Biomarker Development in Endometrial Cancer: Circulating Tumour Cells, Tissue Methylation and Genotyping Studies

Charlotte Rose Lemech

A thesis submitted to the University College London (UCL) for the degree of MD (res) in the field of Clinical Translational Oncology

University College London Cancer Institute
72 Huntley St
London
WC1E 6DD

October 2014

Declaration

I, Charlotte Rose Lemech, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Biomarker Development in Endometrial Cancer: Circulating Tumour Cells, Tissue Methylation and Genotyping Studies

Abstract

Endometrial cancer (EC) is the most common gynaecological malignancy in the developed world and 4th most common women's cancer in the UK. Although approximately 75% of women present with surgically resectable disease, 20% of these will relapse and 25% of initial diagnoses occur with metastatic disease. There are currently no validated biomarkers to assess treatment response or target molecular therapies, despite evidence for targetable aberrations in the literature. The PI3K pathway in particular is commonly mutated in EC and stathmin is a recently identified phosphoprotein associated with PI3K pathway activation and with prognostic significance in EC.

I investigated biomarker development and novel pathways in EC in two ways: firstly, through a feasibility study of circulating tumour cell (CTC) enumeration and molecular profiling (MP) in patients with advanced endometrioid and non-endometrioid EC (EEC and NEEC); and secondly, through methylation and copy number variation (CNV) studies on formalin-fixed paraffin-embedded (FFPE) and fresh frozen (FF) EEC.

CTCs have prognostic and predictive significance in a number of cancers, with increasing evidence on CTC MP. CTC enumeration and assessment of stathmin overexpression was performed on the validated Veridex CellSearch platform and shown to be feasible. In addition, CTC positivity was associated with worse survival and there was a subset of patients for whom a positive CTC count was predictive of outcome on chemotherapy.

The second component focused on methylation and CNV analysis in the different phases of endometrial carcinogenesis from normal endometrium to atypical endometrial hyperplasia (AEH) and EEC. By extracting DNA from FFPE and FF EEC and using the Illumina Infinium HumanMethylation450 BeadChip, I was able to demonstrate these methods were feasible and that differential methylation and CNV was evident in the transition from normal endometrium through to AEH and EEC.

This research provides a basis for further biomarker development and novel target selection in EC.

Table of Contents **Page**

Appendix **Page**

Acknowledgements

I would like to acknowledge my supervisors, Rebecca Kristeleit, Jonathan Ledermann and Tobi Arkenau for their guidance.

In particular, I would like to thank those in the UCL Cancer Institute and UCLH gynae-oncology team, who assisted with patient recruitment and helped train me in the techniques and data analysis required for this research. This includes Anna Karpathakis, Rupali Arora, Harpreet Dibra, Leah Ensell, Helen Lowe, Allison Jones, Dan Reidl, Shohreh Ghazali, as well as Kerra Pearce.

Anna assisted with the techniques involved with handling the FFPE specimens and interpreting the data from the differential methylation analysis. Harpreet also helped with learning the techniques involved with DNA extraction and modification.
Dan and Shohreh were persistent in helping with the fresh tissue collection for the methylation analysis. Kerra ran all the samples for the methylation analysis.
Rupali assisted with identifying appropriate tissue for the methylation studies, as well as in interpreting the EpCAM and stathmin immunohistochemistry.
Leah and Helen processed the CTC samples and assisted me with the interpretation of the CTC data.

My husband, Kelum Kumarasinghe, and newborn baby, Jasmine, kept me positive and motivated through the fluctuations and challenges that characterise part-time research and have been the greatest support over this time.

Thanks to all

List of Tables

List of Figures

Abbreviations

A: adenine
ABCB1: ATP-binding cassette, sub-family B (MDR/TAP), member 1
ACTA1: actin, alpha 1, skeletal muscle
ACTA2: actin, alpha 2, smooth muscle, aorta
AEH: atypical endometrial hyperplasia
AKT: v-akt murine thymoma viral oncogene homolog
ALG10: alpha-1,2-glucosyltransferase
ALK: anaplastic lymphoma receptor tyrosine kinase
APC: adenomatous polyposis coli
ARID 1A/5B: AT-rich interactive domain 1A/5B
ATF2: activating transcription factor 2
ATXN10: ataxin 10
AXIN1/2: axis inhibition protein 1/2
BCL9: B-cell CLL/lymphoma 9
BDNF: brain-derived neurotrophic factor
BIN1: bridging integrator 2
BIS: bisulfite
BMP2/4/5/6/8A/8B: bone morphogenetic protein 2/4/5/6/8A/8B
BRAF: v-raf murine sarcoma viral oncogene homolog B
BRCA: breast cancer associated
C: cytosine
CA125: cancer antigen 125
CACNB2: calcium channel, voltage-dependent, beta 2/4 subunit
CALM2: calmodulin 2
CALML3/5: calmodulin-like 3/5
CARD11: caspase recruitment domain family, member 11
CASP8: caspase 8/9, apoptosis-related cysteine peptidase
CCND3/E1: cyclin D3/E1
CD8B: CD (cluster of differentiation) 8B molecular
CDC42/25A: cell division cycle 42/25A
CDH1: cadherin 1, type 1, E-cadherin (epithelial)
CDH13: cadherin 13
CDKN1A: cyclin-dependent kinase inhibitor 1A
CDKN2A: cyclin-dependent kinase inhibitor 2A (also known as P16INK4A or P14ARF)
CFLAR: CASP8 and FADD-like apoptosis regulator
CGI: 5'cytosine-phosphate-guanine-3' dinucleotide island
ChAMP: Chip Analysis Methylation Pipeline
CHFR: checkpoint with forkhead and ring finger domains
CHUK: conserved helix-loop-helix ubiquitous kinase
CIMP: CpG island methylator phenotype
CLDN4: claudin 4
c-met: c-met proto-oncogene
CN: chromosomal copy number
CNV: chromosomal copy number variation
COL1A1: collagen, type I, alpha 1
COL4A2/4: collagen, type IV, alpha 2/4
COMT: catechol-O-methyltransferase
COX2: cyclooxygenase 2
CpG: 5'cytosine-phosphate- guanine-3' dinucleotide
Cq: quantification cycle
CRH: corticotrophin releasing hormone

CRK: v-crk avian sarcoma virus CT10 oncogene homolog
CSF3R: colony stimulating factor 3 receptor (granulocyte)
CSNK1E: casein kinase 1, epsilon
CT: computed tomography
CTBP1/2: c-terminal binding protein 1/2
CTC: circulating tumour cell
CTGF: connective tissue growth factor
CTNNA2/3: catenin (cadherin-associated protein) alpha a2 1, 102kD/a3
CTNNB1: catenin (cadherin-associated) protein Beta 1
CXCL12: chemokine (C-X-C motif) ligand 12
CYTH2: cytohesin 2
DAAM2: dishevelled associated activator of morphogenesis 2
DAB: 3,3'-Diaminobenzidine tetrahydrochloride
DAXX: death-domain associated protein
DCC: deleted in colorectal carcinoma
DEFB130: beta-defensin 130
DM: diabetes mellitus
DMRs: differentially methylated regions
DNA: deoxyribonucleic acid
DNM2: dynamin 2
dsDNA: double stranded deoxyribonucleic acid
DUSP6: dual specificity phosphatase 6
EBRT: external beam radiotherapy
EC: endometrial cancer
ECOG: Eastern Cooperative Oncology Group
EEC: endometrioid endometrial cancer
EGFR: epidermal growth factor receptor
EOMES: eomesodermin
EpCAM: epithelial cell adhesion molecule
ER: oestrogen receptor
ERG: erythroblast transformation-specific related gene
ERBB2: see HER2
ERBB4: v-erb-b2 avian erythroblastic leukaemia viral oncogene homolog 4
ERK: extracellular signal-regulated kinase
ERN1: endoplasmic reticulum to nucleus signalling 1
ETS1: v-ets erythroblastosis virus E26 oncogene homolog 1
ETX1: sushi-repeat containing protein, X-linked (SRPX)
EXOC2: exocyst complex component 2
FAK: focal adhesion kinase
FAS: Fas cell surface death receptor
FF: fresh frozen
FFPE: formalin-fixed, paraffin-embedded
FGF/R: fibroblast growth factor/receptor
FIGO: International Federation of Gynaecology and Obstetrics
FOXA1/G1/O3: forkhead box A1/G1/O3
G: guanine
GADD45A: growth arrest and DNA-damage-inducible, alpha
GALNT 6/12: polypeptide N-acetylgalactosaminyltransferase 6/12
GATA5: GATA binding protein 5
GCLP: good clinical laboratory practice
GCP: good clinical practice
GFRA1: GDNF family receptor alpha 1
GLI2: GLI family zinc finger 2

GNA12: guanine nucleotide binding protein (G protein) alpha 12
GOG: Gynaecologic Oncology Group
GPR54: G-protein coupled receptor 54; KISS1 receptor
GRB2: growth factor receptor-bound protein 2
GSC: goosecoid homeobox
H2AFY: H2A histone family, member Y
HBA1/2: haemoglobin, alpha 1/2
HBQ1: haemoglobin, theta 1
HDAC1/2/5: histone deacetylase 1/2/5
HER2: human epidermal growth factor receptor 2 (ERBB2)
HGD: homogentisate 1,2-dioxygenase
HGFR: hepatocyte growth factor receptor
HIC: hypermethylated in cancer
HIPK2: homeodomain interacting protein kinase 2
HLA-A/C/G: major histocompatibility complex, class I, A/C/G
HLA-DMB: major histocompatibility complex, class II, DM beta
HNPCC: hereditary non-polyposis colorectal cancer
HOXAII: homeobox A2
HSP90AB1: heat shock protein 90kDa alpha (cytosolic), class B member 1
HSPA1A: heat shock 70kDa protein 1A
ICH: institute for Child Health
IGFALS: insulin-like growth factor binding protein, acid labile subunit
IGR: intergenic region
IHC: immunohistochemistry
IHH: indian hedgehog
IRF4/5: interferon regulatory factor 4/5
ITG A2B/3/A6/AV/B5: integrin, alpha 2b/alpha 3/alpha 6/alpha V/beta 5
ITGB1: integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
JAK1: janus kinase 1
Jak-STAT: janus kinase-signal transducer and activator of transcription
JUN: jun proto-oncogene
KEGG: Kyoto encyclopaedia of genes and genomes
KIAA0240: GLTSCR1-like (glioma tumour suppressor candidate region gene 1 protein-like)
KIFC3: kinesin family member C3
KRAS: kirsten rat sarcoma viral oncogene homolog
KRT8: keratin 8
LAMA3/4: laminin, alpha 3/4
LN: lymph node
LOH: loss of heterozygosity
LUC7L: LUC7-like (S.Cerevisiae)
LVSI: lymphovascular space invasion
MADCAM1: mucosal vascular addresin cell adhesion molecule 1
MAGUK: membrane-associated guanylate kinase (member of the CARD protein family)
MAP2K 3/4/6: mitogen-activated protein kinase kinase 3/4/6
MAP3K 1/2/3/4/6/7/14: mitogen-activated protein kinase kinase kinase 1/2/3/4/6/7/14
MAPK: mitogen-activated protein kinase
MAPK8IP3: mitogen activated protein kinase 8 interacting protein 3
MDFIC: MyoD family inhibitor domain containing
MDS: multidimensional scaling
MeDIP: methylated DNA immunoprecipitation
MEF2D: myocyte enhancer factor 2D
MKNK1: MAP kinase interacting serine/threonine kinase 1

MLH1: mutL homolog 1
MMI: myometrial invasion
MMP: matrix metallopeptidase
MMR: mismatch repair
MRI: magnetic resonance imaging
MSH2/3/6: mutS homolog 2/3/6
MSI: microsatellite instability
MSP: methylation specific PCR
MSRE: methylation-sensitive restriction enzyme
mTOR: mammalian target of rapamycin
MVPs: methylation variable positions
MYC: v-myc avian myelocytomatosis viral oncogene homolog
MYL9: myosin, light chain 9, regulatory
MYOD: myogenic differentiation
NCAM2: neural cell adhesion molecule 2
NDRG2: N-myc downstream-regulated gene family member 2
NEDD4L: neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase
NEEC: non-endometrioid endometrial cancer
NFKB2: nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
NGS: next generation sequencing
NKD1: naked cuticle homolog 1
NOD: nucleotide-binding oligomerization domain
NOD2: nucleotide-binding oligomerization domain containing 2
NPM1: nucleophosmin (nucleolar phosphoprotein B23, numatrin)
NRG1/3: neuregulin1/3
NT5C3: 5'-nucleotide, cytosolic III1
NTRK1: neurotrophic tyrosine kinase, receptor type 1
OPRM1: opioid receptor, mu 1
OR4F4: olfactory receptor, family 4, subfamily F, member 4
OS: overall survival
P14ARF: see CDKN2A
P16INK4A: see CDKN2A
P53: see TP53
P70S6kinase: p70 ribosomal protein S6 kinase
PARVG: parvin gamma
PAX2/8: paired box 2/8
PBC: peak based correction
PBX1: pre-B-cell leukaemia homeobox 1
PC: pyruvate carboxylase
PCSK2: proprotein convertase subtilising/kexin type 2
PDE6B: phosphodiesterase 6B, cGMP- specific rod beta
PDGFR: platelet derived growth factor receptor
PENK: proenkephalin
PET: positron emission tomography
PFS: progression-free survival
PI3K: phosphoinositide 3-kinase
PI4K2A: phosphatidylinositol 4-kinase type 2 alpha
PIK3CA/D: phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha/delta
PIK3R1/R3/R5: phosphoinositide 3-kinase, regulatory subunit 1(alpha)/3(gamma)/5
PIM1: pim-1 oncogene
PIPP4K2A: phosphatidylinositol-5-phosphate 4-kinase, type 2 alpha
PITX2: paired-like homeodomain 2

