hereditary ovarian and breast cancer: what have we learned?. Annals of oncology. 2013 Nov 1;24(suppl_8):viii83-95.
9. Fathalla MF. Incessant ovulation—a factor in ovarian neoplasia. Lancet. 1971 Jul 17;2(7716):163.
10. Luan NN, Wu QJ, Gong TT, Vogtmann E, Wang YL, Lin B. Breastfeeding and ovarian cancer risk: a meta-analysis of epidemiologic studies. The American journal of clinical nutrition. 2013 Aug 21;98(4):1020-31.
11. Luan NN, Wu QJ, Gong TT, Vogtmann E, Wang YL, Lin B. Breastfeeding and ovarian cancer risk: a meta-analysis of epidemiologic studies. The American journal of clinical nutrition. 2013 Aug 21;98(4):1020-31.

12. Havrilesky LJ, Gierisch JM, Moorman PG, Coeytaux RR, Urrutia RP, Lowery WJ, Dinan M, McBroom AJ, Wing L, Musty MD, Lallinger KR. Oral contraceptive use for the primary prevention of ovarian cancer. Evidence report/technology assessment. 2013 Jun(212):1.
13. Schildkraut JM, Schwingl PJ, Bastos E, Evanoff A, Hughes C. Epithelial ovarian cancer risk among women with polycystic ovary syndrome. Obstetrics & Gynecology. 1996 Oct 1;88(4):554-9.
14. Soini T, Hurskainen R, Grénman S, Mäenpää J, Paavonen J, Pukkala E. Cancer risk in women using the levonorgestrel-releasing intrauterine system in Finland. Obstetrics & Gynecology. 2014 Aug 1;124(2):292-9.
15. Beatty MN, Blumenthal PD. The levonorgestrel-releasing intrauterine system: safety, efficacy, and patient acceptability. Therapeutics and clinical risk management. 2009;5:561.
16. Cramer DW, Welch WR. Determinants of ovarian cancer risk. II. Inferences regarding pathogenesis. Journal of the National Cancer Institute. 1983 Oct 1;71(4):717-21.
17. McCartney CR, Eagleson CA, Marshall JC. Regulation of gonadotropin secretion: implications for polycystic ovary syndrome. InSeminars in reproductive medicine 2002 (Vol. 20, No. 04, pp. 317-326). Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662.
18. Schiffenbauer YS, Abramovitch R, Meir G, Nevo N, Holzinger M, Itin A, Keshet E, Neeman M. Loss of ovarian function promotes angiogenesis in human ovarian carcinoma. Proceedings of the National Academy of Sciences. 1997 Nov 25;94(24):13203-8.
19. Wang J, Luo F, Lu JJ, Chen PK, Liu P, Zheng W. VEGF expression and enhanced production by gonadotropins in ovarian epithelial tumors. International journal of cancer. 2002 Jan 10;97(2):163-7.
20. Landen Jr CN, Birrer MJ, Sood AK. Early events in the pathogenesis of epithelial ovarian cancer. Journal of Clinical Oncology. 2008 Feb 20;26(6):995-1005.
21. Kurman RJ, Shih IM. The Origin and pathogenesis of epithelial ovarian cancer-a proposed unifying theory. The American journal of surgical pathology. 2010 Mar;34(3):433.
22. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer biology & medicine. 2017 Feb;14(1):9.
23. Piek JM, Van Diest PJ, Zweemer RP, Jansen JW, Poort-Keesom RJ, Menko FH, Gille JJ, Jongsma AP, Pals G, Kenemans P, Verheijen RH. Dysplastic changes in prophylactically removed Fallopian

tubes of women predisposed to developing ovarian cancer. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland. 2001 Nov;195(4):451-6.

