Project Acronym:	**HUTER**
Project Full Name:	**Human Uterus Cell Atlas**
Call identifier:	H2020-SC1-2019-Single-Stage-RTD
Topic:	SC1-BHC-31-2019
Grant Agreement No:	874867
Start date of Project:	01/01/2020
Project Duration	2.5 years (30 months)
Document due date:	31/12/2021
Submission Date	20/12/2021
Leader of this report:	UEA
Deliverable no:	5.1
Deliverable name:	Methylation atlas for normal endometrium and myometrium tissue
Dissemination level:	**Public**

Version History

Version	Date	Details
1.0	12/11/2021	Template generated
1.1	13/12/2021	Circulating first draft
1.2	20/12/2021	Amended draft to submit

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 874867.

DELIVERABLE 5.1

Methylation atlas for normal endometrium and myometrium tissue

Table of Contents

List of Acronyms

ATAC	Assay for Transposase-Accessible Chromatin		**UEA**	University of East Anglia		**UMI**	Unique Molecule Identifier
RRBS	Reduced Representation Bisulphite Sequencing		**HCA**	Human Cell Atlas		**SMART**	Switching Mechanism At the end of the 5'-end of the RNA Transcript
PBAT	Post-Bisulphite Adaptor Tagging		**scM&T-seq**	Single cell methylation and transcriptome sequencing		**ssDNA**	Single-stranded DNA
PCR	Polymerase chain reaction		**WP**	Work package		**FBS**	Fetal bovine serum

1 PURPOSE OF THIS DOCUMENT

The purpose of this document is to present the first version of the "Methylome for normal endometrial and myometrium tissues, cataloguing cell types". As part of Deliverable 5.1, we will address two main areas:

a) Optimisation of the single-cell methylation methodologies.
b) Acquisition of healthy tissue biopsies.

Due to the ongoing COVID-19 pandemic, tissue recruitment for HUTER has been greatly impacted, which has had implications for D5.1. In the original version, we aimed to profile four myometrial and four endometrial and samples collected at INCLIVA (Spain) and WSI (UK), respectively. However, the organ donor program at CBTM (Cambridge - UK) suspended its activities for 2021, resulting in the cessation of recruitment by CBTM. Therefore, acquisition of samples was amended so that INCLIVA would be the site solely responsible for supplying tissues for this deliverable. As a result, amended delivery deadlines were agreed for WP 5 so that it was re-scheduled for December 31st 2021, an extension of 6 months from the original delivery date.

This document provides a description of the rigorous experimental optimisation of the "TRL3 *(experimental proof of concept lab)*" methods develop during this project, as well as an overview of the samples profiled so far. This document has been developed by UEA, which is the main lead partner of Deliverable D5.1 "Methylome for normal endometrial and myometrium tissues, cataloguing cell types" in collaboration with the additional Work Package partners involved.

1.1 Related documents

Documents linked to future actions to be delivered:

HUTER_WP5_D5.2_Endometrial methylation in sPE

2 EXECUTIVE SUMMARY

The main goal of the HUTER project is to create a reference atlas of the human uterus in health and disease by utilising state-of-the-art technologies allowing transcriptome profiling (scRNA-seq), spatial organisation (spatial transcriptomics), protein profiling and epigenomic analysis. This project is aligned with the Human Cell Atlas (HCA) initiative which aims to profile all cells in the human body and is part of the "Reproductive Network", an area of the HCA focused on profiling reproductive tissues. This document describes sample recruitment and experimental optimisation, as well as implementation, for Work Package 5 of the HUTER project, which aimed to profile DNA methylation in healthy human endometrium and myometrium at single-cell resolution.

We have continued to refine the single-cell methodologies used in the laboratory to help overcome some of the main challenges associated with these technologies, namely low and dispersed sequence alignments, whilst maintaining the ability to unite this epigenetic profiling with genome-wide transcriptomic analysis. We have focused on a comparison of three main methylation screening approaches in a uterine-derived cell line; single-cell Reduced Representation Bisulphite Sequencing, (scRRBS), bisulphite converted library preparations based on the ADAPTASE module, and finally assessing the scalability of Enzymatic Methylation conversion method, EM-seq.

Computationally, we have identified cell-specific marker genes that will be invaluable to assign cell identify, as well as loci most like to show changes in DNA methylation based on bulk methylome profiling. Altogether, our experimental and computational analyses will allow us to generate a comprehensive profile of the single-cell epigenetic signatures associated with human endometrium and myometrium once samples have been obtained.

3 INTRODUCTION

3.1 HUTER project

The HUTER project is a European pilot action to build on the foundations of the Human Cell Atlas (HCA), with the aims of generating a comprehensive reference catalogue of all human cells in order to better understand disease. HUTER is one of six pilot actions (along with BRAINTIME, DISCOVAIR, ESPACE, HCA Organoid, HUGODECA) within the Horizon2020 Research and Innovation Framework Programme, which initiated in January 2020. The HUTER projects aims to provide unprecedented

insights into multiple layers of transcriptomic and spatial patterning of gene expression in the uterus, an important dynamic female organ that exhibits changes in cellular processes and identities, through not only the menstrual cycle, but also across lifespan. The endometrium, the mucosal layer of the uterus, goes through cyclic phases of regeneration, differentiation and shedding during a women's reproductive life from puberty to menopause, whilst the myometrium is the underlying layer consisting mainly of smooth muscle cells, supporting stroma and vascular tissues. The two ovarian hormones, estrogen and progesterone, intricately regulate endometrial growth and differentiation. The endometrium consists of two main layers: the basal layer, which remains attached to the myometrium in all stages of the menstrual cycle, and the functional layer, which develops during the proliferative phase and is shed during menstruation. The basal layer is thought to be enriched in progenitor cells, but the epigenetic regulation controlling expansion and differentiation remains unknown. The main goals of Work Package 5 are to (i) utilise scM&T-seq to generate methylation atlases for endometrium and myometrium, (ii) use the transcriptomic data to deconvolute samples into specific cells types using marker gene expression and determine their unique methylation profiles [combined as Deliverable 5.1], as well as (iii) identify changes in cell-specific methylomes associated with severe pre-eclampsia [Deliverable 5.2].