PLA2G 1B/2C/2E/5: phospholipase A2, group IB/IIC/IIE/V
PMS2: postmeiotic segregation increased 2
PORTEC: post-operative radiation therapy in endometrial carcinoma
POU4F2: POU class 4 homeobox 2
PPP3CB/C: protein phosphatase 3, catalytic subunit, beta/gamma isozyme
PPP3R1: protein phosphatase 3, regulatory subunit B, alpha
PPARD: peroxisome proliferator-activated receptor delta
PPP1R13L: protein phosphatase 1, regulatory subunit 13 like
PR: progesterone receptor
PRKCA/E/Z: protein kinase C, alpha/epsilon/zeta
PROKR1/2: prokineticin receptor 1/2
PRR5: proline rich 5 (renal)
PTCH2: patched 2
PTEN: phosphatase and tensin homologue
PTPN6: protein tyrosine phosphatase, non-receptor type 6
PTPRE: protein tyrosine phosphatase receptor type E
AQP5: aquaporin 5
PTPRN: protein tyrosine phosphatase receptor type N
PXN: paxilin
RAC3: ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)
RAPGEF1: Rap guanine nucleotide exchange factor (GEF) 1
RARA: retinoic acid receptor alpha
RASA1: Ras p21 protein activator (GTPase activating protein) 1
RASGRF2: Ras protein-specific guanine nucleotide-releasing factor 2
RASGRP3: Ras guanyl releasing protein 3 (calcium and DAG-regulated)
RASSF1A: Ras association (RalGDS/AF-6) domain family member 1A
REC: research ethics committee
RET: ret proto-oncogene
RGMA: repulsive guidance molecule family member a
RLGS: restriction landmark genomic scanning
RNA: ribonucleic acid
ROBO1: roundabout axon guidance receptor, homolog 1
ROCK1: Rho-associated, coiled-coil containing protein kinase 1
RPL7L1: ribosomal protein L7-like 1
RPS6KA1/2: ribosomal protein S6 kinase, 90kDa, polypeptide 1/2
RPTOR: regulatory associated protein of MTOR, complex 1
RR: response rate
RRM2: ribonucleotide reductase M2
RSK4: ribosomal protein S6 kinase
RT: radiotherapy
RT-PCR: real-time polymerase chain reaction
RUNX1: runt-related transcription factor 1
RXRB: retinoid X receptor, beta
SERPINB5: serpin peptidase inhibitor, clade B, member 5
SFN: stratifin
SFRP1/2/4/5: secreted frizzled-related protein 1/family2/4/5
SH B/C1: Src homology 2 domain containing adaptor protein B/transforming protein 1
SHH: sonic hedgehog/ smoothened signalling pathway
SKI: v-ski avian sarcoma viral oncogene homolog
SMURF2: SMAD specific E3 ubiquitin protein ligase 2
SNP: single nucleotide polymorphism
SOS1: son of sevenless homolog 1 (Drosophila)
SOX17: SRY (sex-determining region Y)-box 17

SPP1: secreted phosphoprotein 1
SPRY2: sprouty homolog 2 (Drosophila)
ssDNA: single stranded deoxyribonucleic acid
STAT3/5AB/5BA: signal transducer and activator of transcription 3/5A/5B
STMN1: stathmin 1
STXBP4: syntaxin binding protein 4
SUFU: suppressor or fused homolog (Drosophila)
SVD: singular valve decomposition
SWAN: subset-quantile within array normalisation
T: thymine
TAHBSO: total abdominal hysterectomy and bilateral salpingo-oophorectomy
TAOK1/3: TAO kinase 1/3
TBCC: tubulin folding cofactor C
TBL1XR1: transducing beta-like 1 X-linked receptor 1
TCGA: the cancer genome atlas
TERT: telomerase reverse transcriptase
TFAP2A: transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
TGFA/B: transforming growth factor, alpha/beta
TGFB3: transforming growth factor beta 3
TGFBR3: transforming growth factor, beta receptor 3
TGIF: TGFB-induced factor homeobox
THBS2: thrombospondin 2
THY1: thy-1 cell surface antigen
TIMP3: tissue inhibitor of metalloproteinases-3
TIS: tumour immunostaining score
TJP2: tight junction protein 2
TLE1: transducer-like enhancer of split 1 (E(sp1) homolog, Drosophila)
TNF: tumour necrosis factor/receptor
TNXB: tenascin XB
TP53: tumour protein p53
TP53AIP1: tumour protein p53 regulated apoptosis inducing protein 1
TP73: tumour protein p73
TRAF4: TNF receptor-associated factor 4
TRIAP1: TP53 regulated inhibitor of apoptosis 1
Trk: tyrosine kinase receptor
TSG: tumour suppressor gene
TSS: transcription start site
UCL: University College London
UCLH: University College London Hospital
UTR: untranslated region
VBT: vaginal brachytherapy
VEGF(R): vascular endothelial growth factor (receptor)
VHL: von Hippel-Lindau tumour suppressor
WebGestalt: Web-based gene set analysis toolkit
Wnt: wingless-type MMTV integration site family
WNT3/3A/4/6/8B/9B/10A/16: wingless-type MMTV integration site family, member 3/3A/4/6/8B/9B/10A/16
WRN: werner syndrome, RecQ helicase-like
WT1: Wilms' tumour gene
ZNF154: also known as pHZ-92; member of the Kruppel C2H2 zinc finger protein family
ZYX: zyxin

CHAPTER 1: Endometrial Cancer, its Treatment and Molecular Aberrations

1.1. Background on Endometrial Cancer

1.1.1. Incidence

Endometrial cancer (EC) is the most common gynaecological cancer in the developed world and 4th most common cancer in women in the UK [1]. It is the 7th most common cause of death from cancer in Western Europe and is an increasing health problem due to rising incidence in both the UK and Europe [2]. The European age-standardised incidence rate has risen by 46% between 1992-1994 and 2008-2010, with the largest increase occurring in women aged 65 to 79 years [3]. For women aged 40 to 54 years, the incidence rate has been relatively stable with a smaller increase of 16% between 1999-2001 and 2008-2010.

Although an estimated 75% of patients present with surgically resectable early stage disease, approximately 20% have regional metastases and 5% have distant metastases. Of the patients with surgically resectable disease, up to 20% of these cases relapse [4, 5]. For early stage disease, the 5 year survival is 85%, though this decreases to 66% for disease with regional spread and 25% for disease with distant spread [4].

1.1.2. Risk Factors

Risk factors for EC have largely been attributed to hyperoestrogenic and genetic factors, with the increase in incidence thought to be due to more women being overweight or obese, having few or no children and being older at first birth [3, 6].

Historically, Type I EC is associated with increased levels of oestrogen (for example from obesity, diabetes and high-fat diet) as well as early age at menarche, nulliparity, late age at menopause, and tamoxifen use [7]. Recent data however suggests that the two EC types share many common etiologic factors [8] with parity, oral contraceptive use, cigarette smoking, age at menarche and diabetes associated with type I and type II tumours to a similar extent and only body mass index having a greater effect on type I than type II tumours. As such, oestrogenic risk factors may also be associated with type II tumours and the traditional type I and type II distinction may not accurately reflect characteristics of the disease.

Although EC is usually associated with sporadic mutations, about 6% of cases are caused by germline genetic alterations, namely Lynch syndrome/hereditary non-polyposis colorectal cancer (HNPCC) [9]. Lynch syndrome is an autosomal dominant hereditary cancer predisposition syndrome, caused by a germline mutation in one of the mismatch repair (MMR) genes, MLH1, MSH2, MSH6 or PMS2. EC associated with Lynch syndrome tends to occur 10-20 years earlier than sporadic cancers and in women below 50 years, 9% have one of the MMR gene mutations [9].

1.1.3. Symptoms and Diagnosis

The most common presenting symptom is irregular vaginal bleeding, predominantly in postmenopausal patients. As such, patients tend to be investigated and diagnosed at an earlier stage compared to ovarian cancer for example.

If EC is suspected, initial evaluation includes history and physical examination as well as pelvic imaging and endometrial biopsy. Historically, tumour markers including cancer antigen 125 (CA125) have also been performed and used in follow-up, though may be falsely increased with peritoneal inflammation, infection or radiation injury and may not predict recurrence in patients in absence of clinical findings [10, 11], with one study reporting CA125 elevation in only 58% of patients with recurrent disease [12]. Genetic counselling should be considered for patients less than 50 years or with a significant family history of endometrial, colorectal and/or ovarian cancer.

Diagnostic pelvic and/or transvaginal ultrasound assesses endometrial thickness in the anteroposterior dimension. A cut-off of 5mm in postmenopausal women and 15mm in premenopausal women [13] is used and tissue diagnosis via endometrial biopsy is then required. Negative endometrial biopsy in a symptomatic patient should be followed by dilation and curettage under anaesthesia [14]. Hysteroscopy may also be helpful in evaluating endometrial lesions if there is persistent or recurrent undiagnosed bleeding [15].

Additional pre-operative staging generally includes magnetic resonance imaging (MRI) of the pelvis and computed tomography (CT) or positron emission tomography (PET) of the chest/abdomen/pelvis to assess the extent of local and distant disease respectively and to assist with surgical planning. Although CT staging provides limited assessment of cervical invasion, it is useful for extrauterine spread, organ invasion, lymphadenopathy and peritoneal spread [16]. MRI provides the most accurate evaluation of the uterus including information on tumour bulk, depth of myometrial invasion (MMI), cervical involvement and extrauterine spread [17]. Evaluation of MMI is key because invasion greater than 50% is associated with an increased risk of pelvic and para-aortic lymph node metastases [18], and thus can help with planning of lymphadenectomy. PET may also show utility in lymph node staging and early detection of recurrence, as well as in investigation of equivocal lesions on CT [19].

Interestingly, the nature of the treatment provider may also impact on outcomes [20] with evidence that patients treated by gynaecologic oncologists were more likely to undergo complete staging surgery, receive adjuvant chemotherapy for advanced stage disease and have improved survival in high-risk cancers and advanced disease.

1.1.4. Atypical Endometrial Hyperplasia

In many cases, endometrial carcinogenesis is thought to follow a continuum from normal endometrium to endometrial hyperplasia without atypia, endometrial hyperplasia with atypia, through to EC. The presence of nuclear atypia is the most important indicator off EC risk in women with atypical endometrial hyperplasia (AEH), though there can be significant inter-observer variability in reporting with a concordance rate of 38-47% [21, 22].

Most commonly, AEH presents with abnormal uterine bleeding in women aged 50-54 years and is rarely found in women less than age 30 years [23]. The risk factors are the same as those for endometrioid endometrial cancer (EEC) and generally involve exposure of the endometrium to continuous unopposed oestrogen.

The risk of progression from AEH to EC is approximately 23-29% [24]. In addition, for patients diagnosed with AEH on biopsy, between 17-52% are subsequently found to have concurrent EC at hysterectomy [25, 26]. Thus, the current recommendation for management of AEH is total abdominal hysterectomy with bilateral salpingo-oophorectomy (TAHBSO). Some gynaecologic oncologists may attempt conservative hormone therapy to preserve fertility in younger women or to avoid hysterectomy in women who decline surgery or who have high-risk comorbidities.

This current management strategy will significantly over-treat a number of women who do not have and will not develop EC, but there are no current biomarkers to guide selective treatment.

1.1.5. Current Classification and Prognostic Factors in Endometrial Cancer

An expert pathology review is required to classify the tumour into the precise histological subtype, whether epithelial (endometrioid, serous, clear cell adenocarcinoma or carcinosarcoma) or stromal/mesenchymal (endometrial stromal sarcoma, high-grade endometrial sarcoma or uterine leiomyosarcoma).

Historically, the risk of recurrent epithelial EC has been determined according to clinicopathologic criteria that define two groups, Type I and Type II [27]. Type I tumours comprise up to 90% of EC and are associated with low stage, low grade, better prognosis, EEC histology, oestrogen exposure, obesity and concurrent or preceding AEH [28]. On the other hand, type II tumours are typically high stage, high grade, have a poor prognosis, have non-endometrioid (NE) histology, and occur in atrophic endometrium in older women. Although type I cancers are more common, type II cancers account for a high proportion of EC-related deaths with a higher risk of recurrence even when treated at an early stage [29]. As discussed however, the risk factors for type I and II tumours share greater similarities than originally thought [8] and molecular aberrations also cross these two subgroups, as will be discussed in

section 1.4 [30]. The current classification into type I and II tumours is also often of limited prognostic value given that approximately 20% of type I cancers recur and 50% of type II do not [31]. Applicable molecular classification of EC that accurately predicts prognosis and response in defined subgroups is urgently needed. Most clinical trials do not stratify patients according to type and both types are managed with the same chemotherapy regimens with a similar extent of benefit expected [32].

EEC resembles the epithelium of normal, proliferative endometrium and a number of variants exist, as outlined in Table 1.1 [33]**.** Serous carcinoma comprises the majority of type II cancers and histologically resembles serous ovarian cancer, with a more aggressive natural history [34].

Table 1.1: The Histological Subtypes of Endometrial Cancer

Type I: Endometrioid EC	Type II: Non-endometrioid EC
Variant with squamous differentiation Villoglandular differentiation Secretory variant Ciliated-cell variant	Serous adenocarcinoma Clear cell adenocarcinoma Carcinosarcoma Mucinous adenocarcinoma Squamous-cell carcinoma Other (mixed adenocarcinomas, transitional-cell carcinoma, small-cell carcinoma and undifferentiated carcinoma)

The prognostic value of certain laboratory parameters has also been investigated in EC. Anaemia pre-operatively has been associated with NEEC, advanced International Federation of Gynaecology and Obstetrics (FIGO) stage, lymphovascular space invasion (LVSI), cervical involvement, adnexal involvement, positive peritoneal cytology, lymph node involvement and lower 5 year overall survival (OS) compared to patients with haemoglobin levels >12g/dL [35, 36]. Similarly, pre-treatment hypoalbuminaemia has been associated with more advanced FIGO stage and higher grade, as well as with disease-free and progression-free survival (PFS) [37]. However, there is little data for the prognostic utility of anaemia and hypoalbuminaemia in the advanced setting and is not sufficiently discriminatory to be helpful for therapeutic planning.

1.1.6. Staging of Endometrial Cancer

The FIGO staging system is the most commonly used system for staging EC. It has evolved from the 1970 criteria based on pre-surgical evaluation through to the 1988 system and the current 2009 criteria based on surgical and pathologic staging criteria, with separate systems for epithelial tumours and uterine sarcomas [38, 39]. Staging is performed by a multi-

disciplinary team of radiologists, pathologists, surgeons as well as medical and clinical gynae-oncologists.

The 1988 system first modified the staging to incorporate surgical and pathologic data including histologic grade, MMI and the extent and location of extrauterine spread. The 2009 system, outlined in Table 1.2, revised differences in stage I and II disease, based on correlation with survival rates. Stage IA and IB now differentiate between less than or greater than 50% MMI and stage II includes patients with cervical stromal invasion. Stage IIIC disease was subdivided into C1 and C2 based on the absence or presence of para-aortic nodes, as survival is worse if para-aortic nodes are positive.

Compared to the 1998 system, the 2009 system better correlates with prognosis and clinical outcomes [40].