24. Callahan MJ, Crum CP, Medeiros F, Kindelberger DW, Elvin JA, Garber JE, Feltmate CM, Berkowitz RS, Muto MG. Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. Journal of Clinical Oncology. 2007 Sep 1;25(25):3985-90.
25. Soussi T, Ishioka C, Claustres M, Béroud C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. Nature Reviews Cancer. 2006 Jan;6(1):83.
26. Kuhn E, Kurman RJ, Vang R, Sehdev AS, Han G, Soslow R, Wang TL, Shih IM. TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma—evidence supporting the clonal relationship of the two lesions. The Journal of pathology. 2012 Feb;226(3):421-6.
27. Przybycin CG, Kurman RJ, Ronnett BM, Shih IM, Vang R. Are all pelvic (nonuterine) serous carcinomas of tubal origin?. The American journal of surgical pathology. 2010 Oct 1;34(10):1407-16.
28. Many Ovarian Cancers May Start in Fallopian Tubes [Internet]. National Cancer Institute. 2019 [cited 27 March 2019]. Available from: https://www.cancer.gov/news-events/cancer-currents-blog/2017/ovarian-cancer-fallopian-tube-origins
29. Perets R, Wyant GA, Muto KW, Bijron JG, Poole BB, Chin KT, Chen JY, Ohman AW, Stepule CD, Kwak S, Karst AM. Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. Cancer cell. 2013 Dec 9;24(6):751-65.
30. Hellner K, Miranda F, Chedom DF, Herrero-Gonzalez S, Hayden DM, Tearle R, Artibani M, KaramiNejadRanjbar M, Williams R, Gaitskell K, Elorbany S. Premalignant SOX2 overexpression in the fallopian tubes of ovarian cancer patients: Discovery and validation studies. EBioMedicine. 2016 Aug 1;10:137-49.
31. Thaker PH, Han LY, Kamat AA, Arevalo JM, Takahashi R, Lu C, Jennings NB, Armaiz-Pena G, Bankson JA, Ravoori M, Merritt WM. Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma. Nature medicine. 2006 Aug;12(8):939.
32. Shahzad MM, Arevalo JM, Armaiz-Pena GN, Lu C, Stone RL, Moreno-Smith M, Nishimura M, Lee JW, Jennings NB, Bottsford-Miller J, Vivas-Mejia P. Stress effects on FosB-and interleukin-8 (IL8)-driven ovarian cancer growth and metastasis. Journal of Biological Chemistry. 2010 Nov 12;285(46):35462-70.

33. Zink KE, Dean M, Burdette JE, Sanchez LM. Imaging Mass Spectrometry Reveals Crosstalk between the Fallopian Tube and the Ovary that Drives Primary Metastasis of Ovarian Cancer. ACS central science. 2018 Oct 9;4(10):1360-70.
34. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences. 1977 Dec 1;74(12):5350-4.
35. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences. 1977 Dec 1;74(12):5350-4.
36. Becker-Andre M, Hahlbrock K. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a P CR aided t ranscipt t itration assay (PATTY). Nucleic acids research. 1989 Nov 25;17(22):9437-46.
37. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 1991 Jun 21;252(5013):1651-6.
38. Nelson NJ. Microarrays have arrived: gene expression tool matures. Journal of the National Cancer Institute. 2001 Apr 4;93(7):492-4.
39. Affymetrix I. Statistical algorithms description document. Technical paper. 2002;62:110.
40. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K. mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods. 2009 Apr 6;6(5):377.
41. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, Schwartz S. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature. 2014 Jun;510(7505):363.
42. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. science. 2010 Jan 29;327(5965):542-5.
43. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental & molecular medicine. 2018 Aug 7;50(8):96.
44. Staszewski RO. Cloning by limiting dilution: an improved estimate that an interesting culture is monoclonal. The Yale journal of biology and medicine. 1984 Nov;57(6):865.