3.2 Epigenetics and DNA methylation

Epigenetic modification cause heritable changes in phenotype without changes in the DNA sequence itself. One of the best examples of epigenetic regulation is the control of tissue-specification and differentiation during which lineage inappropriate genes are repressed. DNA methylation in mammals is an important epigenetic mark in which a methyl group ($-CH_3$) is added to cytosines to form 5-methylcytosine, normally within a CpG dinucleotide sequence context. Most CpGs are methylated throughout the genome, with the exception of CpG-dense regions called CpG islands that are generally associated with promoter regions (Deaton & Bird, 2011). These are frequently unmethylated and permissive to transcription. Being such an important modification (it has been referred to as the fifth DNA base), any aberrant profile is generally associated with a disease state, including cancer (Robertson, 2005).

3.3 Single-cell DNA methylation

The vast majority of DNA methylation profiles in tissue have been generated using "bulk" samples of heterogeneous cell populations (ENCODE Project Consortium), with only a few studies scrutinising this DNA modification at single-cell resolution. These single-cell epigenomic tools present an exciting opportunity to profile DNA methylation in unprecedented detail (Monk & Kelsey, 2018). While some recently developed methods have limited applicability, others have become gold-standard for assessing genome-wide methylation profiles. The majority of methods currently utilise sodium bisulphite treatment of DNA to discriminate between methylated and unmethylated cytosines. Upon exposure to bisulphite, unmethylated cytosines deaminate to uracils, while methylated cytosines remain unaltered. During the subsequent PCR stages, unmethylated cytosines are read as thymines, whereas methylated cytosines remain unchanged. This technique therefore offers base-pair resolution and is an invaluable tool when coupled to deep sequencing in revealing the entire methylome of a sample. However, a high level of genome coverage is expensive and mappability is challenging, as the bisulphite converted genomes possess less nucleotide complexity due to the inherent lack of cytosines.

Over the past five to six years several groups have attempted to adapt the bisulphite sequencing methods for single-cell applications. However, sodium bisulphite treatment is harsh and results in fragmentation of the DNA and sample loss, as well as the varying conversion rates, making bisulphite-based methods challenging for use with low input samples and single-cells (reviewed in Karemaker & Vermeulen 2018). Furthermore, high duplication rates are routinely observed due to the PCR-induced amplification required, however this has partially been addressed through the introduction of unique molecular identifiers (UMIs) (Wang et al., 2015). Some research teams, including ourselves, have attempted to improve coverage rates (due in part to sample loss) by utilising post-bisulphite adaptor tagging (PBAT) (Smallwood et al., 2014), where the harsh fragmentation associated with bisulphite treatment is used to our advantage in order to reduce the size of the DNA (as opposed to sonication) preceding adaptor ligation and PCR amplification. However, whilst we have obtained coverage rates of up to 18%, this method does not facilitate genome-wide coverage. More recently studies have utilised random priming and extension, after which the samples are tailed and ligated to a second adaptor in a single step, resulting in inline barcoding and increased multiplexing capacity (Luo et al., 2017). This single-nucleus methylcytosine

sequencing (snmC-seq) technology has recently been marketed by Swift BioSciences, being called Accel-NGS Adaptase Module. The methods described so far have limited scalability since the majority are not compatible with droplet-based microfluidic technologies and are restricted to plate-based library preparations. However, a highly-scalable assay for genome-wide methylation profiling based on single-cell combinatorial indexing for methylation analysis (sci-MET) has recently been described in which the authors produced 3282 single-cell bisulphite sequencing libraries (Mulqueen et al., 2017). With such increased multiplexing capacity, DNA loss is reduced, although such improvements do not compensate for loss of material due to the bisulphite conversion itself which has detrimental consequences for genome coverage and limits downstream bioinformatic analysis. Therefore, in an attempt to generate consistent fragments of the genome to assess methylation, methods such as Reduced Representation Bisulphite Sequencing (RRBS) have been adapted for low input and single-cell use (Gu et al., 2011). RRBS utilises restriction enzyme digestions (classically *MspI*, which cuts at CCGG motifs) and size fractionation (between 160-350 bp) to reduce the complexity of DNA, thereby generating smaller, yet reproducible libraries. This comes at a cost however, as low CpG density regions, that often include enhancers and cis-regulatory elements, are eliminated from the library preparation.

Recently, strategies that integrate single-cell multi-omic approaches have revealed the power of generating epigenetic and transcriptomic data simultaneously from the same cell. Apart from the recently released 10x Genomics Chromium Single Cell Multiome platform that combines ATAC-seq and gene expression, the majority of techniques are in their infancy. Methods such as scM&T-seq (Angemueller et al. 2016), that reveals single-cell DNA methylation and expression profiles, or scNOMe-seq (Clark et al. 2017), that also reports nucleosome occupancy, have been described in their seminal publications, but very few studies have utilised the techniques due to their specialist requirements and high costs.

4 Experimental pipeline

4.1 Overview of experimental plan

We have performed robust optimisation of genome-wide single-cell methylation protocols that can be incorporated in the single-cell genome and transcriptome (G&T-seq) (Macaulay et al., 2015)

pipeline to allow for simultaneous characterisation of DNA methylation and gene expression (termed scM&T-seq). This relies on physically separating mRNA and genomic DNA using poly(A)$^+$ mRNA hybridized to biotinylated oligo (dT) reverse transcription (RT) primer on streptavidin beads (please note, the transcriptome part of the scM&T-seq protocol is highly optimised and routinely performed, so we focused on developing the genome-wide methylation aspect). This procedure was classified as "TRL3 *(experimental proof of concept lab)*" experiment requiring additional optimisation in the original application as the methylation aspect of the protocol was known to be inefficient. As a consequence of our efforts, we have decided to generate the methylation data using a different method, swapping the scPBAT-seq protocol to scEM-seq (see description in subsequent section) as the C>T conversion rate was superior and the duplicated reads within the test libraries significantly less. These experiments were performed whilst waiting for samples to be recruited by WP3, which is ongoing. It is planned that healthy endometrial samples will be collected using Pipelle® suction curette which allows for a biopsy to be taken without the need of anaesthesia. The myometrial samples have already been collected from uteri obtained following hysterectomy. For several of these samples, paired endometrial samples were also obtained. However, due to the age of the patients, they cannot be classified as healthy controls, rather they may be interesting in their own right as they were all from post-menopausal women (which could be a future aim).