Table 1.2: The 2009 FIGO Staging System

FIGO stage	TNM	Surgical-Pathologic Findings
	TX/NX	Primary tumour/regional lymph nodes cannot be assessed
	T0/N0/M0	No evidence of primary tumour/regional lymph nodes/distant metastases
	Tis	Carcinoma in situ (preinvasive carcinoma)
I	T1	Tumour confined to the corpus uteri
IA	T1a	Tumour limited to endometrium or invades less than one-half of myometrium
IB	T1b	Tumour invades one-half or more of the myometrium
II	T2	Tumour invades stromal connective tissue of the cervix but does not extend beyond the uterus
IIIA	T3a	Tumour involves serosa and/or adnexa (direct extension or metastasis)
IIIB	T3b	Vaginal involvement (direct extension or metastasis) or parametrial involvement
IIIC		Metastases to pelvic and/or para-aortic lymph nodes
IIIC1	N1	Regional lymph node metastasis to pelvic lymph nodes
IIIC2	N2	Regional lymph node metastasis to para-aortic lymph nodes, with or without positive pelvic lymph nodes
IV		Tumour involves bladder and/or bowel mucosa and/or distant metastases
IVA	T4	Tumour involves bladder mucosa and/or bowel
IVB	M1	Distant metastases (includes metastasis to inguinal lymph nodes, intra-peritoneal disease, or lung, liver or bone. It excludes metastasis to para-aortic lymph nodes, vagina, pelvic serosa or adnexa)

TNM: tumour node metastasis staging classification system

1.2. Treatment

This section covers the surgery, radiotherapy (RT) and chemotherapy options for EEC and NEEC. Less common pathologic subtypes of EC including uterine sarcomas have a different natural history and as this thesis does not analyse samples from this patient group, uterine sarcomas are not discussed further.

1.2.1. Surgery

Surgery is the mainstay of treatment for localised EC, though the optimal techniques are still debated. Traditionally, surgical treatment involves TAHBSO with or without lymphadenectomy for early stage disease and including lymphadenectomy for higher stage disease.
Hysterectomy may be performed via laparotomy, vaginally or via laparoscopy. Laparoscopy for stage I-IIA disease, compared to laparotomy, is associated with shorter hospital stay, few moderate-to-severe postoperative adverse events, improved body image, and similar 3 year recurrence rate and estimated 5 year OS [41, 42]. Thus it is a reasonable alternative to laparotomy for treatment and staging of early-stage EC.
Interestingly, timing of surgery also appears to impact survival with recent evidence demonstrating that wait times of <2 weeks and >12 weeks were adversely prognostic for 5 year survival [43]. Wait time <2 weeks was generally associated with surgery being performed for acute issues or more advanced disease, contributing to poorer outcomes, while wait times beyond 12 weeks resulted in a decreased 5 year OS of 71.9% compared to 81.8% if the wait time was 2-6 weeks.
As part of surgery, visual assessment of the peritoneal, diaphragmatic and serosal surfaces with biopsy of any suspicious lesions and peritoneal cytology is important to exclude extrauterine disease. Pelvic lymph node dissection for staging may include the external iliac, internal iliac, obturator and common iliac nodes as well as para-aortic nodal dissection for high-risk tumours. The role of lymphadenectomy is to stage disease, define prognosis and decide on adjuvant therapy. There is conflicting data however on its therapeutic role [44, 45] as well as its prognostic role [46] and as such, a selective approach to pelvic and para-aortic lymphadenectomy is recommended, to avoid systematic overtreatment and associated morbidity [47]. In early stage disease confined to the uterus, there is no demonstrated benefit in disease-free nor OS between pelvic lymphadenectomy compared to standard surgery and no lymphadenectomy [48] [44]. One approach is to base decisions regarding lymphadenectomy on preoperative and intraoperative findings, including degree of MMI, tumour size less than 2cm and grade 1-2 histology [49], though this may be difficult to assess intraoperatively [50]. Sentinel lymph node mapping has also been investigated and appears feasible [51] but its utility needs to be established in clinical trials. Thus, the identification of

reproducible, validated biomarkers to guide optimal treatment from the earliest stages of disease is imperative.

In the setting of locally advanced or metastatic disease, maximal debulking should be performed and palliative surgery should still be considered in patients with good performance status. For suspected or gross cervical involvement, radical or modified radical hysterectomy is recommended along with lymph node dissection, bilateral salpingo-oophorectomy and peritoneal lavage for cytology. For intra-abdominal extra-uterine disease, surgical intervention is still warranted for maximum debulking aiming to eradicate measurable residual disease, which may improve survival [52]. In this scenario, pelvic external beam radiotherapy (EBRT), vault brachytherapy (VBT) and/or chemotherapy may be used to reduce tumour burden and allow delayed surgery. For extra-abdominal disease, palliative TAHBSO may still be considered and combined with chemotherapy, RT and/or hormone therapy.

Pathologic assessment of any surgical specimen should include tumour size, location (fundus vs lower uterus), histology subtype and grade, LVSI, ratio of depth of myometrial/stromal invasion to myometrial thickness, cervical stromal or glandular involvement, fallopian tubes/ovarian involvement, peritoneal cytology and nodal status. Screening with immunohistochemistry (IHC) and/or microsatellite instability (MSI) for inherited MMR gene mutations should be considered in patients deemed at risk of Lynch syndrome [9].

1.2.2. Adjuvant Treatment and Surveillance

Risk of recurrence after surgery for early-stage disease is currently defined by a number of factors including higher age, histological subtype, grade 3 histology, LVSI, lower uterine (cervical/glandular) involvement and tumour diameter >2cm. Adjuvant RT and/or chemotherapy are then considered based on these factors. In general, as grade, MMI and risk factors increase, more aggressive adjuvant treatment is given. However, criteria for guiding these treatment decisions remain poorly defined.

Overall, adjuvant RT improves disease control in the pelvis but does not appear to improve OS [53] [54, 55]. For stage I disease, although grade 1 disease can be managed with a watchful waiting approach, VBT or pelvic EBRT or both for higher grade disease is recommended and chemotherapy should also be considered for grade 3, stage IB disease [56]. The post-operative radiation therapy in endometrial cancer (PORTEC) 2 study showed that for high-intermediate risk stage I or IIA disease, VBT was as effective as EBRT for local control with less gastrointestinal toxicity [55]. Importantly, for women less than 60 years and with stage I disease, there is no difference demonstrated in OS between VBT and EBRT but women younger than 60 years have higher mortality rates and a greater risk of secondary cancer after EBRT [57]. With stage II disease, either VBT or EBRT may be used at a lower grade but both are

required with the consideration of chemotherapy for higher grade disease [56]. A number of trials have investigated, yet not definitively answered the question regarding the role of combination chemotherapy and pelvic radiotherapy for stage I-III disease. Many of these were small and underpowered to detect an OS difference, though a combined analysis of the ILIADE-III and NSGO-EORTC [56] trials showed an improvement in PFS with the addition of adjuvant chemotherapy to radiation. The role of concurrent chemoradiotherapy versus radiotherapy alone in high risk stage I (stage IA with invasion, grade 3, LVSI; stage IB grade 3), stage II, stage III and serous/clear cell stage IA (invasive), IB, II or III will be reported in PORTEC-3 [58]. This study will also help answer the question regarding the role of adjuvant chemotherapy for Stage IB and II, grade 3 disease, as these patients continue to have a risk of distant metastases despite pelvic RT. Previous studies have demonstrated improved PFS but no OS benefit and further studies are pending [56, 59]. Although the role of adjuvant chemotherapy in high-grade, uterine-confined disease is yet to be confirmed, for stage III disease, the overall evidence suggests a benefit from adjuvant chemotherapy with RT [60].

After adjuvant therapy is completed, post-treatment surveillance occurs every 3-6 months for the first 3 years and then every 6-12 months thereafter [61]. The value of more intensive surveillance and ancillary testing has not been demonstrated [62] and there remains a significant risk of recurrence for later stage disease. If disease recurs, it tends to do so within the first 3-5 years and is symptomatic in up to 75% of patients. In up to 60% of patients, disease recurrence will occur at a distant location and prognosis is limited, regardless of timing.

In terms of follow-up interventions, asymptomatic recurrences may be detected by physical exam in 5-33%, vaginal vault cytology in 0-4%, chest X-ray in 0-14%, abdominal ultrasound in 4-13%, abdominal/pelvic CT in 5-21% and CA125 in up to 15% [62]. There are no validated tumour markers however used for monitoring patients routinely during follow-up for EC. CA125 has been demonstrated to be increased in approximately 50% of patients at the time of recurrence [12] while up to 35% of women in follow-up may have elevated CA125 but no evidence of recurrence [10, 63]. Thus, although the use of CA125 for monitoring in EC has been extrapolated from the ovarian cancer setting and may be used by clinicians, there is in fact limited evidence of its utility.

1.2.3. Treatment of Recurrent or Metastatic Disease

If there is a suspicion of recurrent disease, it is important to establish whether this is localised and still potentially curable or whether there are distant metastases in which case treatment will be palliative.

Isolated vaginal recurrence can be treated with RT and/or surgery and 5-year survival may be up to 50-70% [64]. Radical surgery, specifically pelvic exenteration, has been reported to have a 5-year survival up to 20% [65]. Hormonal and/or chemotherapy should also be considered for isolated relapse, either as primary treatment if the disease is unresectable or as additional treatment if surgical resection is achieved. For patients who have received adjuvant RT however, it is uncommon for recurrences to be confined to the pelvis.

There have been a number of studies of the optimal combination of cytotoxic therapy, though prognosis is poor with limited response rates (RR) [60, 66] often of short duration and with a reported median OS between 4.1–17.6 months [66-68]. Among women who are chemotherapy-naïve, RR over 40% have been reported with doublet chemotherapy [66]. Historically, doxorubicin/cisplatin was the combination of choice with a reported RR of 42%, PFS of 5.7 months and OS of 12 months [69]. Three-drug regimens were compared to this doublet but did not consistently show superiority and have higher toxicity, therefore are not routinely used [70, 71]. A taxane and platinum combination is now the preferred first-line chemotherapy option, with a reported RR of 40-67% and OS up to 32 months [72, 73]. Again, comparison with a triplet regimen in the Gynaecologic Oncology Group (GOG) 209 study showed similar outcomes and more tolerable toxicity with the doublet arm of paclitaxel and carboplatin, compared to doxorubicin, paclitaxel and cisplatin [73]. Single agent chemotherapy is also an option if there is concern regarding tolerability of doublet treatment and paclitaxel, cisplatin, carboplatin, doxorubicin, liposomal doxorubicin, topotecan and docetaxel are all options, though paclitaxel on a weekly schedule is often preferred due to better tolerance and some evidence suggesting improved efficacy [74]. These agents can also be considered in the second line setting however disease recurring after first-line chemotherapy is more chemoresistant. This is an area of significant unmet clinical need and there are efforts in clinical trials to define groups of patients that may benefit from novel therapies in this setting.

Hormonal therapy is used most frequently in endometrioid histologies, namely progestogens agents, but also daily tamoxifen with intermittent weekly megestrol [75] and more recently, aromatase inhibitors [76]. Single-agent progestogens have demonstrated a RR of approximately 20%, with a higher response in progesterone receptor (PR) positive and lower-grade tumours [77]. Reported RR with megestrol and tamoxifen is 33% with a PFS up to 3 months and OS up to 5 years in those women with high expression of oestrogen receptor (ER) on their tumours. Retrospective review of aromatase inhibitors demonstrated a clinical RR up to 70%, though radiological data was limited in this study. These agents may play a role particularly for well-differentiated tumours with expression of ER and PR, if there is a long disease-free interval and if there are pulmonary metastases.

More recently, the focus of drug development has shifted to molecularly targeted agents, though correlation with predictive and prognostic biomarkers has been limited. Anti-angiogenic agents, mammalian target of rapamycin (mTOR) pathway inhibitors and fibroblast growth factor receptor (FGFR) inhibitors have been studied in EC with early evidence of efficacy. Bevacizumab, a vascular endothelial growth factor (VEGF) inhibitor, has shown a RR of 13.5% and OS of 10.5 months in a phase II trial for persistent or recurrent EC [78] while aflibercept, another anti-angiogenic agent, has also demonstrated clinical activity in a phase II study but there was significant toxicity with two treatment-related deaths [79]. There has been limited assessment of small molecular inhibitors of angiogenesis.

The use of mTOR inhibition has been attractive in EC due to the frequency of aberrations in the phosphoinositide 3-kinase (PI3K) pathway. Both temsirolimus and everolimus have demonstrated efficacy and acceptable tolerability in phase II studies with partial response rate (PR) of 4-5% and stable disease (SD) at 4 months being 48% for temsirolimus and for everolimus, 36% at 3 months in the recurrent disease setting [80-82]. Ridaforolimus has also been assessed as single agent treatment in recurrent or persistent EC in a phase II study. Clinical benefit (PR or prolonged SD) was seen in 29% with a 6 month PFS rate of 18% [83]. There are also planned studies of adding an mTOR inhibitor to chemotherapy or hormone therapy in the advanced or recurrent setting.

Phase II studies have also assessed agents targeting epidermal growth factor receptor (EGFR) over-expression and human epidermal growth factor receptor 2 (HER2) amplification, which occur in 50-80% and 20% of ECs respectively, though limited efficacy has been seen at this stage [84, 85].

Targeting the FGFR pathway may also show promise with evidence of FGFR2 mutations in 12-16% of EC and pre-clinical data that FGFR2 blockade leads to cell death [86]. A phase II study is underway with Dovitinib, an anti-VEGFR/FGFR tyrosine kinase inhibitor, as second-line therapy in patients with FGFR2-mutated or non-mutated advanced and/or metastatic EC [87].

To improve understanding of the molecular aberrations driving endometrial tumorigenesis and better develop translational research and drug trials, GOG has created a large biospecimen repository to provide tissue, serum and epidemiologic survey data as well as clinical information. This aims to define predictive and prognostic biomarkers and identify novel targets to improve therapy [68]. Developing the optimal techniques to analyse tissue from this repository from FF, FFPE and blood samples will be key in maximising its therapeutic relevance and utility.

The identification and validation of biomarkers from blood and tumour tissue is a priority in EC to allow better selection and administration of targeted agents and define prognosis. For

example, aberrations in the PI3K pathway are well documented and yet the studies with temsirolimus and everolimus failed to correlate markers of PI3K pathway activation with patient response [80, 81] . Understanding the tumour biology and identifying biomarkers that accurately define this will allow better translation of targeted therapies into the clinical setting with the overall aim of improved efficacy. This thesis focuses on two methods to define novel biomarkers in EC, one based on peripheral blood samples and one on archival FFPE tissue analysis.