45. Guo F, Li L, Li J, Wu X, Hu B, Zhu P, Wen L, Tang F. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell research. 2017 Aug;27(8):967.
46. McCarthy RC, Breite AG, Green ML, Dwulet FE. Tissue dissociation enzymes for isolating human islets for transplantation: factors to consider in setting enzyme acceptance criteria. Transplantation. 2011 Jan 27;91(2):137.
47. Carvalho PP, Gimble JM, Dias IR, Gomes ME, Reis RL. Xenofree enzymatic products for the isolation of human adipose-derived stromal/stem cells. Tissue Engineering Part C: Methods. 2013 Jan 31;19(6):473-8.
48. He M, Huang H, Wang M, Chen A, Ning X, Yu K, Li Q, Li W, Ma L, Chen Z, Wang X. Fluorescence-Activated Cell Sorting Analysis of Heterotypic Cell-in-Cell Structures. Scientific reports. 2015 Apr 27;5:9588.
49. Esteve-Codina A, Arpi O, Martinez-García M, Pineda E, Mallo M, Gut M, Carrato C, Rovira A, Lopez R, Tortosa A, Dabad M. A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. PloS one. 2017 Jan 25;12(1):e0170632.
50. Datta S, Malhotra L, Dickerson R, Chaffee S, Sen CK, Roy S. Laser capture microdissection: Big data from small samples. Histology and histopathology. 2015 Nov;30(11):1255.
51. Marcus JS, Anderson WF, Quake SR. Microfluidic single-cell mRNA isolation and analysis. Analytical chemistry. 2006 May 1;78(9):3084-9.
52. Hedlund E, Deng Q. Single-cell RNA sequencing: technical advancements and biological applications. Molecular aspects of medicine. 2018 Feb 1;59:36-46.
53. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nature protocols. 2014 Jan;9(1):171.
54. Smart-Seq2 - Enseqlopedia [Internet]. Enseqlopedia. 2019 [cited 27 March 2019]. Available from: http://enseqlopedia.com/wiki-entry/rna-sequencing-methods/low-level-rna-detection/smart-seq2/
55. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nature protocols. 2014 Jan;9(1):171.
56. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015 May 21;161(5):1202-14.

57. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. Bioinformatics.babraham.ac.uk. 2019 [cited 27 March 2019]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
58. Bioinformatics - Trim Galore! [Internet]. Bioinformatics.babraham.ac.uk. 2019 [cited 27 March 2019]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
59. Vladimir Kiselev (wikiselev) M. Analysis of single cell RNA-seq data [Internet]. Hemberg-lab.github.io. 2019 [cited 27 March 2019]. Available from: https://hemberg-lab.github.io/scRNA.seq.course/processing-raw-scrna-seq-data.html#trimming-reads
60. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.
61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21.
62. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology. 2016 May;34(5):525.
63. Kessler M, Hoffmann K, Brinkmann V, Thieck O, Jackisch S, Toelle B, Berger H, Mollenkopf HJ, Mangler M, Sehouli J, Fotopoulou C. The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. Nature communications. 2015 Dec 8;6:8989.
64. Mora-Castilla S, To C, Vaezeslami S, Morey R, Srinivasan S, Dumdie JN, Cook-Andersen H, Jenkins J, Laurent LC. Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. Journal of laboratory automation. 2016 Aug;21(4):557-67.
65. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013 Nov 13;30(7):923-30.
66. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017 Jan 14;33(8):1179-86.
67. Ning G, Bijron JG, Yamamoto Y, Wang X, Howitt BE, Herfs M, Yang E, Hong Y, Cornille M, Wu L, Hanamornroongruang S. The PAX2-null immunophenotype defines multiple lineages with common expression signatures in benign and neoplastic oviductal epithelium. The Journal of pathology. 2014 Dec;234(4):478-87.