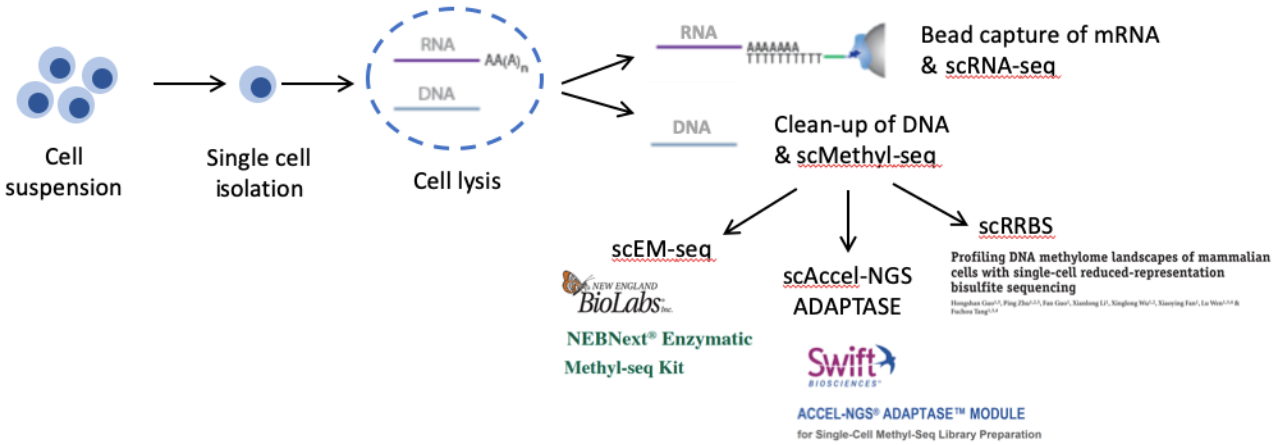

Figure 1. Overview of methodology.

4.2 Processing of samples (to date)

Samples were processed by the INCLIVA (Spain) team using established methods as described by the Vento-Tormo group (WSI) or the Simon team (INCIVA). Briefly, after transport in HypoThermosol, the tissue was minced up, washed in ice-cold RPMI supplemented with 10% FBS prior to a two-step enzymatic digestion to liberate single cells.

Myometrial dissociation (protocol is adapted from Mas et al. 2012 *Fertil Steril* with minor modifications): The endometrial and laser-burnt areas were roved from the myometrial layer using scalpel. The myometrial was then minced thoroughly into small pieces prior to digestion with Collagenase IV and DNase I over-night at 37 ºC and passed throught 100-50µm filters. If there was noticeable contamination with red blood cells, the dissociated cells were subjected to hypotonic shock using ACKL lysis buffer.
https://www.protocols.io/view/enzymatic-disaggregation-of-human-myometrium-for-1-bb5miq46.html

Endometrial dissociation (protocol is adapted from Vento-Tormo et al. 2018 *Nature* with minor modifications): A sequential digestion protocol was used to ensure cells from endometrium and epithelial glands were obtained. Following thorough washing and sample mincing, the tissue was digested with Collagenase V and DNase I for 45 minutes at 37 ºC and passed throught 100-50µm filters. The remaining tissues was then re-digested with Trypsin/EDTA for 20 minutes at 37 ºC and passed throught 100-50µm filters. The trypsin digestion was stopped by the addition of cell media containing 10% FBS. If there was noticeable contamination with red blood cells, the dissociated cells were subjected to hypotonic shock using ACKL lysis buffer.
https://www.protocols.io/view/endometrium-dissociation-with-collagenase-76thren
https://www.protocols.io/view/endometrium-dissociation-with-trypsin-72dhqa6

For both endometrial and myometrial dissociated samples, the individual cells were loaded into a in wells of 96 cell plates containing Qiagen RTL buffer, with the exception of wells A1, E7 (mini-bulk 100 cells) and C3, G9 (empty controls) using florescent-activated cell sorting (FACS). Plates were then immediately frozen at -80ºC until couriered on dry ice from INCLIVA (Spain) to UEA (UK).

Sample ID	Characteristics
Whole uterus samples	
SPA1-47 - myometrium and endometrial cells	Live donor, 74-year-old post-menopausal patient. Sample collected after hysterectomy. Cells from posterior zone of uterus.
SPA1-48 - myometrium and endometrial cells	Live donor, 67-year-old post-menopausal patient. Sample collected from hysterectomy. Cells from posterior zone of uterus.
SPA1-54 - myometrium and endometrial cells	Live donor, 62-year-old post-menopausal patient. Sample collected from hysterectomy. Cells from posterior zone of uterus.
SPA3-55 - myometrium and endometrial cells	Deceased donor, 62-year-old post-menopausal patient. Sample collected via the organ donor programme. Cells from posterior zone of uterus.
SPA1-56 - myometrium and endometrial cells	Live donor, 66-year-old post-menopausal patient. Sample collected from hysterectomy. Cells from posterior zone of uterus.
Endometrial biopsies	
No samples available as of 13/12/2021	

Table 1. Characteristics of samples available to study for Deliverable 5.1

4.3 Optimisation single-cell methylome methods (Pilot study)

Whilst waiting for processed cells from the primary samples, we performed extensive optimisation of both bisulphite and enzymatic DNA conversion for methylation analysis using the Ishikawa endometrial cell line (ECACC 99040201). This adenocarcinoma cell line was derived from a 39-year old woman and importantly retains its ability to respond to both estrogen and progesterone stimulation. These adherent cells were grown in basal media (DMEM supplemented with 10% FBS and 1% penicillin/streptomycin), then distributed into 96-well plates by FACS. Cells were sorted into wells containing single cells as well as mini-bulks of both ten and one hundred cells, prior to DNA and RNA separation and protocol optimisation.