1.3. The Rationale for Circulating Tumour Cells in Endometrial Cancer

1.3.1. Background

Circulating tumour cells (CTCs) are epithelial cancer cells known to circulate in the peripheral blood of patients including in breast, prostate, ovarian, colorectal and lung cancers, and are not detected in healthy individuals [88]. One of the earliest reports of their presence in the blood of a patient with cancer was in 1869 by Thomas Ashworth, an Australian physician [89]. In a post-mortem examination, he identified circulating cells in the lower limb venous system that were morphologically similar to the cancer at a different site and concluded that "cells identical with those of the cancer itself being seen in the blood may … throw some light upon the mode of origin of multiple tumours existing in the same person".

CTCs are thought to be a mixture of cells including cancer stem cells capable of seeding, recirculation and evolution in metastatic clones, as well as non-viable cells shed from primary or metastatic tumours [90]. They are rare events, estimated to account for 1 cell in 10^9 nucleated cells [91].

As CTCs can be analysed from a peripheral blood draw, they are an attractive option for further investigation and validation as a surrogate marker across cancer types. A surrogate biomarker is defined as a biomarker intended to substitute for a clinical endpoint, while a biomarker is defined as a characteristic that is objectively measured and evaluated as an indication of a normal biologic, pathogenic process or pharmacologic response to a therapeutic intervention [92]. If biomarkers, such as CTC count or molecular profile, can be substituted for longer-term endpoints such as response to therapy, PFS or OS, then they may offer greater efficiency in clinical studies, with the generation of useful results based on shorter study duration or smaller sample size [93].

A number of methods for detecting CTCs have been described based on different physical and molecular properties of these cells, but the only regulatory approved and validated detection method is the Veridex CellSearch platform. This detection technique relies on the use of a ferrofluid epithelial cell adhesion molecule (EpCAM)-based capture reagent and a further

series of immunofluorescent reagents to create cell images that are classified by the operator [88], described in detail in Chapter 2.

The expression of EpCAM on the CTCs forms the basis of the enrichment method used in the CellSearch system. EpCAM is an epithelial cell adhesion molecule, frequently overexpressed in primary and metastatic adenocarcinomas, and has been demonstrated across a broad range of tumours, particularly colorectal, pancreatic, gastric and prostate cancers [94] with a poor prognostic association in gallbladder, ovarian [95] and pancreatic cancers [96]. If the tumour either does not express EpCAM or there is a reduced expression of EpCAM with metastases or on peripheral circulating cells, than there will be limitations to CTC detection with this method. While there is some evidence for correlation of EpCAM expression between primary tumour tissue and metastatic tissue [97], other studies demonstrate variation in its gene expression and loss of epithelial antigens as occurs during epithelial to mesenchymal transition (EMT) in the development of metastases [98]. As such there is an argument for testing EpCAM expression in metastatic lesions rather than primary tissue to determine the utility of CTC enumeration in certain cancer types and patients.

In regards to EC, 65- 80% of tumours demonstrate strong EpCAM staining, across all grades and stages [97, 99].

Initial analysis of the CellSearch technology demonstrated that CTCs were detectable in approximately 44% of patients with metastatic carcinoma with CTCs ≥2 in 57% of prostate cancers, 37% of breast and ovarian cancers, 30% of colorectal cancers and 26% of other cancers [88]. Importantly, only 5.5% of healthy subjects demonstrated 1 CTC and only 7.5% of subjects with non-malignant disease demonstrated 1 CTC. 0.3% of healthy subjects or with non-malignant disease had ≥2 CTCs per 7.5ml blood.

1.3.2. Current CTC Data in Breast, Prostate and Colorectal Cancer

The CellSearch Epithelial Kit and CellSearch Analyser were approved for use by the US Food and Drug Administration (FDA) in 2004 based on data generated in breast, prostate and colorectal cancers, and a summary of these data are presented here.

In breast cancer patients both in the metastatic and adjuvant setting, a CTC count ≥5 per 7.5ml of blood, occurring in just under 50% of patients, has been shown to be associated with a shorter median PFS and OS [100-103]. On multivariate analysis in the metastatic breast cancer setting, the CTC count at baseline and first follow-up was the strongest predictor of PFS and OS compared to other clinical and pathological factors including ER and PR status, ECOG performance status, time to metastasis, HER2 status and type of therapy. This predictive value persisted not only with newly diagnosed disease [104], but also at each follow-up time point

during therapy [105] and CTCs have been demonstrated to predict OS better than imaging [101]. In the adjuvant setting in early breast cancer, the presence of CTCs both before and after chemotherapy was also associated with poor disease-free survival, breast cancer-specific survival and OS [103].

Similarly, in castrate-refractory prostate cancer, the presence of CTCs has been demonstrated to predict OS and response to therapy and is increasingly being incorporated in clinical trials, both as an eligibility criteria and as a predictive marker to correlate with patient symptoms and radiographic findings.

A CTC count ≥5 has been shown to discriminate between favourable and unfavourable groups and identify patients with a shorter OS across studies in both chemotherapy and hormonal agents [106, 107]. Those patients whose CTC count decreases from ≥5 to <5 also have a better prognosis and survival while those whose CTC counts increase from <5 to ≥5 do worse. Moreover, changes in CTC counts may predict OS better than PSA decrement [108]. CTCs measured during treatment with abiraterone for patients with prostate cancer have also been shown to fulfil Prentice's criteria for surrogacy for OS [109] based on the following parameters; that the treatment was prognostic for survival, that the treatment was prognostic for the CTC count, that CTCs were prognostic for survival and that the full effect of the treatment on survival was captured by the CTC count [109, 110].

The CTC count in colorectal cancer has also demonstrated prognostic and predictive significance, with unfavourable and favourable prognostic groups defined based on CTC levels of ≥3 or <3 respectively [111], though only 26% of patients have ≥3 CTCs at baseline. Patients with CTCs ≥3 at baseline have a shorter median PFS and OS with an improvement in PFS and OS if CTCs decrease from ≥3 to <3 after 3-5 weeks on treatment [112]. The association of CTC count ≥3 with inferior PFS and OS is also consistent across a number of subgroups including patients receiving first- or second-line therapy, having liver involvement, age ≥65 years and ECOG performance status [112].

The evidence for CTCs as a surrogate marker for survival has the potential to streamline clinical trial and drug development, both to improve efficiency and shorten timelines in development of novel anticancer agents and assessing at earlier time points whether patients should continue with a particular therapeutic strategy or not. Assessment of CTCs may allow earlier detection of relapse as well as molecular analysis, potentially allowing earlier intervention with an appropriately targeted therapy. Similarly, if early assessment of CTC count after initiation

of treatment indicates a favourable or unfavourable outcome, it raises the argument that this should be used in clinical practice to trigger a change in treatment.
Thus, clinical trials are underway in a number of solid tumours to assess the utility of CTC changes to drive therapeutic changes and where possible incorporate molecular analysis [113].

1.3.3. CTCs in Gynaecologic Malignancies

CTCs in gynaecologic malignancies, particularly ovarian cancer, have been demonstrated in a limited percentage of patients and have not shown an association with survival according to current data. While some studies report correlation between CTC counts and prognosis in recurrent ovarian cancer, others report no such correlation and investigation is ongoing into other molecular techniques both for enumeration and characterisation. CTC analysis, for example, in a phase III study of pegylated liposomal doxorubicin with trabectedin demonstrated baseline CTCs >2 in 14.4% of patients (31/216), which correlated with a higher risk for progression and decreased survival and was independent of established factors such as CA125, platinum sensitivity status and ECOG performance status [114]. However, an analysis of 50 patients with platinum resistant and sensitive recurrent ovarian cancer at Dana-Farber Cancer Institute showed low numbers of CTCs in these patients with no correlation with CA125, treatment response, type or duration of chemotherapy [115].
There are limited published data regarding CTC enumeration in endometrial and cervical cancers and there is a need in both cancer types to develop biomarkers to better guide therapeutic management.

1.3.4. The Role of CTCs in Clinical Trials

With the increasing evidence for the predictive and prognostic role for CTCs across a number of tumour types, longitudinal assessment of CTCs is being incorporated into clinical trials to determine their utility as a biomarker with novel targeted therapies at different stages of disease. A longitudinal approach may also enable detection of molecular changes that may be driving resistance or progression.
CTC enumeration has been investigated in regards to decision making for patients entering phase I oncology trials for example, and patients with a higher CTC count have been shown to have a greater risk of death [116, 117].
In the breast cancer setting, a number of phase III trials are ongoing that incorporate CTC count into the treatment decision algorithm. In the metastatic setting, CTCs are under evaluation to guide both change of treatment and use of chemotherapy versus endocrine therapy [118]. In the adjuvant setting, trials are assessing HER2-positive CTCs in HER2-

negative primary tumours and testing the role of HER2-directed therapies in these patients [118].

In the prostate cancer setting, some trials of novel agents are including CTC count ≥5 as an eligibility criteria to select a poorer prognostic group that may demonstrate the utility of the agent in a more time-efficient manner [119].

1.3.5. Molecular Characterisation of CTCs

Molecular characterisation of CTCs is the next step beyond enumeration that could guide clinical decision making and targeted drug development, with its ability to provide longitudinal assessment of a tumour's molecular profile and possible causes of drug resistance and tumour progression. This is currently an experimental approach with much promise but not routinely used in any solid tumour type.

There are a number of different technologies that describe isolation of DNA from CTCs and subsequent molecular characterisation. Within the CellSearch system, there is a fourth channel in which a fluorescein-conjugated antibody can be applied that is then analysed along with the routine CTC morphological and staining features. Using a fourth channel antibody has been investigated in this study.

1.4. Molecular Aberrations in Endometrial Cancer

1.4.1. Genetic Aberrations

1.4.1.1. Background

The most common genetic aberrations in EC have been found in the PI3K pathway, but changes in other molecular pathways do co-exist, as outlined in Table 1.3 [33]. This demonstrates the differences between Type I and II EC, though more recent studies suggest the reclassification of EC based on genomic characterisation beyond these two groups to assist with treatment stratification [30].

In general, type I cancers express ER and PR, are associated with hyperoestrogenic risk factors and demonstrate MSI and mutations of kirsten rat sarcoma viral oncogene homolog (KRAS), phosphatase and tensin homologue (PTEN), phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (PIK3CA) and β-catenin. Type II tumours are generally aneuploid and historically, were not thought to be related to hyperoestrogenic factors nor changes in the ER pathway. They have been associated with genetic aberrations in the tumour suppressor genes CDKN2A and p53 and the HER2 oncogene [33]. There is significant overlap however, so that these definitions cannot be routinely clinically applied.

As previously outlined, EC as a result of Lynch syndrome accounts for up to 6% of all cases and patients with this syndrome have a 50-fold lifetime risk of developing EC. MSI can be tested on

archival tumour tissue, and in hereditary cancer, is associated with a germ line mutation in the MMR genes, while in sporadic tumours, is associated with hypermethylation of the MLH1 promoter.

Although some of these alterations have demonstrated prognostic implications, they generally have not assisted with treatment choice and unlike in breast, lung and colorectal cancers, there are no genetic aberrations identified that guide treatment decisions.

Table 1.3: Molecular Aberrations in Endometrial Cancer

Alteration	Role	Type I (%)	Type II (%)
PI3K/AKT/mTOR pathway PIK3CA mutation (Exon 9/20) [120, 121] PIK3CA amplification [122] AKT mutation [123] PTEN loss of function [124, 125]	Oncogene	 ~30 (7-15.5/10-34) 2-14 3 83	 ~20 (0/21) 46 0 5
KRAS mutation [126, 127]	Oncogene	11-26	2
FGFR2 mutations [86, 128, 129]	Oncogene	12-16	1
HER2 overexpression [130] HER2 amplification [131]	Oncogene	3-10 1	32 17
Microsatellite instability [132, 133]	MMR genes	20-45	0-5
Nuclear accumulation of β-catenin [134]	Oncogene	18-47	0
E-cadherin loss [135, 136]	TSG	5-50	62-87
p53 mutation [127, 137]	TSG	~20	~90
Loss of function of p16 [138]	TSG	8	45

TSG: tumour suppressor gene, MMR: mismatch repair
[139]

The majority of studies outlined focus on reporting single genetic or pathway based aberrations, such as in the PI3K or WNT pathways, MSI or hormone receptor signalling, and are outlined in further detail below. However, with improvement in next generation sequencing techniques and other technologies, there is increasing information regarding the crosstalk between multiple pathways and how this may impact on treatment delivery and drug resistance.

For example, analysis of multiple oncogenic pathways in patients selected from the PORTEC-2 trial demonstrated that the presence of two or more aberrations was associated with decreased disease-free survival [140]. The presence of two or more aberrations affecting the PI3K, Wnt/β-catenin, p53 and MSI pathways identified patients with clinically aggressive disease, compared to patients with no alterations in these pathways that had a very low risk of disease progression.

The Cancer Genome Atlas (TCGA) Research Network [30] has performed integrated genomic, transcriptomic and proteomic characterisation of EC fresh frozen (FF) tissue using whole-

exome sequencing, whole-transcriptome sequencing, genome-wide copy number (CN) analysis, expression profiling, reverse-phase protein array, methylation profiling and MSI assessment. They interrogated 186 endometrioid, 42 serous and 4 mixed-histology EC in an integrated manner and suggested a novel classification system based on this analysis. Some of their findings were similar to those previously reported in terms of the frequency of PI3K pathway aberrations and in fact, EC was found to have more frequent mutations in the PI3K pathway than any other tumour type studied by the TCGA to date. EEC was also associated with PTEN, CTNNB1, PIK3CA and KRAS mutations as well as ARID1A and ARID5B, involved in chromatin-remodelling and controlling gene expression. They confirmed the frequency of p53 mutations in NEEC, found in 91%.

However, they also proposed a new classification based on four categories: polymerase epsilon (POLE) ultramutated (7%), MSI/hypermutated (28%), CN low (endometrioid, 39%) and CN high (serous-like, 26%). POLE, a catalytic subunit of DNA polymerase epsilon involved in DNA replication and repair, was a novel mutation identified and occurred in both EEC and NEEC with an association with better PFS. Interestingly, serous-like EC clustered into the CN high group, as did 25% of tumours classified as high-grade EEC, which may have treatment implications for some EEC patients, as this group had a worse PFS. This CN high group showed focal amplifications in MYC (8q24.12), HER2 (17q12), CCNE1 (19q12), FGFR3 (4p16.3) and SOX17 (8q11.23). The CN low group showed 1q amplification and was made up of predominantly EEC.