68. Terré B, Piergiovanni G, Segura-Bayona S, Gil-Gómez G, Youssef SA, Attolini CS, Wilsch-Bräuninger M, Jung C, Rojas AM, Marjanović M, Knobel PA. GEMC1 is a critical regulator of multiciliated cell differentiation. The EMBO journal. 2016 May 2;35(9):942-60.
69. Rasmussen SM, Bilgrau AE, Schmitz A, Falgreen S, Bergkvist KS, Tramm AM, Bæch J, Jacobsen CL, Gaihede M, Kjeldsen MK, Bødker JS. Stable phenotype of B-cell subsets following cryopreservation and thawing of normal human lymphocytes stored in a tissue biobank. Cytometry Part B: Clinical Cytometry. 2015 Jan;88(1):40-9.
70. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015 Mar 6;347(6226):1138-42.
71. Paul F, Arkin YA, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, David E. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell. 2015 Dec 17;163(7):1663-77.
72. Guillaumet-Adkins A, Rodríguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A, Keren-Shaul H. Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome biology. 2017 Dec;18(1):45.
73. Wang W, Penland L, Gokce O, Croote D, Quake SR. High fidelity hypothermic preservation of primary tissues in organ transplant preservative for single cell transcriptome analysis. BMC genomics. 2018 Dec;19(1):140.
74. Januszyk M, Rennert R, Sorkin M, Maan Z, Wong L, Whittam A, Whitmore A, Duscher D, Gurtner G. Evaluating the effect of cell culture on gene expression in primary tissue samples using microfluidic-based single cell transcriptional analysis. Microarrays. 2015 Nov 4;4(4):540-50.
75. Kim T, Echeagaray OH, Wang BJ, Casillas A, Broughton KM, Kim BH, Sussman MA. In situ transcriptome characteristics are lost following culture adaptation of adult cardiac stem cells. Scientific reports. 2018;8.
76. Dey P. Epigenetic changes in tumor microenvironment. Indian journal of cancer. 2011 Oct 1;48(4):507.
77. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2017 Nov 6;19(4):562-78.

78. Perets R, Wyant GA, Muto KW, Bijron JG, Poole BB, Chin KT, Chen JY, Ohman AW, Stepule CD, Kwak S, Karst AM. Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. Cancer cell. 2013 Dec 9;24(6):751-65.
79. Dehring DA, Vladar EK, Werner ME, Mitchell JW, Hwang P, Mitchell BJ. Deuterosome-mediated centriole biogenesis. Developmental cell. 2013 Oct 14;27(1):103-12.
80. Mitra A, Fillmore RA, Metge BJ, Rajesh M, Xi Y, King J, Ju J, Pannell L, Shevde LA, Samant RS. Large isoform of MRJ (DNAJB6) reduces malignant activity of breast cancer. Breast Cancer Research. 2008 Apr;10(2):R22.
81. Chedid S, Abbeel EV, Steirteghem AC. Effects of cryopreservation on survival and development of interphase-and mitotic-stage 1-cell mouse embryos. Human Reproduction. 1992 Nov 1;7(10):1451-6.
82. Bałakier H, Zenzes M, Wang P, MacLusky NJ, Casper RF. The effect of cryopreservation on the development of S-and G 2-phase mouse embryos. Journal of in vitro fertilization and embryo transfer. 1991 Apr 1;8(2):89-95.
83. Zhou H, Gao J, Lu ZY, Lu L, Dai W, Xu M. Role of c-Fos/JunD in protecting stress-induced cell death. Cell proliferation. 2007 Jun;40(3):431-44.
84. Lim CP, Jain N, Cao X. Stress-induced immediate-early gene, egr-1, involves activation of p38/JNK1. Oncogene. 1998 Jun;16(22):2915.
85. Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, Baguley BC, Finlay GJ, Leung EY. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. PloS one. 2014 Jul 9;9(7):e102140.
86. Han L, Zhang W, Zhang B, Zhan L. Long non-coding RNA SOX2OT promotes cell proliferation and motility in human ovarian cancer. Experimental and therapeutic medicine. 2018 Feb 1;15(2):2182-8.
87. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Reviews Genetics. 2013 Sep;14(9):618.
88. Analyzing single-cell RNA-seq data containing read counts [Internet]. Bioconductor.org. 2019 [cited 27 March 2019]. Available from: https://bioconductor.org/packages/release/workflows/vignettes/simpleSingleCell/inst/doc/work-1-reads.html#1_overview

89. Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, Araúzo-Bravo MJ, Lee J, Fishman M, Robbins GE, Lin X. RNA-sequencing from single nuclei. Proceedings of the National Academy of Sciences. 2013 Dec 3;110(49):19802-7.
90. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. Journal of the American Society of Nephrology. 2019 Jan 1;30(1):23-32.