To determine the best method to profile global DNA methylation we compared three additional protocols with the results from our standard scPBAT workflow, these were scRRBS, snmC-seq using the AcceI-NGS Adaptase Module and EM-seq from New England Biolabs. The first two procedures were dependent upon bisulphite conversion, which required further optimisation and down-scaling, the latter methods requiring down-scaling to ensure compatibility with single-cell analysis.

Comparison of bisulphite conversion conditions:

We compared reagents from two suppliers, Zymo Research (EZ DNA Methylation-Direct) and Qiagen EpiTect Fast Bisulfite kit) for three important parameters; DNA fragmentation; C>T conversion efficiency and the ability to capture post-bisulphite treated DNA using magnetic beads rather than columns (which we believed results in considerable sample loss contributing to the low and unpredictable genome coverage). To determine the extent of fragmentation, 500ng of DNA was exposed to the bisulphite reagents for varying amounts of time (10 minutes to 3.5hrs, depending on kit specifications). As can be seen in Figure 2, all kits fragmented high molecular weight DNA to a smear ranging from 400bp-5kb. It was evident that increased incubation times with the Qiagen EpiTect kit resulted in increased linear fragmentation, whilst the Zymo EZ Direct kit was largely stable over time. Therefore, we concluded that on bulk samples, incubation time made little difference of DNA integrity.

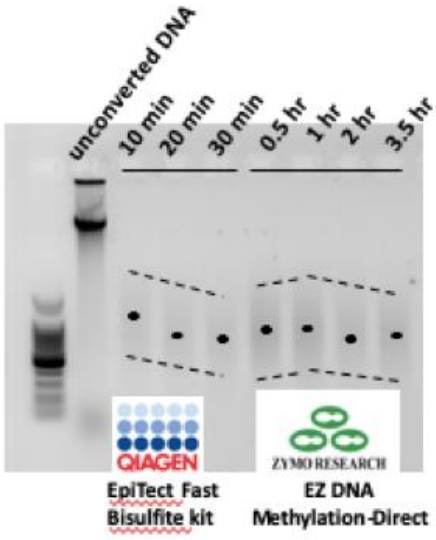

Figure 2. Gel electrophoresis of genomic DNA subjected to bisulphite conversion.

To determine conversion efficiency (which may vary in incubation time), we used a two-step approach. Firstly, following bisulphite conversion of DNA from 1000, 100 and 1 cells we performed

quantitative PCR (qPCR) using primers designed to the converted reference genome. If conversion was incomplete, any C>T mismatch in the primer sequence would impede primer binding and amplification rates. In all cases, multiple PCR amplifications were comparable (targeting methylated and unmethylated loci) suggesting that conversion had occurred efficiently (Figure 3 & 4). To determine the conversion rates of additional cytosines (not just under the primer binding sites), we also sequenced the amplicons. Aside from the shortest incubation using the Zymo EZ Direct kit, conversion was highly efficient.

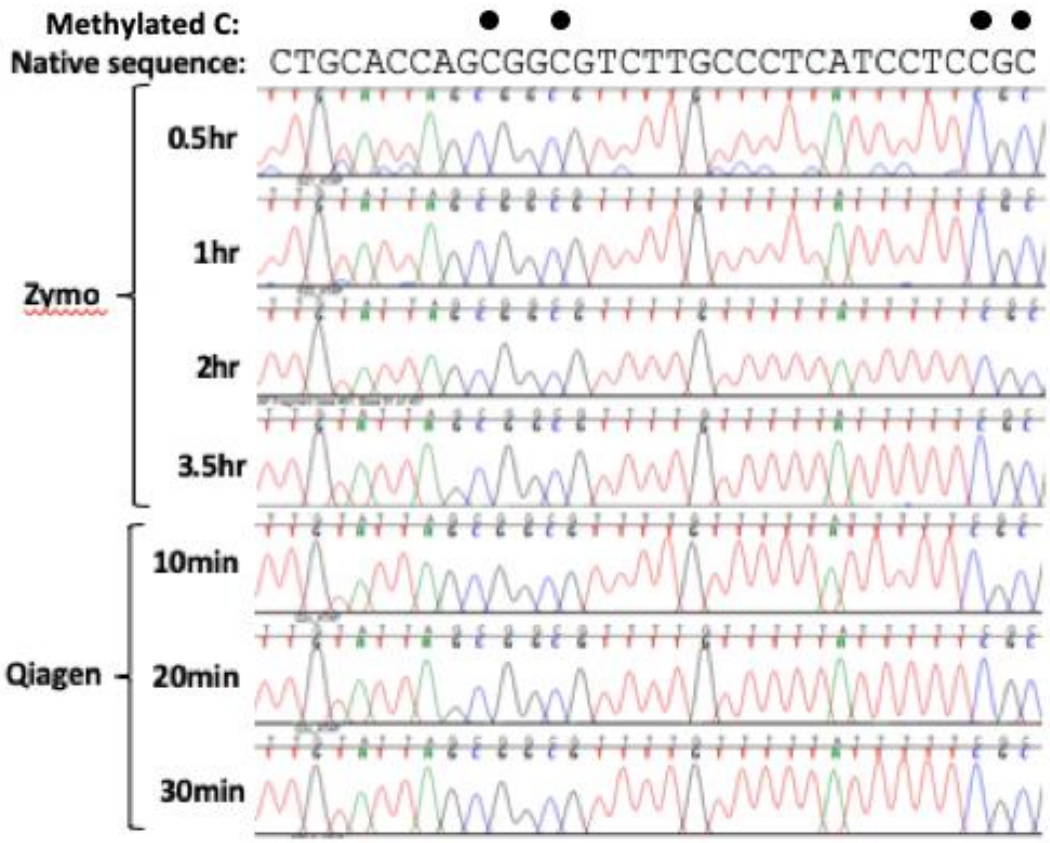

Figure 3. Sanger sequencing of amplicons revealing the partial conversion for 0.5 hr using the Zymo EZ Direct kit.