Thus, an integrated approach to molecular characterisation rather than single pathway analysis may yield novel and more clinically useful information on classification and appropriate targeted therapies. Much of the literature in EC to date however has focused on specific pathways and these are outlined in further detail in the following 5 sections.

1.4.1.2. The PI3K pathway in Endometrial Cancer

The PI3K/AKT/mTOR pathway is one of the most studied pathways in EC, with evidence of aberrations in both type I and II cancers including oncogenic PIK3CA mutations, PTEN loss of function and upstream changes in KRAS, FGFR2, HER2 and EGFR leading to PI3K pathway activation. Reproducible, predictive, time- and cost-efficient markers of PI3K pathway activation however remain elusive. Changes in the PI3K and upstream pathways are summarised in Table 1.4 and Figure 1.1.

PTEN loss of function and subsequent activation of the PI3K pathway is the most common aberration currently reported in EC. This loss of function can be due to PTEN mutations as well as PTEN gene promoter methylation, microRNA regulation of the PTEN gene or alterations of PTEN protein stability and degradation mechanisms [124].

There are contradictory findings regarding the prognostic implications of PTEN loss of function and challenges with methodology and interpretation of PTEN IHC may contribute to this. PTEN expression by IHC can yield varied results due to difficulties with sensitivity and specificity of anti-PTEN antibodies when applied to FFPE tumours [141]. Although there are limitations to the reproducibility of PTEN loss by IHC, detection of functional PTEN loss by sequencing alone will not detect all causative mechanisms [142]. Thus, alternatives to PTEN IHC in detecting the spectrum of changes that can contribute to PI3K pathway activation have been investigated.

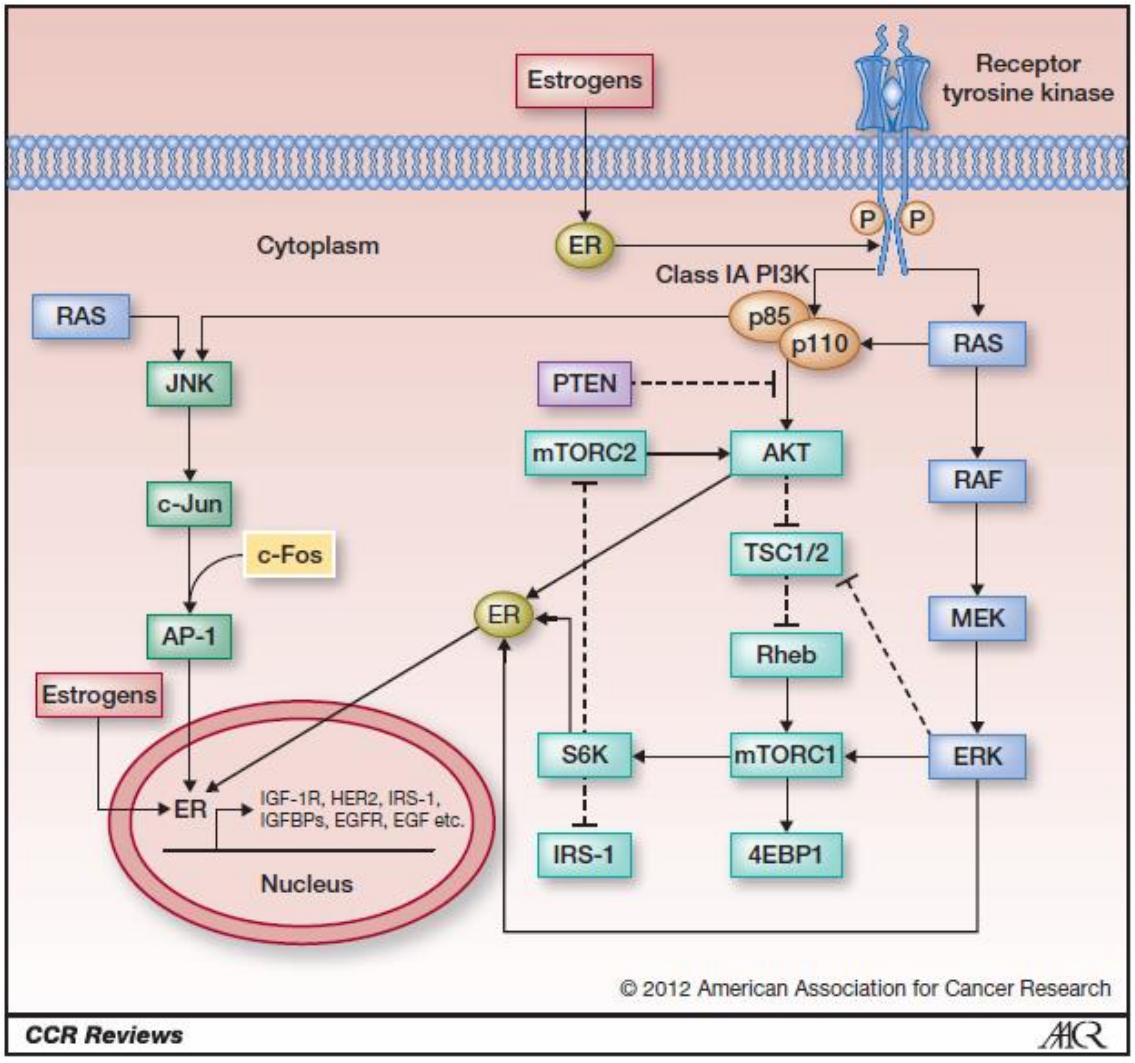
Table 1.4: Examples of PI3K Pathway and Upstream Pathway Aberrations

Author [ref]	Aberration	Frequency	EC type	Significance	n	Associated with
Catasus [121]	PIK3CA mtn Exon 9 (Helical) Exon 20 (kinase domain)	29%	EEC	Exon 9 ass'd with G1 & MMI<50% Exon 20 ass'd with G3, MMI>50%, cerv invasion	109	MSI Mtns in PTEN, CTNNB1, KRAS, BRAF
Oda [120]	PIK3CA mtn Exon 9 Exon 20	36%	No ass'n	No ass'n with grade, stage or prognosis	66	Mtns in PTEN
Rudd [143]	PIK3CA mtn Exons 1-7	16%	EEC > NEEC		108	Mtns in PTEN & KRAS
Mutter [124]	PTEN mtn/loss	83/61% vs 55/75% vs 0	EEC vs AEH vs normal	PTEN loss of function is an early event in EC & ass'd with oestrogen exposure	84	-
Mackay [144]	PTEN loss (IHC)	55%	57.9% EEC 42.9% NEEC	PTEN loss ass'd with improved survival in advanced disease	128	No ass'n with MSI
Kanamori [145]	PTEN loss or mixed staining (IHC)	65.3%	EEC only	PTEN loss ass'd with worse survival in advanced disease. No ass'n with age, FIGO stage, MMI or grade	98	ne
Birkeland [146]	KRAS mtn KRAS amp'n	14.7% 3 vs 18%	EEC>NEEC P vs M	Ass'd with low grade & obesity Ass'd with poor outcome, FIGO stage, NEEC, high grade.	414 P 61 M	ne

Dutt [86]	FGFR2 mtn	12%	EEC>NEEC	Oncogenic in cell lines	40	Co-occur with CTNNB1,PTEN & PIK3CA, not with KRAS mtns
Konecny [131]	HER2 amp'n (FISH) EGFR exp'n (IHC)	17 vs 1% 34 vs 46%	NEEC vs EEC	No ass'n with survival Ass'd with poor survival in NEEC	279	ne

EEC: endometrioid endometrial cancer, NEEC: non-endometrioid endometrial cancer
Mtn: mutation, ne: not evaluated, ass'n: association, amp'n: amplification, exp'n: expression, P: primary, M: metastatic. Gene abbreviations are outlined in the Abbreviations section.

Figure 1.1: An overview of PI3K/Akt/mTOR pathway signalling and cross-talk with other pathways relevant to endometrial cancer [147]



Solid lines represent activation and dashed lines represent inhibition.

1.4.1.3. The Role of Stathmin in Endometrial Cancer and the PI3K pathway

Stathmin is a cytosolic phosphoprotein expressed in all tissues and plays an important role in the regulation of the microtubule cytoskeleton and cell cycle [148]. It is involved in protein activity and cell division, intracellular transport, cell motility and regulation of apoptosis [149] and MAPK and PI3K pathway kinases can activate stathmin [150]. Stathmin is overexpressed on IHC in several malignancies and has been demonstrated to have prognostic significance in ovarian [151], cervical [152], breast [153, 154], prostate [155], urothelial [156], gastric [157] and oral squamous cell carcinoma [158] and predictive value in preclinical studies of breast, prostate and leukemic cancer cells [159-161]. Stathmin has been shown to act as a surrogate marker of PI3K pathway activation, is associated with PTEN loss as well as PIK3CA amplification and/or overexpression [122, 162] and there is also evidence that it is a better correlate of PI3K pathway activation than PTEN loss [162].

In the clinical setting, stathmin is overexpressed in between 27-57% of EC, can predict EC lymph node metastases [163] and is associated with NEEC, higher grade as well as lower recurrence-free and disease-specific survival [164]. Even in the otherwise low-risk EEC subgroup, it still carries prognostic value [122]. As there is evidence that it can change between primary and recurrent tumours, reassessment on progression may be important in treatment selection [165]. There is also evidence that stathmin may be a predictive marker for response to taxane chemotherapy [151, 166, 167], including in EC, where high stathmin levels have been associated with a poor response to paclitaxel containing chemotherapy and reduced disease specific survival [168].

In addition, phospho-stathmin (pStathmin, serine38) has been shown to be a biomarker of increased tumour cell proliferation and poor prognosis in endometrial cancer and in fact may be a stronger prognostic indicator than stathmin [169]

Thus, stathmin is a highly relevant marker in EC, with evidence as both a prognostic and predictive biomarker when overexpressed on IHC. Based on current data, it may play a role in stratifying patients to lymph node sampling, though this requires further validation. Furthermore, as it can potentially change over time and may reflect PI3K pathway activation in this fashion, longitudinal assessment of it and pStathmin's activity may be important to further establish its role as a predictive marker of PI3K activation and in monitoring response to PI3K inhibition. In this study, a method to detect stathmin in CTCs has been established.

1.4.1.4. The Wnt pathway

Wnt pathway aberrations are also common in EC, with different aberrations documented in EEC and NEEC involving the cadherins, cell-surface glycoproteins, and catenins, both of which combine to mediate cell-cell adhesion. However, specific drug targeting of this pathway is not

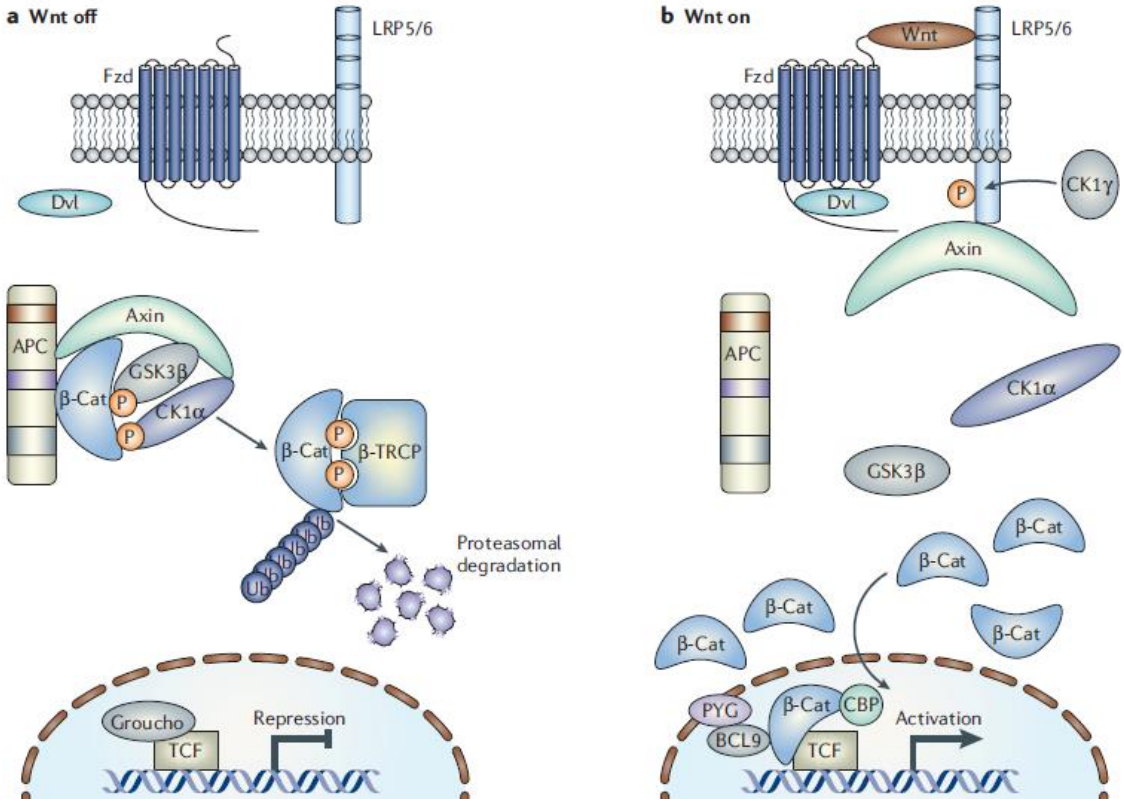
defined, unlike PI3K signalling. E-cadherin is linked to cytoskeletal actin filaments through α, ß or γ catenins, key in the Wnt signal transduction pathway [134]. Reduced E-cadherin expression, measured by IHC, is found in 5-53% of EEC and 62-87% of NEEC [135, 136] and such tumours are more likely to be poorly differentiated, have cervical extension, positive peritoneal cytology and adnexal spread and be of more advanced stage when compared with E-cadherin positive tumours [135, 136].

CDH1 is a tumour suppressor gene that codes for E-cadherin [170]. LOH of the gene occurs in 57.1% of NEECs and 22.5% of EECs, with CDH1 promoter hypermethylation in a further 21.2% of ECs [136]. Hypermethylation of the CDH1 promoter with subsequent gene silencing, is associated with low expression of E-cadherin in EC as well as higher grade, MMI, higher FIGO stage and shorter 5-year survival [171]. Aberrations of ß- and γ-catenins and resultant Wnt pathway disruption has also been implicated in EC [172].

Although genetic abnormalities in the Wnt pathway are common in EC, so far they have not translated into novel therapies nor treatment stratification.

Figure 1.2 illustrates an overview of the Wnt signalling pathway.