Appendix

A)

Genes	Cryopreserved	O.N. Cultured	2-day cultured	6-day cultured
GAPDH	ns	****	****	****
EPCAM	**	*	****	****
KRT7	****	****	****	****
PAX8	****	ns	**	****
CCDC17	ns	*	ns	*
CCDC78	****	ns	ns	*
CCNE1	****	ns	ns	**
NXF1	****	****	****	****
TRMT2A	ns	****	***	**
DNAJB6	****	****	ns	*
CCNF	****	ns	ns	****
CDCA8	****	****	****	****
TOP2A	**	*	****	****
CDK1	ns	ns	****	****
FOS	****	****	****	****
JUN	****	****	****	****
EGR1	****	****	****	****
SOX2-OT	ns	ns	ns	ns
TP53	****	****	ns	ns
FOSB	****	****	****	****
IL8	***	ns	ns	ns
ADRB2	****	ns	ns	ns

B)

Symbol	level of significance
ns	$P > 0.05$
*	$P \leq 0.05$
**	$P \leq 0.01$
***	$P \leq 0.001$
****	$P \leq 0.0001$

Table S3.1: Wilcox significance level across preservation methods compared to fresh cells. Table B depicts a table legend which defines the meaning of the symbols found in figure A. Interpretations of significance levels are the following: (ns) not significant, (*) little evidence for significance, (**) moderately significant(***) mostly significant and (****) highly significant.

	batch	New.wells	Patient	Sample.Nam	Well	Concernration	CT.GAPDH	Tm1.GAPDH	CT.SOX2	Tm1.SOX2	Tm2.SOX2	Tm3.SOX2	Medium	Note	Sort.date	RT.PCR.date	Nextera.date	Sc
11519L-P1-(	sc01	A04	11519L	11519L-P1	E10	2.46	13.11	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	A07	11519L	11519L-P1	F02	3.87	9.56	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	A08	11519L	11519L-P1	F03	3.97	10.69	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	A10	11519L	11519L-P1	F06	8.82	7.73	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	A11	11519L	11519L-P1	F07	6.55	9.55	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	A12	11519L	11519L-P1	F08	6.2	9.96	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B01	11519L	11519L-P1	F09	2.57	11.34	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B02	11519L	11519L-P1	F10	8.52	8.81	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B03	11519L	11519L-P1	F12	3.55	10.94	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B04	11519L	11519L-P1	G01	2.13	14.82	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B05	11519L	11519L-P1	G02	2.57	13.94	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B07	11519L	11519L-P1	G05	2.41	13.72	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B08	11519L	11519L-P1	G07	2.15	12.52	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B10	11519L	11519L-P1	G10	4.63	9.39	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P1-(	sc01	B11	11519L	11519L-P1	G11	7.14	9.35	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C01	11519L	11519L-P2	C03	2.28	11.34	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C02	11519L	11519L-P2	C04	3.02	10.41	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C03	11519L	11519L-P2	C06	5.65	10.46	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C04	11519L	11519L-P2	C07	3.58	12.22	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C06	11519L	11519L-P2	C09	6	9.13	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C07	11519L	11519L-P2	C10	1.77	13.21	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C09	11519L	11519L-P2	D03	2.08	11.18	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C10	11519L	11519L-P2	D04	1.73	14.16	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	C12	11519L	11519L-P2	D07	3.46	9.65	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D02	11519L	11519L-P2	D12	3.62	10	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D03	11519L	11519L-P2	E01	4.5	10.94	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D04	11519L	11519L-P2	E02	3.09	13.1	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D06	11519L	11519L-P2	E05	3.87	9.89	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D07	11519L	11519L-P2	E06	5.67	9	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D09	11519L	11519L-P2	E09	1.24	13.89	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D10	11519L	11519L-P2	E11	2.96	10.61	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D11	11519L	11519L-P2	E12	3.15	13.26	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	D12	11519L	11519L-P2	F01	4.75	10.21	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	E01	11519L	11519L-P2	F02	2.72	10.63	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	E02	11519L	11519L-P2	F03	2.66	11.92	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P
11519L-P2-(	sc01	F03	11519L	11519L-P2	F06	3.8	10.79	NA	NA	NA	NA	NA	BM2	cell	28/03/2017	03/04/2017	05/04/2017	P