In order to limit a potential source of sample loss, we explored the use of Zymo MagBinding beads (magnetic DNA capture beads) to capture post-bisulphite treated DNA. We first aliquoted varying low-input amounts (10pg to 500ng) of bulk DNA to simulate near single-cell levels, then performed bisulphite conversion using the Zymo EZ Direct kit. Subsequent desulphonation and clean-up steps were performed using either the Zymo-Spin IC Columns (as supplied with the kit) or Zymo MagBinding

beads. We assessed recovery following clean-up by performing qPCR in a region of the *OOEP* promoter, using primers specifically targeting methylated sequences. Figure 4 shows that for each amount of input DNA tested, the bead purification method out-performed the columns for recovery of bisulphite-converted DNA. Importantly, for 10pg to 2ng input DNA, the levels of OOEP promoter is undetectable when columns are used, but become easily detachable by qPCR when beads are used for the clean-up steps.

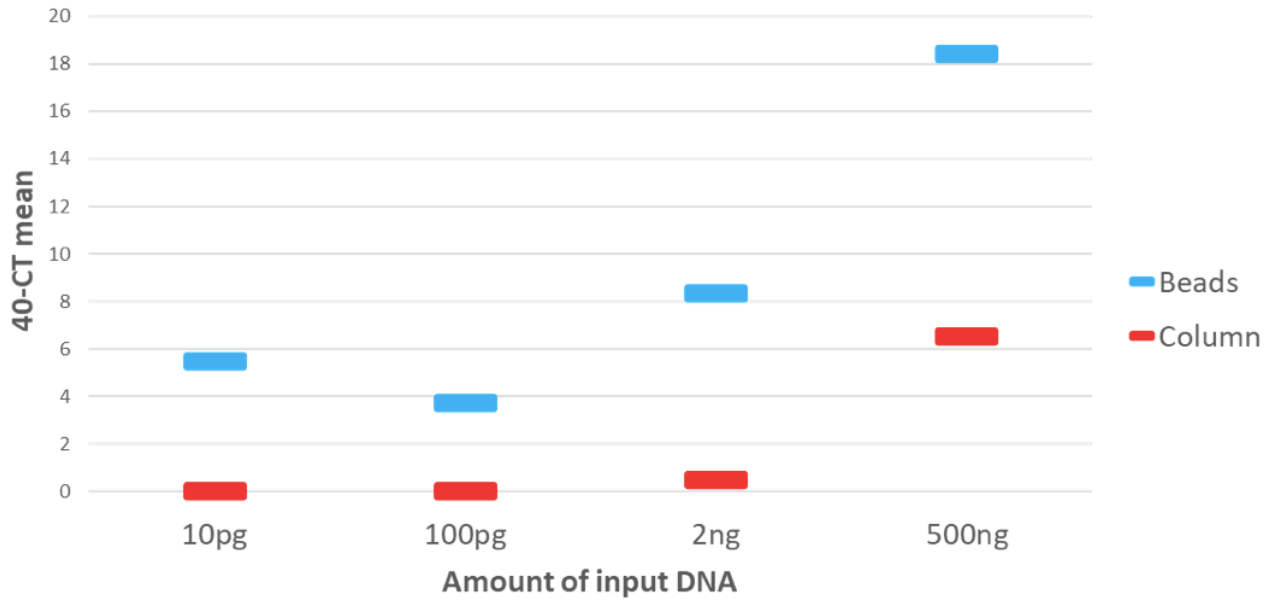

Figure 4. Quantitative PCR levels of *OOEP* promoter regions performed on varying amounts of bisulphite-converted DNA purified by either standard Zymo-Spin IC Columns or Zymo MagBinding beads.

Comparison of scRRBS, AcceI-NGS Adaptase Module and EM-seq library preparations methods:

- **scRRBS**

We followed the scRRBS protocol from Guo and colleagues (Guo et al., 2015) using mini-bulk and single cell inputs. In our hands the methods worked well for low-input amounts (as low as 100pg or ~15 cells) but was extremely variable when using single cells. This was especially apparent when characterising the "recovery" of DNA fragments using qPCR (using carefully designed PCRs to amplify between and across M*spI* sites). This method does not pool samples until the latter stages of the protocol, which may contribute to material loss during the clean-up stages. Whist we were adapting this method, the Gnirke & Meissner laboratories published an improved protocol compatible with dual transcriptome analysis (smart-RRBS, Gu et al., 2021) as so we made several changes to incorporate their improved procedure.

Briefly, after physically separating poly(A)$^+$ mRNAs and genomic DNA, the captured mRNAs from each single cell are reverse transcribed in the presence of a template-switching oligonucleotide (TSO) to enable subsequent PCR amplification using a single primer. These PCR products are purified, quantified and converted into pools of Illumina single-cell RNA-seq libraries using the Nextera XT DNA tagmentation kit.

Separately, the mRNA-depleted DNA fractions were cleaned up on AMPure XP SPRI beads and digested by MspI or by MspI and HaeIII (double-digest). The resulting fragments were blunt-ended and dA-tailed by Klenow exoDNA polymerase and ligated to indexed methylated adapters. Next, 12 (but this can be increased to 24) indexed ligation reactions are pooled together and subject to bisulphite conversion, followed by amplification with PCR primers carrying pool-specific indices (using different cycle numbers ranging from 10 to 25 cycles). The resulting multiple library pools were subject to TapeStation analysis and used for PCR-based QC steps, but ultimately would be sequenced on the same lane of an Illumina sequencer.

As a result of the inconsistent library size and quantities obtained, we decided that we would focus on the other methodologies, especially as the incorporation of the scRRBS protocol into the scM&T workflow had just been published.