Figure 1.2: An overview of the Wnt signalling pathway [173]



Interaction of a Wnt ligand with a receptor complex including a Frizzled family member triggers downstream phosphorylation. B-catenin can accululate and enter the nucleus, interacting with coactivator proteins and leading to cell proliferation.

1.4.1.5. Microsatellite Instability

MSI is another well-described molecular aberration in EC with Lynch Syndrome accounting for 1.8-6% of cases [174] and another 15-31% of sporadic cancers also having high MSI [132, 175]. Lynch syndrome results from germline mutations in DNA MMR genes, MLH1, MSH2, MSH6 and PMS2 whereas in sporadic cases, methylation and transcriptional silencing of the MLH1 gene promoter occurs [176].

MSI may occur concurrently with PTEN mutations, particularly in NEEC [176, 177] and interestingly, MLH1 promoter methylation appears to be an early event during EC tumorigenesis [178]. There is no clear correlation between MSI and survival [133, 144].

1.4.1.6. Hormone Receptor Pathway Signalling

Although hormonal therapies have some demonstrated efficacy in EC, their use is limited by a lack of predictive biomarkers for response. ER and PR expression is observed in approximately 80% of sporadic ECs and is associated with low histological grade, good prognosis and response to hormonal therapies [179]. Expression of ERα, PR-A and PR-B are associated with earlier stage EC and lower grade tumours, and absence of ERα and PR-A is associated with increased risk of death [180]. Interestingly, a recent study of ERα gene amplification demonstrated similar frequencies in both EEC and NEEC [181]. Although this was hypothesis generating in terms of the role of hormonal signalling in EC, the methodology and reproducibility of the findings was questioned [182].

There is evidence for the interaction between hormone receptor signalling and other signalling pathways. Oestrogen exposure may contribute to PTEN loss in AEH and subsequent development into EC [183]. Thus, combination therapies targeting PI3K pathway and hormone pathway aberrations may be of interest.

1.4.2. Epigenetic Aberrations

1.4.2.1. Background

Epigenetics refers to the alteration in gene expression potential during development and cell proliferation, without any change in gene sequence [184]. DNA methylation is one of the most well documented mechanisms of epigenetic aberrations and refers to the reversible covalent modification of cytosines through the addition of a methyl group to the carbon-5 position of cytosine bases by one or more DNA methyltransferase (DNMT) enzymes, occurring predominantly in the context of cytosine-guanine dinucleotides (CpGs). Other epigenetic mechanisms include histone modification, and non-coding, small-interfering RNAs, all of which can affect transcript stability, DNA folding, nucleosome positioning, chromatin compaction and overall nuclear organisation, determining whether a gene is silenced or activated [185, 186].

The epigenetic silencing of a gene that is not normally suppressed or activation of a gene that is not normally activated is called an epimutation [187].

CpGs can occur randomly throughout the genome and tend to be methylated. They can also occur in clusters, known as CpG islands (CGIs), which range from 500 up to 2000 base pairs and tend to be unmethylated, unequally distributed and localized within and around the promoter regions of mammalian genes [188]. Hypermethylation in CGIs has proven to be a significant event in carcinogenesis [188, 189].

Global hypomethylation and/or CGI hypermethylation occurs in many cancers [190, 191] and was first described in colorectal cancer in 1983 [192]. Recent studies suggest that within each tumour, hundreds of genes may be silenced by DNA methylation, compared to 10-15 genes that may be silenced by genetic mutations [193]. DNA methylation in cancer also appears to be tissue and tumour-type specific [190], thus DNA methylation profiles may act as biomarkers for diagnosis, prognosis and response to treatment.

More recently, there is data to suggest that considerable DNA methylation occurs in CpG shores and in the gene body and can also be of significance in carcinogenesis [189, 194]. A representative diagram of CpG content is detailed in Figure 1.1 [195].

Figure 1.3: Representative diagram of CpG content [195]

Open Sea (white): isolated CpGs in the genome
Shore (blue): up to 2kb from CpG island
Shelf (brown): 2-4kb from CpG island
CpG Island (green): >200bp in length

1.4.2.2. Epigenetic Aberrations in Endometrial Cancer

Epigenetic alterations may be more frequent than genetic alterations in EC and there are documented epigenetic abnormalities in genes encoding tumour suppressor genes, oncogenes, apoptosis inhibitors, cell cycle regulators, steroid receptors, transcription factors and angiogenesis modulators [196].

A high frequency of methylation in tumours and cell lines has been reported for APC (-40%), CASP8 (-35%), CDH1 (-25%), ERα-promoter-C (94%), hMLH1 (41%), PR-promoter-B (75%), RASSF1A (33%) and THBS2 (62%) with a much lower frequency (-15%) of hypermethylation of p16INK4A (CDKN2A), p14ARF, p73 and PTEN [197-201]. Other genes silenced by hypermethylation that have been reported in EC include E-cadherin, CHFR, TGFBR3, p73, HOXAII, COMT, SPRY2, GPR54 and RSK4.

Most recently, TCGA has identified EC as having four subgroups as described previously [30], with one subgroup, the MSI/hypermutated group, displaying MLH1 promoter hypermethylation and a CpG island methylator phenotype (CIMP). CIMP refers to extensive DNA hypermethylation of a subset of CpG islands that normally remain unmethylated [202]. This subgroup consisted of both EEC and NEEC and did not contain TP53 somatic mutations nor extensive somatic CN alterations.

The subgroup of EC that was CN high had a normal-like DNA methylation profile and was found particularly in serous-type tumours and with TP53 somatic mutations.

The methylation analysis performed by TCGA using genome-wide bead chip technology provides a more extensive assessment of methylation changes with clinical correlation, compared to historical literature which tended to focus on a specific area of the genome through candidate gene methylation methods. These historical data however provide a framework for the current methylation data and a summary of some of these studies in EC is presented in Table 1.5. Of note, there is often limited correlation with clinical outcomes, such that the significance of the epigenetic change may only be partially understood. How best to use this data in the clinical setting has also not been established.

Table 1.5: Summary of studies of Epigenetic Aberrations in Endometrial Cancer

Author (year)	Gene (function)	Implication	Frequency	n
Ignatov [203]	APC (TSG)	Promoter hypermethylation found in AEH and early EC	77.4% I/II 24.2% III/IV 23.5% AEH 0 normal	86
Kang [204, 205] Pallares [206] Jo [207] Fiolka [208]	RASSF1A (TSG)	Common in solid tumours Negative regulator of RAS-MAPK Found in early and late stage EC Occurs with KRAS mtn in MSI+ EC Promoter hypermethylation in EC Associated with advanced stage, high grade, LN + and poor DFS Present in AEH and EC	33.3%	75
Velasco [209]	SPRY2 (TSG)	Antagonist of FGF & RAS-MAPK pathways Promoter hypermethylation correlated with RASSF1A and cell proliferation	53.4%	58
Dewdney [210]	RSK4 (TSG)	ERK substrate; inhibits FGFR2 & RAS/ERK pathways Frequently hypermethylated in EC	63%	158
Catasus [211]	TIMP3 (TSG)	Promoter hypermethylation ass'd with high stage EEC & extrauterine spread Also ass'd with MSI+ & MLH1 hypermethylation	25%	60

Wang [212]	CHFR (TSG)	Potential marker of microtubule-targeting drugs Promoter hypermethylation leads to weak CHFR expression & greater taxane sensitivity Hypomethylation of CHFR & high expression leads to resistance	Cell lines	6
Sanchez-Vega [213]	ZNF154 VHL (TSG) CASP8 (TSG)	Promoter hypermethylation of ZNF154 (little known on biological function) Hypomethylation in CASP8 and VHL.	ne	26
Dvorakova [214]	Multiple TSGs	Promoter hypermethylation of CDH13, WT1 and GATA5 genes	80%, 20%, 10%	59
Wu [215]	PAX2 (oncogene)	Methylated and silenced in normal endometrium; hypomethylated in malignant cells	75%	53
Teodoridis [216] Risinger [217] Whitcomb [200]	MSI	MSI+ ass'd with widespread gene mutations and CIMP MSI epigenetic repression of MMR genes (hMLH1, MGMT, WRN, BRCA1) more common than mutations of same genes MSI+ ass'd with promoter hypermethylation of hMLH1, PTEN, APC, SFRP1/4, RASSF1A MSI+ ass'd with HOXA11& THBS2 methylation	ne 38%	Ne 24
Kang [218]	GPR54	Receptor of KISS1; suppressor of cancer metastasis; high expression ass'd with increased survival Promoter hypermethylation reduces GPR54 expression	Cell lines only	ne
Katoh [219]	SFRP1	Negatively regulates Wnt signalling Promoter hypermethylation leads to its downregulation	ne	ne
[220]	HOXA10	Promoter hypermethylation in EC. No clinicopathologic correlation.	20%	19
Di Domenico [221]	Multiple	Promoter hypermethylation in ERα, PRα, hMLH1, CDKN2A, SFRP1/2/5 in EEC Hypomethylation of SFRP4.	38,38,84, 92,54,15, 53% (1 of 7 in 100%)	16
Zhang [222]	Multiple	Promoter hypermethylation p14, p16, ER, COX-2 & RASSF1A. CIMP+ (hypermethylation in >2 genes) in EC	40, 23, 29, 43 & 34% 46%	35
Powell [223]	Ribosomal DNA genes	Hypomethylation ass'd with shortened OS, esp in African-American women	74%	215

Mtn: mutation, ass'd: associated, ne: not evaluated, esp: especially

1.4.2.3. Methods of Methylation Analysis

There are a number of available techniques to characterise methylation profiles, the choice of which depends on the type of information required, whether global or locus-specific, genome-wide or candidate gene, quantitative or qualitative [224]. Methods including methylation-sensitive restriction enzyme digestion, sodium bisulfite (BIS) conversion and immunoprecipitation all provide different information, vary in time-intensity and require differing amounts of DNA.

Gene-specific methylation analysis includes both genome-wide and candidate gene approaches. Candidate gene approaches rely on the use of primers that detect methylated and unmethylated alleles of interest, whereas genome-wide analysis uses either array based or non-microarray based methods. Table 1.6 summarises some of the available techniques for methylation analysis.

Table 1.6: Techniques for Methylation Analysis

Technique	Genome-wide or candidate gene	Description	Pros	Cons
Methylation-sensitive restriction enzyme (MSRE) digestion [225]	Candidate gene	Analyses the methylation status of CpGs within the recognition sequences of specific REs	Little starting DNA required	Only CpGs within the RE sites are analysed False+ result if incomplete digestion
Methylation specific PCR (MSP) e.g. Methylight [226]	Candidate gene	PCR amplification of BIS converted DNA. Methylight uses fluorescence based PCR and specifically designed primers and probes for genes of interest	Highly sensitive Low cost Lower starting DNA conc Quantitative	Only CpG dinucleotides covered by the primers and probe will be assessed for methylation
BIS sequencing & restriction analysis (COBRA) [227]	Candidate gene	Provides data only for the specific restriction enzyme cutting sites		Time-consuming Large DNA starting conc
Methylated DNA immunoprecipitation (MeDIP) [228]	Genome wide	Antibodies target the methyl-CpG binding domain or the methylated cytosines, then hybridizes the	Does not require restriction digestion or BIS conversion	Requires large amounts of genomic DNA Lacks sensitivity in

		immunoprecipitated DNA to microarrays		regions with a low density of CpG sites
Restriction landmark genomic scanning (RLGS) [229]	Genome wide	DNA gel electrophoresis in combination with MSRE	Analyses thousands of loci at once.	Labour intensive Requires large DNA conc
Bead chip technology [230]	Genome wide	BIS conversion with fluorescence detection at individual CpG sites	Quantitative Can use fragmented DNA	

BIS: bisulfite, Conc: concentration

The TCGA methylation analysis is based on bisulfite-sequencing techniques on the Illumina HumanMethylation 450K BeadChip array, carried out on FF tumour specimens. Treatment of DNA with bisulfite converts cytosine to uracil and leaves methylated cytosine unaffected. Further analysis is then performed to extract this information regarding the methylation status of a segment of DNA, as described in Methods section 2.2.5.

The 450K array can assess methylation signatures in both FF and FFPE tissue, as evidenced by Bibikova et al [231] and Thirlwell et al [232] and analyses over 450 000 CpG sites on 12 samples in parallel. It was designed with the guidance of a consortium of methylation researchers and is comparable to whole-genome bisulfite sequencing data generated on a HiSeq2000 using NGS technology, including coverage of 99% RefSeq genes with multiple probes per gene, 96% of CpG islands from the UCSC (University of California, Santa Cruz) database, CpG island shores and additional content from whole genome bisulfite sequencing data and input from DNA methylation experts [233] [234].

1.5. Study Aims and Objectives

The current prognostic and predictive factors used to identify high-risk groups and guide therapeutic decisions in EC have been unchanged and unchallenged for many years, despite an increasing body of literature on the genetic and epigenetic aberrations that contribute to endometrial carcinogenesis. The recent TCGA data on EC molecular aberrations support its reclassification beyond the Type I and II subdivision, to better assist with risk stratification and treatment selection for EC patients, and to translate evidence on the molecular changes into the clinical trial setting with concurrent biomarker development. However, this data set does not interrogate the transition from normal endometrium to AEH to EC and was carried out in FF tissue which is not easily obtainable from patients. This latter point means large retrospective analyses and longitudinal analyses using historical tissue cannot be carried out unless banks of FF tissue exist.

The two main components of this thesis investigate the feasibility and clinical correlation in EC firstly with CTCs as a biomarker strategy and secondly with epigenetic analysis. The platforms used for both biomarker analyses are well-established. However, this work examines important novel aspects of these technologies that can be incorporated easily into clinical practice. These novel approaches are epigenetic analysis of methylation using EC FFPE tissue on the Illumina 450K array and detection of stathmin as a novel marker in CTC applied to EC. Thus, determining feasibility was one of the main objectives. Further to testing on FF tissue alone, as has been done by TCGA, profiling FFPE endometrial tissue would provide a far larger body of samples for study. In addition, identifying the aberrations which occur early in endometrial tumorigenesis in AEH, may assist with improving treatment strategies at an earlier stage of disease.

CTC enumeration and molecular profiling, if feasible, may not only provide one of the first predictive biomarkers to monitor patients' treatment in EC but also may allow for longitudinal assessment of molecular changes to assist with targeted therapy development.

Thus, through these two studies, a platform for further research in EC could be established into further practical, clinically applicable biomarker development and treatment strategies to improve patient outcomes.

They key objectives of both study components are outlined in Figures 1.1 and 1.2 below.