Figure S3.1.: Extract from cell info (sce). This excel provides information details of the cells, such as dates for the implemented experiments and batch origin

Geneid	Chr	Start	End	Strand	Length	rank_counts	n_cells_coun	pct_dropout	is_feature_control	mean_counts	log10_mean_cou	n_cells_by_counts	pct_dropout_l	total_counts	log10_total_counts
LOC643837	LOC643837	chr1;chr1;chr1;chr1;chr	762971;763178;764383;	763155;7632	+;+;+;+;+;+;+	9233	12853	492	89.84100764	FALSE	8.228527212	0.965132397	426	89.01212278	31902	4.503
NOC2L	NOC2L	chr1;chr1;chr1;chr1;chr	879583;880437;880898;	880180;8805	-;-;-;-;-;-;-;-	2800	19997	1762	63.6175924	FALSE	76.02992004	1.886659447	1625	58.08614908	294768	5.469
KLHL17	KLHL17	chr1;chr1;chr1;chr1;chr	895967;896673;897009;	896180;8969	+;+;+;+;+;+;+	2564	7707.5	165	96.59302085	FALSE	0.924941965	0.284417641	150	96.13102915	3586	3.554
PLEKHN1	PLEKHN1	chr1;chr1;chr1;chr1;chr	901877;902084;905657;	901994;9021	+;+;+;+;+;+;+	2436	9751	221	95.43671278	FALSE	2.417590921	0.533720077	198	94.89295847	9373	3.971
HES4	HES4	chr1;chr1;chr1	934342;934906;935072	934812;9349	-;-;-	1040	10915	651	86.55791865	FALSE	3.887799845	0.689113413	583	84.96259995	15073	4.178
ISG15	ISG15	chr1;chr1	948847;949364	948956;9499	+;+	666	18155.5	1175	75.73817881	FALSE	31.6631416	1.514057954	1028	73.48465308	122758	5.089
AGRN	AGRN	chr1;chr1;chr1;chr1;chr	955503;957581;970657;	955753;9578	+;+;+;+;+;+;+	7326	12598	1263	73.92112327	FALSE	7.01857106	0.904096982	1144	70.49264896	27211	4.434
C1orf159	C1orf159	chr1;chr1;chr1;chr1;chr	1017198;1019733;1019	1018367;101	-;-;-;-;-;-;-;-	2104	14295	500	89.67582077	FALSE	12.5439773	1.131746217	435	88.77998452	48633	4.686
TTLL10	TTLL10	chr1;chr1;chr1;chr1;chr	1109286;1109665;1109	1109306;110	+;+;+;+;+;+;+	3129	9116	146	96.98533967	FALSE	1.897085375	0.461961294	135	96.51792623	7355	3.866
SDF4	SDF4	chr1;chr1;chr1;chr1;chr	1152288;1153838;1154	1153184;115	-;-;-;-;-;-;-	2138	19936	2229	53.974809	FALSE	73.27263348	1.870828823	2077	46.42765025	284078	5.453
B3GALT6	B3GALT6	chr1	1167629	1170420	+	2792	11164	354	92.69048111	FALSE	4.573123549	0.746098671	308	92.05571318	17730	4.248
FAM132A	FAM132A	chr1;chr1;chr1;chr1;chr	1177826;1178215;1178	1178026;117	-;-;-;-;-;-;-	1043	8360	61	98.74045013	FALSE	1.086922879	0.3195064	50	98.71034305	4214	3.624
UBE2J2	UBE2J2	chr1;chr1;chr1;chr1;chr	1189292;1191425;1192	1190867;119	-;-;-;-;-;-;-	2437	19580	1723	64.42287838	FALSE	63.70544235	1.810940811	1581	59.2210472	246986	5.39
SCNN1D	SCNN1D	chr1;chr1;chr1;chr1;chr	1215816;1216606;1216	1216046;121	+;+;+;+;+;+;+	3054	12563	316	93.47511873	FALSE	7.43229301	0.925945689	297	92.33943771	28815	4.45
ACAP3	ACAP3	chr1;chr1;chr1;chr1;chr	1227764;1229203;1229	1229088;122	-;-;-;-;-;-;-	3759	9403	202	95.82903159	FALSE	2.007995873	0.478277236	167	95.69254578	7785	3.89
PUSL1	PUSL1	chr1;chr1;chr1;chr1;chr	1243994;1244295;1244	1244100;124	+;+;+;+;+;+;+	1240	15343	607	87.46644642	FALSE	16.84137219	1.251428253	527	86.40701573	65294	4.814
CPSF3L	CPSF3L	chr1;chr1;chr1;chr1;chr	1246965;1247398;1247	1247304;124	-;-;-;-;-;-;-	2629	20969	2216	54.24323766	FALSE	130.5194738	2.118990063	2057	46.94351303	506024	5.704
GLTPD1	GLTPD1	chr1;chr1;chr1	1260143;1262216;1262	1260482;126	+;+;+	2193	12359	480	90.08878794	FALSE	6.478978592	0.87384229	419	89.19267475	25119	4.400
DVL1	DVL1	chr1;chr1;chr1;chr1;chr	1270658;1273357;1273	1271895;127	-;-;-;-;-;-;-	2924	11322.5	510	89.46933719	FALSE	4.799845241	0.763416405	450	88.39308744	18609	4.269