- **AcceI-NGS Adaptase module**

This method was described as snmcC-Seq (Luo et al., 2017) and the first step is bisulphite conversion of DNA, for which we used 100, 10 and individual Ishikawa cells using the Zymo direct kit with optimised reduced volumes (65µl CT conversion reagents added to DNA in 10 µl, with incubations at 64^{o}C for 3 hours) and subsequent magnetic clean-up. Once converted, bisulphite treated DNA is single-stranded and the Adaptase module adds a truncated P7 adapter sequence to the 3' end ssDNA products resulting from random primed synthesis. This is followed by an enzymatic step using ExonucleaseI and Shrimp Alkaline phosphatase to remove unused random primers and inactivate dNTPs, and a SPRI bead clean-up. Since the random primer incorporates a truncated P5 adapter to 5' ends, the subsequent Adaptase step ligates a truncated P7 adapter to the 3'end of the ssDNA products which is followed directly by PCR amplification with indexed primers to enable multiplexing to complete library construction. As described above, the multiple library pools were subject to TapeStation analysis and used for PCR-based QC steps. This revealed that the libraries were complex and sample quantities were high, helped partially by the early multiplexing. When assessing the size

artifacts distribution of the library using TapeStation analysis, this occasionally revealed high molecular weight smears well above the expected library size (greater than 2kb) products well above the expected library size. These turned out to be bubble products, which are well-known artifacts of over-amplication of next-generation sequence libraries when PCR reagents are exhausted. Bubble products are hetroduplexes composed of partially homologous library fragments containing double-stranded complementary adaptor sequences flanking single-stranded noncomplementary inserts. This was remedied by the inclusion of three to five additional PCR cycles using a fresh master mix. As expected, the library size was dictated by the bisulphite DNA fragmentation, hence libraries were generally in the 300bp-1kb range (Figure 5). All optimisation libraries were sequenced on MiSeq to determine bisulphite conversion rates (as judged by non-CpG C's in sequence) and library duplication rates (see section 4.4 for details).

- **EM-seq scRRBS**

This technique relies upon an enzymatic conversion of unmethylated cytosines as opposed to bisulphite conversion to map CpG methylation and has the advantage that library size is manipulable using highly-specific ultrasonication. This is a commercial kit from New England Biolabs, although it is only recommended for starting amounts of 100pg-50ng DNA. To determine if this technology could be used for very low-input and single-cell applications, we performed the protocol on 100pg and 10pg bulk DNA, as well as 100, 10 sorted cells, plus individual Ishikawa cells. The initial step of the protocol is the controlled fragmentation of DNA using a Covaris LE220 Focused-ultrasonicator instrument, using optimised settings (450W peak incident power, 15% duty factor, 200 cycles per burst, for 160 sec at 5-15^{o}C, in 55µl volume) to obtain fragments with a peak of 300bp. Subsequently, DNA ends were trimmed and adaptors were ligated to each end of the fragments followed by a SPRI bead clean-up. The DNA was then subject to TET2 convertion followed by APOBEC-induced cytosine deamination and a further SPRI neat clean-up. The final steps of the preparation are library amplification with barcoded oligonucleotides, followed by pooling of individual samples, SPRI bead clean-up, elution and standard QCs. When assessing the distribution of the libraries using TapeStation, we observed fragments consistent 300bp-1kb size range, with a peak size of 400-500bp (Figure 5), perfect for 150bp paired-end sequencing which would ensure maximal coverage of the genome. Furthermore, qPCR of the libraries reassuringly revealed excellent amplifications, indicative of complex libraries. All optimisation libraries were subsequently sequenced on MiSeq to determine bisulphite conversion rates and library duplication rates (see section 4.4 for details).