Figure 1.4. CTC Enumeration and Molecular Profiling in EC (EEC and NEEC)

- Feasibility of CTC enumeration and molecular profiling in EC on the Veridex CellSearch platform
- Correlation with clinical data and treatment
- Correlation of tumour tissue IHC with CTC profiling

Figure 1.5. Epigenetic and Genetic studies in EEC only

- Feasibility and reproducibility of epigenetic and CNV analysis from FFPE and FF EC tissue on the Illumina 450K array
- Analysis and comparison of methylation patterns in normal endometrium, AEH and EEC and correlation with clinical data
- Correlation of epigenetic and CNV data with TCGA and other published literature

CHAPTER 2: Methods

2.1. Development of the CTC Protocol: from approval to enumeration and profiling

2.1.1 Protocol Approval

The Research Ethics Committee (REC) submission, Site-Specific Information (SSI) form and Research and Development (R&D) submissions were uploaded and completed via the Integrated Research Approval System (IRAS) and were approved by the local REC (National Research Ethics Service Committee, NRES, London, Bloomsbury). There were two major and one minor amendment made to the protocol. The first involved an increase in the blood volume taken at sample collection from one vial (10ml) to two vials (20ml) to allow adequate blood volume for CTC enumeration and molecular analysis via the 4th channel. The second amendment added Sarah Cannon Research Institute UK (SCRI-UK) as an additional site for sample collection to improve patient recruitment. The third amendment increased the number of patients to be evaluated in the study, as some consented patients did not have evaluable samples or had limited follow-up. These documents are available for review via the NRES Committee London, Bloomsbury.

Approval for IHC on the corresponding archival tumour tissue blocks was approved by the University College London (UCL) Cancer Institute Biobank REC and was performed on tissue from the UCL hospital (UCLH) histopathology department only.

2.1.2. Patient Selection and Consent

Patients with locally advanced or metastatic EEC or NEEC were identified from the weekly UCLH Gynae-oncology MDT meetings, as well as from the weekly Gynae-oncology outpatient clinics. Patients older than 18 years who had a confirmed histopathological diagnosis of EC and signed informed consent were eligible for the study. Patients were excluded if there was a prior history of another cancer within the last 5 years, except non-melanoma skin cancer, or if there was a high risk that the patient would not comply with protocol requirements.

At SCRI-UK, patient referrals were reviewed weekly and potential patients identified.

The study was discussed with identified patients at the consultation with their treating team. If patients were interested in the study, they were approached by myself or the treating team, given the patient information sheet and this was discussed in detail. Once all questions were answered, patients gave their informed consent and blood collection was arranged thereafter. Patients with advanced stage EC were recruited at any timepoint of their treatment, whether pre or post-surgery, or pre, during or post chemotherapy or RT. Patients were followed over time and further tests taken for CTC enumeration every 3-6 months, or if there was a change in the clinical status, with up to 4 samples taken per patient.

2.1.3. Sample and Data Collection

Following consent, a peripheral blood sample of up to 20ml for CTC isolation and enumeration was collected in two CellSave Preservative tubes (Cat No 7900005), either by a phlebotomist or chemotherapy nurse. The samples were pseudoanonymised and transported from the UCLH outpatients department to the UCL Cancer Institute, where they were received by a member of staff that was GCP and GCLP trained. Each specimen was processed within 72 hours of being received, as per the Veridex and UCL Cancer Institute lab protocols.

Clinical data were collected on these patients from the UCLH patient information system including demographic, clinicopathologic information, details of treatment, disease recurrence and survival. These data were also pseudo-anonymised and stored on a password-controlled UCL Cancer Institute computer. The last time-point of data and CTC collection was February 2014.

If they had tissue collection, either biopsy or surgery, performed at UCLH, this was retrieved from the pathology department archive to be sectioned for EpCAM, Stathmin and p70S6kinase IHC.

2.1.4. Sample Preparation

From the peripheral blood draw, 7.5ml of blood was combined with 6.5ml of Dilution Buffer (CellSearch CTC Kit, Cat No 7900001) in a 15ml conical tube, centrifuged at 800xg for 10 minutes at room temperature and processed on the CellTracks AutoPrep system (Cat No 9541) within one hour of sample preparation.

A Veridex control sample (CellSearch CTC Control Kit, Cat No 7900003) was prepared for each kit run. The control sample was spiked with two different cell lines at a low and high concentration. This demonstrated at each run that the automated system was detecting CTC counts correctly, based on what was detected in the control specimen. This was stored at 4°C and warmed to room temperature 20 minutes prior to the analysis.

After inverting 5 times and vortexing for 5 seconds, the control sample was added to a 15ml conical tube and placed on the CellTracks AutoPrep system.

The Veridex platform also has a 4^{th} channel (FITC) that allows molecular analysis to be performed. Stathmin antibody (Biorbyt catalogue number orb 104103) was used in the 4^{th} channel in this study, after having been validated for use in cell lines and CTCs at the UCL Cancer Institute, as described in section 2.1.7. For the Stathmin solution, 30µl of antibody mix was diluted with 570microl PBS in a 1:20 dilution to 50µg/ml. 300µl was required dead volume

for each kit run and 150µl was required per specimen analysis. This total volume was then placed in the 4th channel.

Once on the CellTracks Autoprep system, the plasma and buffer layer were aspirated from the blood sample. Ferrofluids containing nanoparticles with a magnetic core surrounded by a polymeric layer coated with antibodies to EpCAM were then added and incubated. Magnetic separation would then occur and the remaining unbound cells and plasma were aspirated out. Staining reagents and permeabilization buffer were then added to fluorescence label the immunomagnetically labelled cells and identify the presence of CTCs.
The fluorescent reagents that were added were cytokeratins 8, 18, 19 (CK-PE), 4'6 –diamidino-2-phenylindole (DAPI) and an antibody to CD45 conjugated to allophycocyanim (CD45-APC). CK-PE is specific for the intracellular protein cytokeratin, characteristic of epithelial cells, while DAPI stains the cell nucleus and CD45-APC is specific for leukocytes.
Cells were then resuspended in the MagNest cell presentation fixture. A strong magnetic field was generated by the MagNest and caused the magnetically-labelled target cells to move to the surface of the cartridge, distribute uniformly over the analysis surface of the cartridge and orient for analysis at a single focal depth. The cartridge containing stained CTCs was then placed onto the CellTracks Analyzer II (Cat No 9555 RUO), a four-colour semi-automated fluorescence microscope, for scanning.
This process is illustrated in Figure 2.1 below.

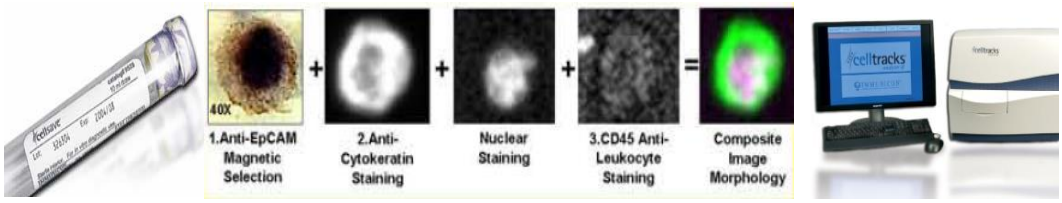


Figure 2.1: The Veridex CellSearch CTC Detection System

2.1.5. CTC Analysis and Enumeration

On the CellTracks Analyzer II, image frames covering the entire surface of the cartridge were captured. Once the cartridge had been scanned, the system displayed tumour cell candidates that were positive for cytokeratin and DAPI, and could then be reviewed by the operator.

A CTC was defined as positive for CK-PE and DAPI, negative for CD45-APC and having the correct morphology to be characterised as a tumour cell. The CellTracks Analyzer II presented the images with overlays of CK-PE and DAPI signals, to show whether the nuclear and cytokeratin staining were overall consistent with a tumour cell. The object in the CK-PE filter

channel needed to be an intact cell, at least 4 microns in diameter and generally round or oval, though could be polygonal or elongated. The nuclear area should be smaller than the cytoplasmic area and more that 50% of the nucleus needed to be visibly surrounded by the cytoplasm.

An image that was very bright in the CK-PE channel could result in spectral spillover and create a visible cytoplasmic image in the CD45-APC channel. This could still be classified as a tumour cell and differentiated from leukocytes that would be positive for CD45-APC and DAPI but negative for CK-PE.

Squamous cells and other artefacts were also often present in samples. Squamous cells may have been due to contamination from the needle stick or from handling the specimen. They were identified by their low nuclear to cytoplasmic ratio and their large, polygonal appearance. Artefacts usually appeared in all channels with the same shape.

All samples were reviewed by two trained laboratory staff, as well as by myself. Examples of CTCs and other artefacts are illustrated in Figure 2.2.

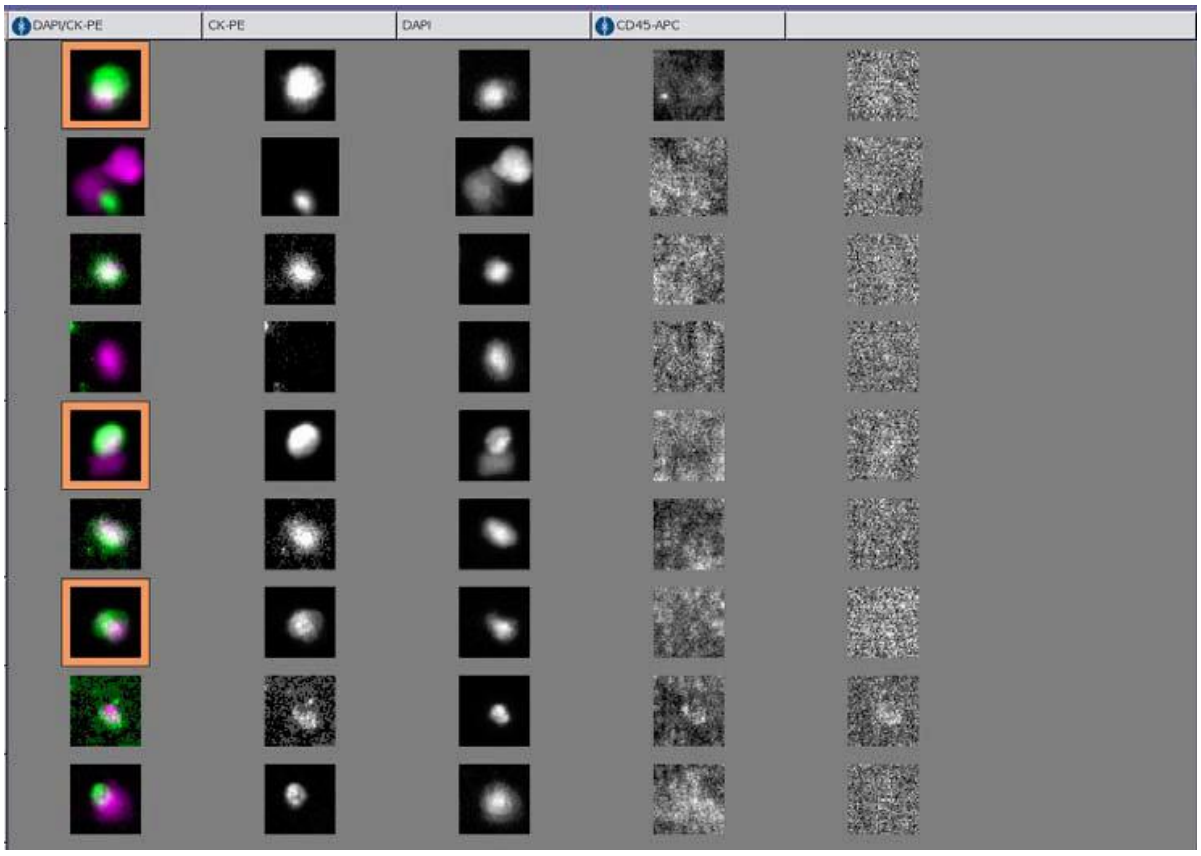


Figure 2.2. A screenshot from the CellTracks Analyzer demonstrating the Criteria for CTC Analysis

The first column is the CK-PE and DAPI overlay, the second column is CK-PE staining only, the third column is DAPI staining only and the 4th column is CD45-APC. The fifth column was not used in this diagram but when used, would show the 4th channel staining for molecular characterisation.

2.1.6. FFPE Tissue Processing and IHC assessment

For patients who had a biopsy or surgery performed at UCLH, EpCAM (1/4 dilution; mouse anti-human, RTU-ESA; Novocastra, Newcastle, UK), stathmin (1/30 dilution; 3352 CellSignalling, Hertfordshire, UK) and p70S6kinase (1/25 or 1/50 dilution; Clone E135, Genetex, Cat No GTX 61056, California, USA) IHC was performed on the available FFPE tumour blocks, obtained from the UCL Histopathology Department.

Each archival tumour block was sectioned at 3 micron thickness onto charged slides, drained and baked at 60°C for 60 minutes. Unstained slides were labelled and placed for IHC staining of EpCAM, stathmin or p70S6kinase primary antibodies on the Leica Bond III autostainer platform, using the Bond Polymer Refine Detection Kit (Cat No DS9800).

The general methodology is summarised below while the specific protocols for each primary antibody are outlined in Table 2.1.

On-board dewax and rehydration were performed followed by epitope retrieval by either heat-induced (HIER with ER2, EDTA-based, Cat No AR9961) or enzymatic digestion (with Enzyme 1 diluted 1/2000, Bond Enzyme Pretreatment Kit, Cat No AR9551) to break the formalin-protein crosslinks that occur in FFPE tissues. Peroxide block provided in the Bond Polymer Refine Detection Kit (Cat No DS9800) was then applied to neutralise the endogenous peroxidase in human tissue.

The primary antibody was applied for 15 or 30 minutes at room temperature, as detailed in Table 2.1, followed by a rabbit anti-mouse post primary antibody and goat anti-rabbit horseradish peroxidase conjugated polymer (components of DS9800).

Antigens were visualised with the substrate chromogen 3,3'-Diaminobenzidine tetrahydrochloride (DAB, Cat No 9800) which formed an insoluble brown precipitate at the antigen site on reaction with horseradish peroxidase conjugated to the polymer. Bond DAB Enhancer (Cat No AR9432) was then added to darken the DAB precipitate.

Slides were counterstained with haematoxylin (component of DS9800) to allow the visualization of cell nuclei. All washes between reagents were with either Bond Wash solution 1x (AR9590) or deionised water. The staining was run over a 2.5hr period.