MXRA8	MXRA8	chr1;chr1;chr1;chr1;chr	1288071;1289228;1289	1289009;128	-;-;-;-;-;-;-	2273	11510	175	96.38653727	FALSE	3.504771731	0.653672789	146	96.2342017	13588	4.13
AURKAIP1	AURKAIP1	chr1;chr1;chr1;chr1	1309110;1309380;1310	1309282;130	-;-;-	1147	20730	2822	41.73033244	FALSE	104.7340727	2.024214961	2625	32.29301006	406054	5.608
CCNL2	CCNL2	chr1;chr1;chr1;chr1;chr	1321091;1323157;1323	1322962;132	-;-;-;-;-;-;-	3631	20621	2227	54.01610572	FALSE	98.8225432	1.99922863	2035	47.51096208	383135	5.583
LOC148413	LOC148413	chr1;chr1;chr1	1334910;1335538;1335	1335069;133	+;+;+	1748	13892	366	92.44270081	FALSE	11.25818932	1.088426325	323	91.66881609	43648	4.63
MRPL20	MRPL20	chr1;chr1;chr1;chr1	1337276;1341189;1342	1337636;134	-;-;-;-	733	20301	2899	40.14040884	FALSE	85.54784627	1.937256265	2716	29.94583441	331669	5.520
ANKRD65	ANKRD65	chr1;chr1;chr1;chr1	1353800;1355432;1356	1354929;135	-;-;-;-	2192	18478	1092	77.45199257	FALSE	41.68970854	1.63032319	991	74.43899923	161631	5.208
VWA1	VWA1	chr1;chr1;chr1	1370903;1372307;1374	1371201;137	+;+;+	4659	12260	666	86.24819327	FALSE	6.285530049	0.862461154	592	84.7304617	24369	4.386
ATAD3C	ATAD3C	chr1;chr1;chr1;chr1;chr	1385069;1387426;1387	1386138;138	+;+;+;+;+;+;+	3859	9710	574	88.14784225	FALSE	2.179262316	0.502326362	501	87.07763735	8449	3.926
ATAD3B	ATAD3B	chr1;chr1;chr1;chr1;chr	1407164;1412654;1414	1407469;141	+;+;+;+;+;+;+	2433	13542	805	83.37807144	FALSE	9.482073768	1.020447212	725	81.29997421	36762	4.565
ATAD3A	ATAD3A	chr1;chr1;chr1;chr1;chr	1447523;1447910;1451	1447853;144	+;+;+;+;+;+;+	2807	12848	671	86.14495148	FALSE	7.660046428	0.93752022	597	84.601496	29698	4.472
SSU72	SSU72	chr1;chr1;chr1;chr1;chr	1477053;1479249;1480	1477547;147	-;-;-;-	1303	21171	2869	40.75985959	FALSE	153.1663657	2.187989635	2669	31.15811194	593826	5.773
MIB2	MIB2	chr1;chr1;chr1;chr1;chr	1550795;1551245;1551	1551013;155	+;+;+;+;+;+;+	3574	11321	482	90.04749122	FALSE	4.48052618	0.738822257	441	88.62522569	17371	4.239
CDK11B	CDK11B	chr1;chr1;chr1;chr1;chr	1571100;1571299;1571	1571218;157	-;-;-;-;-;-;-	2527	8788	272	94.3836465	FALSE	1.546040753	0.405865351	244	93.70647408	5994	3.777
SLC35E2B	SLC35E2B	chr1;chr1;chr1;chr1;chr	1592939;1599766;1601	1597458;159	-;-;-;-;-;-;-	6003	10638	582	87.98265538	FALSE	3.547846273	0.657805776	528	86.38122259	13755	4.138
SLC35E2	SLC35E2	chr1;chr1;chr1;chr1;chr	1656277;1658824;1666	1656861;166	-;-;-;-;-;-;-	7297	9528	571	88.20978732	FALSE	2.168171266	0.500808651	519	86.61336085	8406	3.924
NADK	NADK	chr1;chr1;chr1;chr1;chr	1682671;1685006;1685	1684499;168	-;-;-;-;-;-;-	4226	14428	818	83.10964278	FALSE	12.55119938	1.131977735	735	81.04204282	48661	4.687
GNB1	GNB1	chr1;chr1;chr1;chr1;chr	1716725;1718761;1720	1718492;171	-;-;-;-;-;-;-	3163	17910	1758	63.70018584	FALSE	33.94325509	1.543363359	1614	58.36987361	131598	5.119
KIAA1751	KIAA1751	chr1;chr1;chr1;chr1;chr	1884752;1888059;1890	1887289;188	-;-;-;-;-;-;-	4710	7740	104	97.85257072	FALSE	0.960536497	0.292374932	92	97.62703121	3724	3.571
PRKCZ	PRKCZ	chr1;chr1;chr1;chr1;chr	1981909;1986880;1987	1982140;198	+;+;+;+;+;+;+	3079	12163	421	91.30704109	FALSE	6.315708022	0.864256364	373	90.37915914	24486	4.388
C1orf86	C1orf86	chr1;chr1;chr1;chr1;chr	2115899;2117443;2118	2116952;211	-;-;-;-;-;-;-	6992	14167	1420	70.67933099	FALSE	11.20221821	1.086438787	1300	66.46891927	43431	4.637