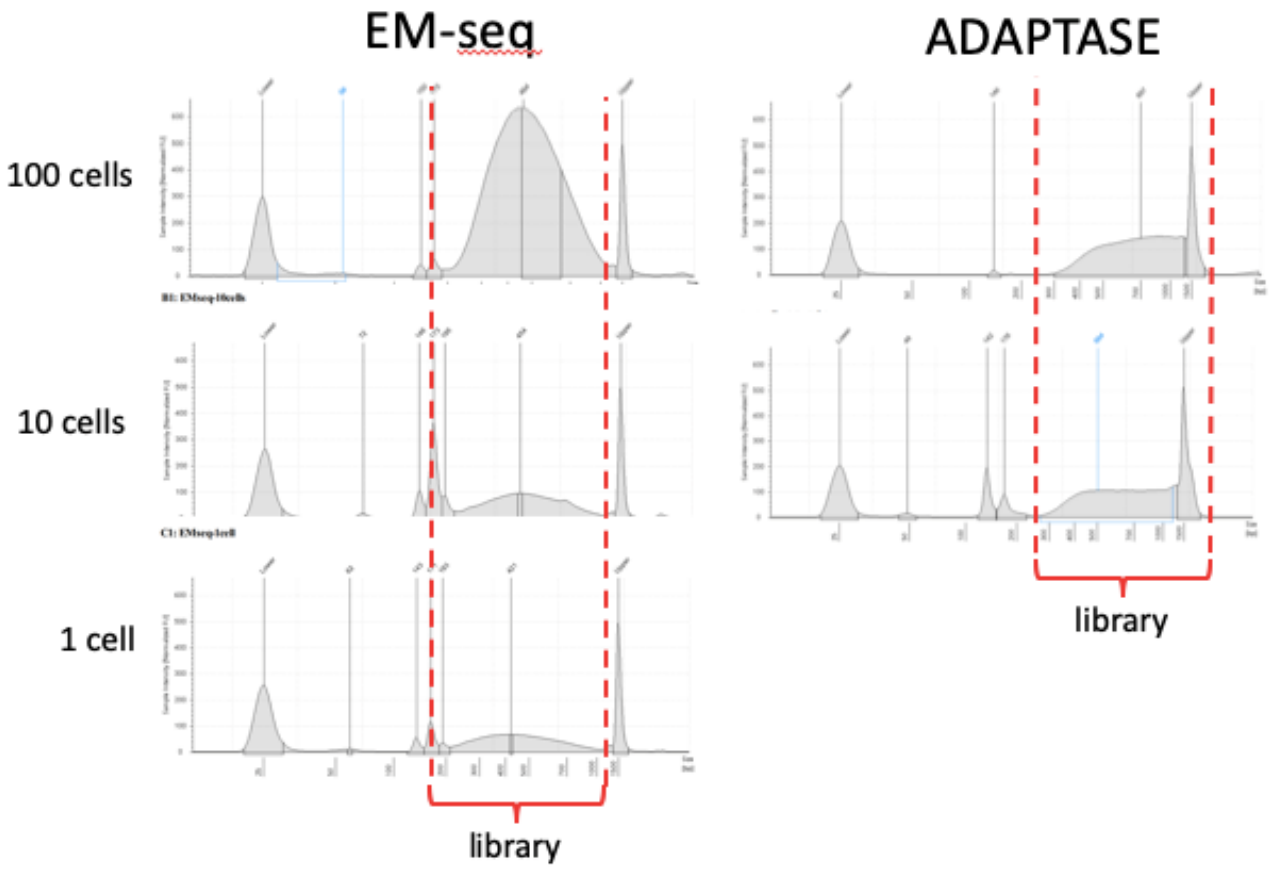

Figure 5. Comparison of TapeStation results for scEM-seq and snmcC-Seq ADAPTASE module. Note that gDNA was fragmented by Covaris sonication for EM-seq, whereas sample fragmentation in snmcC-Seq is dependent upon bisulphite degradation.

4.4 Bioinformatic QC for single-cell methylation techniques (Pilot study)

The resulting methylation libraries were bioinformatically assessed using standard pipelines. The raw sequencing reads were trimmed using Trim Galore (v0.6.5) to remove adapters and poor-quality base calls. Trimmed reads were then aligned to the reference human genome using Bismark in single-end non-directional mode. After removal of duplicate alignments, methylation calls were extracted and coverage was calculated. All the steps after trimming were performed using Bismark (v0.22.3). R was used for processing of the output files and the generation of plots using in-house scripts. As can be clearly observed in Figure 6, the scEM-seq method out performed the snmcC-Seq ADAPTASE module and scPBAT (previous study), and has been selected as the method to incorporated into the scM&T-seq protocol.

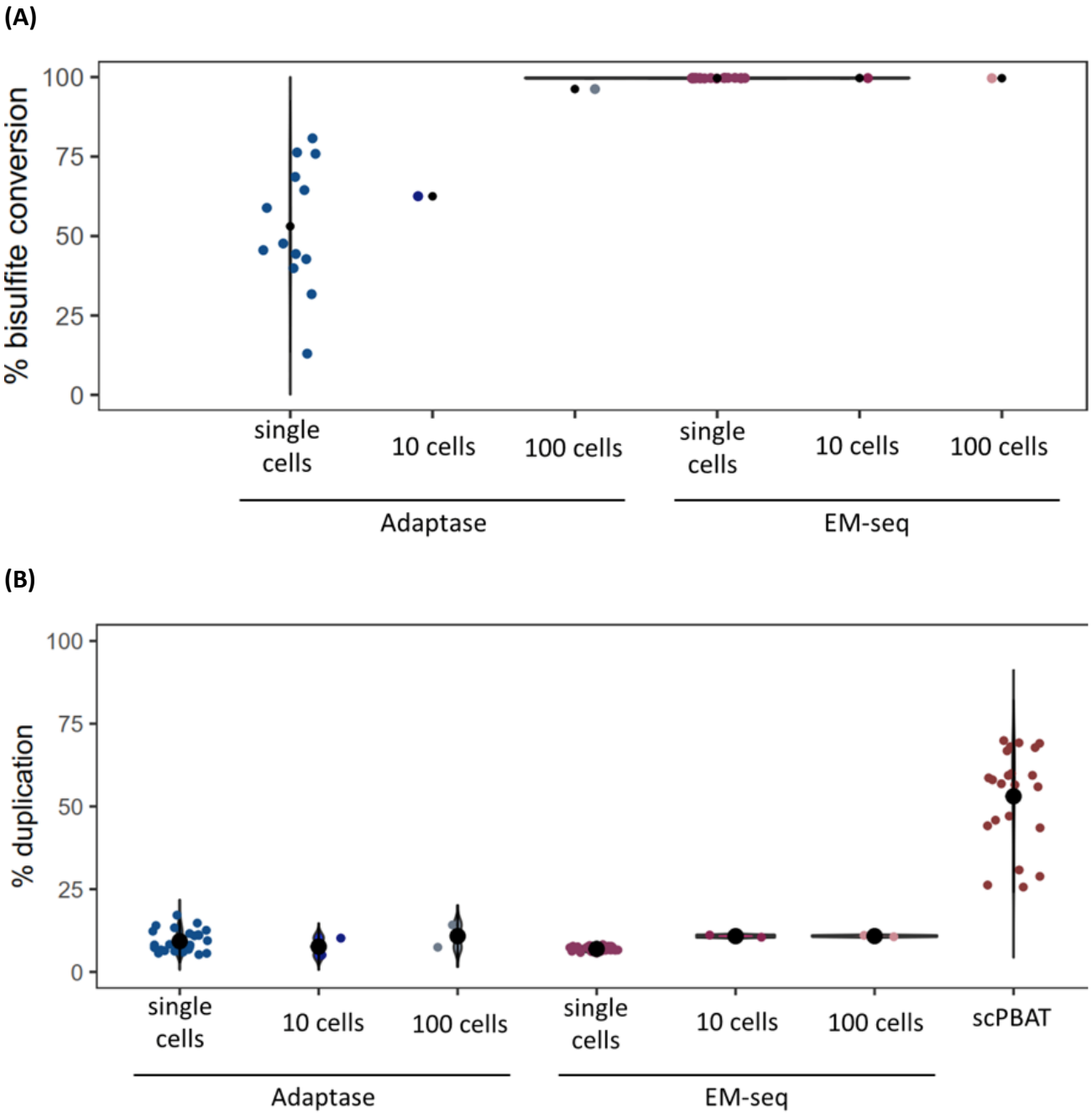

Figure 6. Sequence QC comparison for (A) conversion rates for single-cell and mini-bulks of 10 and 100 cells for ADAPTASE and EM-seq libraries and (B) duplication rates for single-cell and mini-bulks of 10 and 100 cells for ADAPTASE and EM-seq libraries compared with those generated using the original scPBAT method.

In addition to analysing the data, the UEA bioinformatician has been interrogating data generated by the HUTER partners to identify cell-type specific marker genes that we can use to subsequently identify cells following cluster analysis. Table 2 highlights some of the selected markers for both endometrial and myometrial cell identity.

Stromal cells	**Endothelium**	**Unciliated epithelium**	**Ciliated epithelium**	**Myometrial muscle cells**
DCN	ADGRL4	EPCAM	CDHR3	ACT2
COL6A1	VWF	WFDC2	SNTN	MYH11
CRISPLD2	PCDH17	KLF5	DYNLRB2	MYL6
COL6A3	PECAM1	SDC4	FAM183A	MYL9
LUM		UCA1	PIFO	MCAM
COL5A1		ELF3	ARMC3	PDGFRB
MMP11		DSP	CAP5L	
SFRP1		CLDN4	MS4AB	
WNT5A		KRT18	CFAPS3	
DKK1		CLDN3	MMP7	
FOXO1		GPX3	SCGB1D2	
CRYAB		CXCL14	GPX3	

Table 2. List of markers used to determine cell identity. Analysis was performed on data from Wang et al., 2020; Wu et al., 2021 and Vento-Tormo et al., 2018.

Furthermore, in anticipation of performing both genome-wide (binned) and locus-specific methylation analyses, we have examined datasets from recent studies with the expectation of identifying loci that respond to hormonal stimulation to determine the role of DNA methylation in regulating estrogen and progesterone signalling (based on differential gene expression in organoids following stimulations from Garcia-Alonso et al. 2021, with confirmation in Ishiwaka cells) (i.e. *PGR*, *CDC20B, CCNO, HES6, FOXJ1* and *PIFO*), as well as genes that have dynamic expression and methylation profiles in endometrial biopsies during the transition from pre-receptive to receptive phase of the cycle (Kukushkina et al., 2017) e.g. *ARL15* and *TRPM1.*

5 Data and public dissemination

Although no publications have been generated to date, the UEA team have published a review article discussing the recent advances in single-cell multi-omic technologies related to fertility.

- Dagnė Daškevičiūtė, Marta Sanchez-Delgado, David Monk. Epigenetics from oocyte to Embryo. New Genetic Diagnostic Technologies in Reproductive Medicine, Second Edition. Taylor & Francis Group.

Furthermore, members of the UEA team participated in the Earlham Institute Single-Cell symposium in 2021 and all relevant HCA Network and Annual meetings.

With regards interacting with non-academic stakeholders and the general public, the UEA team participated in the Norwich Science Festival in October 2021, presenting both informative posters and an activity stand, during which we explained the application of single-cell technologies in placenta-uterine research. The event was attended by more than 10,000 members of the public.

6 Conclusions

D5.1 has two main goals:

(i) To optimise and utilise scM&T-seq to generate methylation atlases for endometrium and myometrium.

(ii) To use the associated transcriptomic data to deconvolute samples into specific cells types using marker gene expression and determine their unique methylation profiles.

The COVID-19 pandemic has hugely affected the number of donors recruited on this project. Due to the shortage of samples, during the first 15 months we have focused on comparing different genome-wide methylation methods, ensuring they are compatible with physical separation of genomic DNA from polyA mRNA used for scRNA-seq using SMART-seq2 protocols. As described in this report, we have successfully compared numerous protocols, which has revealed that EM-seq has not only the most streamlined protocol but the best library coverage and lowest duplication rate using an endometrial cell line. Whilst we have not fully processed the myometrial samples collected to date, for fear of introducing unwanted batch effects during the NGS stages, we will proceed once

the normal endometrial samples have been collected. Based on our experience, we are confident we will be able to deliver the full dataset proposed by the end of this grant (June 2022) once samples are recruited and processed by INCLIVA (Spain) and WP3.

7 Bibliography

- Angermueller et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat. Methods 13: 229–232.
- Clark et al. (2017). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. 9: 781
- Eaton & Bird (2011). CpG islands and the regulation of transcription. Genes Dev. 25:1010-22.
- ENCODE Project Consortium (2012). An integrated encyclopaedia of DNA elements in the human genome. Nature. 489: 57-74.
- Garcia-Alonso et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro.* BioRxiv.
- Gu et al. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat. Protocols 6: 468–481.
- Gu et al. (2021). Smart-RRBS for single-cell methylome and transcriptome analysis. Nature protocol 16: 4004-4030.
- Guo et al. (2015). Profiling DNA methylome landscape of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nature Protocol 10:645-659.
- Karemaker & Vermeulen (2018). Single-Cell DNA Methylation Profiling: Technologies and Biological Applications. Trends in Biotechnology. 36: 952-965.
- Kukushkina et al. (2017). DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Scientific Reports &:3916.
- Luo et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science 604: 600–604.
- Macaulay et al. (2015). G&T-seq: parallel sequencing of single-cell genome and transcriptomes Nature Methods 12: 519-522.
- Mas et al. (2012). Identification and characterisation of the human leiomyoma side populations as putative tumour-initiating cells. Fertility & Sterility 98: 741-751.

- Monk & Kelsey. Laboratory Molecular Methodologies to Analyse DNA Methylation. Epigenetics and Assisted Reproduction. 1st Edition, 2018. CRC Press.
- Mulqueen et al. (2017). Scalable and efficient single-cell DNA methylation sequencing by combinatorial indexing. Nature Biotechnology 36: 428-431.
- Robertson (2005). DNA methylation and human disease. *Nature Reviews Genetics* 6: 597–610.
- Smallwood et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods 11: 817–820.
- Vento-Tormo et al. (2018) Single-cell reconstruction of the early maternal-fetal interface. Nature 563: 347-353.
- Wang et al. (2015). Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analyses. Epigenetics 10: 775–783.
- Wang et al. (2020). Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. Nat Med 26: 1644-1653.
- Wu et al. (2021) Cell Atlas of human uterus. bioRxiv.