Figure S3.2.: Extract from gene info (sce). Each row represents a gene, this excel provides brief information of the analyzed genes for this experiment, such as the chromosome origin (Chr) and location of start coordinate (Start).

Sample ID	Patient ID	Source	Age	Diagnosis	Cells for analysis
11511L&R	11511	cryo-preservation & overnight cultured	69	Endometrial cancer	63
11519L&R	11519	cryo-preservation & overnight cultured	50	HGSOC	610
11528L	11528	cryo-preservation & overnight cultured	56	HGSOC	258
11529L	11529	cryo-preservation & overnight cultured	78	HGSOC	146
15062L&R	15062	cryo-preservation & overnight cultured	73	Endometrial cancer	100
11543L&R	11543	fresh	73	Advanced ovarian cancer	225
11545L&R	11545	fresh	66	Primary peritoneal cancer	518
15066L&R	15066	fresh	52	High-grade endometrial cancer	606
11553L&R	11553	fresh	77	HGSOC	464
15072L&R	15072	fresh	62	squamous cell carcinoma of endometrium	319
11553-LT	11553	long-term cultured	\	\	26
15072-LT	15072	long-term cultured	\	\	135
11553-ON	11553	overnight cultured	\	\	229
15072-ON	15072	overnight cultured	\	\	178

Note:

L - left tube

R - right tube

LT- long-term

ON - overnight

Figure S 3.3: Patient details. (figure provided by supervisor)