In addition to the standard protocol, 10% goat serum was used after stathmin antibody application for those slides sectioned from surgical resection samples. This blocked non-specific antibody binding that occurred in these samples, though was not required for the biopsy samples. Further details of stathmin antibody validation are detailed in section 2.1.7.

Different protocols were also used for biopsy and surgical resection tissue for p70S6 kinase antibody to optimise staining. A 1/50 dilution was used for the biopsy samples and a 1/25

dilution, with longer retrieval time and antibody incubation time, was used for the control and surgical resection specimens.

Table 2.1: Staining protocols for EpCAM, Stathmin and p70S6kinase

Step	EpCAM		Stathmin B/S		p70S6 k B/S	
Part 1	Time (min)	Temp (°C)	Time (min)	Temp (°C)	Time (min)	Temp (°C)
Bake	60	60	60	60	60	60
Bond dewax solution	0.5 x1 0 x1 0 x1	72 72 Room	0.5 x1 0 x1 0 x1	72 72 Room	0.5 x1 0 x1 0 x1	72 72 Room
Alcohol	0 x3	Room	0 x3	Room	0 x3	Room
BWS	0 x2 5 x1	Room	0 x2 5 x1	Room	0 x2 5 x1	Room
*ER2 ER2 ER2 BWS BWS OR *Enzyme 1 BWS	- 10 0 x3	- 37 Room	0 x2 20 12 0 x3 3 - 	Room 100 Room 35 Room - 	0 x2 **30/40** 12 0 x3 3 - 	Room 100 Room 35 Room -
Peroxide block	5	Room	5	Room	5	Room
BWS	0 x3	Room	0 x3	Room	0 x3	Room
1° Ab	15	Room	30	Room	**30/60**	Room
BWS	2 x3	Room	0 x3	Room	0 x3	Room
10% goat serum	-	-	**20** **(S only)**	**Room** **(S only)**	-	-
BWS	-	-	-	-	0 x3	Room
Post-1° Ab	8	Room	20	Room	20	Room
BWS	2 x3	Room	2 x3	Room	2 x3	Room
Polymer	8	Room	20	Room	20	Room
BWS	2 x2	Room	2 x2	Room	2 x2	Room
DW	0	Room	0	Room	0	Room
Mixed DAB refine	0 & 10	Room	0 & 10	Room	0 & 10	Room
DW	0 x3	Room	0 x3	Room	0 x3	Room
Bond DAB enhancer	5	room	5	Room	5	Room
BWS	0 x3	Room	0x3	Room	0x3	Room
Haematoxylin	1	Room	1	Room	1	Room
DW	0	Room	0	Room	0	Room
BWS	0	Room	0	Room	0	Room
DW	0	Room	0	Room	0	Room

Time 0 indicates a wash was performed, X3 indicates the wash occurred 3 times
BS2: bond solution 2, BWS: bond wash solution, DW: deionised water,
B: biopsy specimen, S: surgical resection specimen, k: kinase
*This row indicates the epitope retrieval method; either with the ER2 HIER technique or enzymatic epitope retrieval technique with Enzyme 1.
Bolded entries refer to where there was a difference between protocols for biopsy and surgical resection specimens.

Once the staining procedure was complete, slides were rinsed in tap water and dehydrated through graded alcohols as follows for permanent mounting:

- Industrial methylated spirit (IMS; Barrettine Industrial Ltd, Bristol, UK) 100% x3 for 2 min
- Absolute ethanol x1 for 2 min (Analar NORMAPUR 20821.321 VWR, BDH)
- Xylene x4 for 2 minutes (Analar NORMAPUR 28975.325 VWR, BDH)

Then the xylene-based ClearVue mountant (Thermo Fisher, 23-425-401) was used to attach the coverslip and the slides were baked at 60°C for 20 minutes.

The EpCAM, stathmin and p70S6 kinase IHC analyses were interpreted with a consultant histopathologist, specialised in gynae-oncology, who was blinded as to the CTC status of each patient.

The Tumour Immunostaining Score (TIS) was used to evaluate EpCAM expression, similar to the Allred score in the evaluation of oestrogen receptor positivity [97, 163, 235]. The TIS is a product of a proportion score and an intensity score. The proportion score describes the estimated fraction of positively stained tumour cells (0: none; 1: <10%; 2: 10-50%; 3: 51-80%; 4: >80%). The intensity score represents the estimated staining intensity as compared with a normal FFPE control, in this case appendix (0: no staining; 1: weak; 2: moderate; 3: strong). The TIS ranges from 0-12 and is classified into 4 subgroups as follows: no expression (TIS 0), weak expression (TIS 1-4), moderate expression (TIS 6 and 8), intense expression (TIS 9 and 12).

The reported TIS score for stathmin overexpression was different to that used for EpCAM IHC [163]. It is also defined by a semiquantitative grading system that incorporates staining intensity (score 0-3), as per the EpCAM scoring system, and area of tumour with positive staining (0: no staining, 1: <10%, 2: 10-50%, 3: >50% of tumour cells). The staining index is then calculated as the product of staining intensity and staining area, with a range 0-9 [163]. Some consider the upper quartile for the data set as positive [138, 163], while others define moderate/high expression as an index ≥4/9 and absent/minimal expression as an index <4 [138, 235]. This quantification method is well documented [163, 235] and was used in this study.

Similarly, p70S6 kinase staining scores vary within the literature depending on the specific antibody and tissue type. For consistency and based on reported data, the same staining score was used as for EpCAM, based on the product of staining intensity (0-3) multiplied by the proportion of immunoreactive cells in the areas of interest (1-4), with a score of ≥6 representing positive staining [236].

2.1.7. Validation of Stathmin Antibody for Detection on CTCs and FFPE IHC

2.1.7.1. Stathmin Antibody Validation on CTCs

Stathmin antibody (Biorbyt catalogue number orb10403) validation for detection on CTCs was performed between February to May 2012 in UCL Cancer Institute ECMC GCLP Facility laboratories by Helen Lowe and Natalie Griffin [237]. Firstly, immunofluorescence work was performed on cell lines to confirm their stathmin expression. Secondly, the suitable cell lines were spiked into whole blood from healthy donors along with different concentrations of the stathmin antibody for analysis on the CellSearch system.

For this validation study, 3 cell lines, Hela (epithelial cervical cancer), A549 (epithelial lung cancer) and SKOV3 (epithelial ovarian adenocarcinoma), all known to express stathmin [238-240], were analysed to confirm the binding of stathmin antibody. SKOV3 was found to be a higher expressing cell line for stathmin compared to A549, while the Hela cells autofluoresced in the absence of stathmin antibody. Thus, the A549 and SKOV3 cell lines alone were taken forward for analysis on the CellSearch system. Stathmin antibody was used at a concentration of 500µg/ml and demonstrated binding to the cell lines, with no batch variation.

For analysis on the CellSearch system, whole blood was collected from healthy donors and spiked with the cell lines, either Hela or A549. Blank samples were also assessed to rule out any false positives. A control CTC sample was used with all experiment on the CellSearch system to ensure that it was performing correctly. The only parameters that could be altered on the CellSearch Analyser II were the exposure time for the 4^{th} channel and the concentration of stathmin antibody.

Nine experiments were performed to determine the optimal stathmin concentration and exposure time and this was set at 1:20 dilution (50µg/ml) and 0.8 seconds respectively.

The percentage of CTCs positive for stathmin expression in the cell lines was between 0.2-1%, consistent across 2 experiments and thus deemed suitable for clinical use.

2.1.7.2. Stathmin antibody validation for FFPE IHC

Use of stathmin antibody (3352 CellSignaling) for FFPE IHC was validated by Jennifer Paterson and Gabrielle Elshtein at UCL-Advanced Diagnostics.

The Biorbyt stathmin antibody used for CTCs did not demonstrate overexpression on FFPE tissue, therefore the CellSignalling antibody was used based on published evidence off its utility for FFPE IHC [163]. Tonsil was used as a control tissue due to documentation of positivity and experience with staining and analysis [241]. Different dilution and retrieval times and techniques were assessed to determine optimal staining, first in tonsillar then in EC tissue. Based on local assessment, biopsy specimens were best analysed with protocol 30-20-

20, retrieval H2(20) and dilution 1:30. The same parameters were used for the surgical resection specimens but with the addition of 10% goat serum to block non-specific antibody binding, as previously detailed in Table 2.1.

2.2. Endometrial Sample Collection and DNA Modification

2.2.1. Protocol Approval

The protocol and standard operating procedures for this study were approved by the UCL Cancer Institute Biobank REC. Standard operating procedures covered both FFPE and FF tissue collection, transportation and use. Tissue processing only commenced once the study was approved. This approval allowed archival FFPE to be used for research purposes without individual patient consent, provided the tissue was collected from the living and was surplus to diagnostic requirements.

For FF tissue, a consent form approved by the UCL Biobank REC was used by the Gynae-oncology medical and surgical team to allow the collection of tissue, surplus to diagnostic requirements, for research under the ethical approval for the gynaecological biobank at UCL. This was approved by the REC for the UCL/UCLH Biobank for Studying Health and Disease, based at Pathology-Rockefeller Building and UCL Cancer Institute.

2.2.2. Patient Selection and Consent

For collection of FFPE tissue, a retrospective search of the UCLH archival histopathology database was performed by the histopathology team for cases of grade 1 and grade 3 EEC, surgically resected between January 2009 and December 2013. Once identified, the slides were retrieved from storage and reviewed with a consultant Gynae-oncology histopathologist to select the appropriate blocks for DNA extraction.

FF samples were obtained through collaboration with the gynaecological surgical team at UCLH. No patient underwent surgery or biopsies specifically for the purpose of this study and only tissue surplus to diagnostic requirements was used. The patients were identified during review in the gynaecological surgical clinics, MDT and via review of the weekly surgical list. Patients were approached by a member of the clinical or research team during their pre-admission clinic or on the morning of surgery to discuss consent for collection of normal endometrial and cancer tissue, if present in excess to diagnostic requirements. Donation of tissue samples under the Biobank approval was discussed and they received the UCL/UCLH Biobank for Studying Health and Disease written patient information sheet with consent form attached. Written, informed consent was then obtained after the patient had read and understood the information sheet and all questions were answered. An original of the consent

was given to the patient and a copy scanned into the patient's secure electronic medical notes. A paper original was filed in the patient's medical notes and a paper copy was securely stored in a locked cabinet on the ground floor of the UCL Cancer Institute.

2.2.3. Sample and Data Collection

For FFPE samples, the most recent specimens of each tissue type surplus to diagnostic requirements were identified, selecting specimens that contained histologically normal tissue as well as EC and where possible, AEH. The haematoxylin and eosin (H&E) stained slides of each case were reviewed with Consultant Histopathologist, Dr Rupali Arora, to identify for which cases there was adequate amounts of normal endometrium, AEH and grade 1 or 3 EEC within the specimen. At this time, representative areas for sampling within the tumour block were marked out on the corresponding H&E slides. Each case was pseudoanonymised and allocated a number based on the order of identification. All subsequent analysis used only the study-specific number and whether normal, atypical or cancer tissue was present.
For FF samples, the tissue sample was collected at the time of the surgical procedure by the research nurse or fellow and transported to the UCL pathology department at the Rockefeller Building. Tissue was reviewed by a senior histopathologist for diagnostic purposes, and where possible, surplus tissue, both normal and cancer from the same specimen were identified and dissected for collection. This tissue was snap frozen in cryovials within a Dewar flask of liquid nitrogen using appropriate precautions. The frozen tissue was taken to the ground floor of the UCL Cancer Institute (Room G12 allocated to the Institute for Women's Health) and stored in drawer 9 of the designated -80 °C freezer.

The FF specimens were logged and pseudo-anonymised in LabVantage and appropriately labelled to ensure ease of tracking and traceability. LabVantage is a password-secured, electronic database that records sample identification details as well as the precise location of each individual sample to include room number, shelf, rack, box number and co-ordinates of the boxed samples where applicable. The sample log was updated each time a sample was added, moved, transferred or returned to the storage freezer and recorded the date, name of receiving researcher, name of lead researcher, description of tissue, details of consent and location. A unique identifier was assigned for use in the database and a number assigned in order of recruitment. This identifier was held separately on a password secured computer by the two staff members who updated LabVantage.

All tissue analyses fulfilled requirements of the Human Tissue Act 2004 and the Human Tissue Authority and was conducted in accordance with the ethical principles in the Declaration of Helsinki, Good Clinical Practice and applicable regulatory requirements.

Clinical data from the UCLH patient information system was collected for the patients who had FFPE or FF specimens processed on the arrays including demographics, clinicopathologic information, details of recurrence and survival. This data was pseudoanonymised and stored on a password secured database at UCL Cancer Institute, accessible only to relevant members of the research team at UCL/UCLH.

2.2.4. DNA Extraction and Modification for the Illumina 450K array

2.2.4.1. Formalin-Fixed Paraffin-Embedded Tissue

2.2.4.1.1. DNA Extraction

Each block, corresponding to the selected slide for each histopathology sub-type, was sectioned at 6 micron thickness onto between 20 and 40 plain glass slides, depending on the amount of tissue available per slide, and were left to air dry.

Slides were de-waxed using a standard protocol, whereby a series of washes were performed, first in Histoclear (National Diagnostics UK, AGTC Bioproducts, Yorkshire, HU13 9LX, UK) for 10 minutes, repeated three times; then in industrial methylated spirits (IMS) 100% for 10 minutes, repeated twice; then in IMS 70% for 5 minutes and finally in double-distilled (dd) water for 5 minutes. Slides were then left to air dry at room temperature.

Once the slides were dry, the marked sections were macro-dissected from the slide using a sterile 21-gauge needle with buffer ATL on its tip and placed in a 1.5ml microcentrifuge tube with 180µl buffer ATL. After the addition of 20-40µl of proteinase K, the sample was mixed by vortexing and incubated on a spinning cycle overnight to digest.

The process of DNA extraction using the QIAmp DNA mini-kit (Qiagen GmbH D-40724 Hilden, Cat No 51304) was then performed using the manufacturer's instructions and is summarised as follows:

i) The 1.5ml tube was centrifuged to remove drops from inside the lid

ii) 200µl Buffer AL was added, pulse vortexed for 15 seconds and incubated at 70°C for 10minutes then briefly centrifuged

iii) 200µl ethanol (96-100%) was added, pulse vortexed for 15 seconds and briefly centrifuged

iv) This mixture was applied to the QIAmp mini spin column, placed in a 2ml collection tube, without wetting the rim. The cap was closed and it was centrifuged at 6000xg for 1 minute, the placed in a clean 2ml collection tube, discarding the filtrate

v) 500µl Buffer AW1 was added and centrifuged at 6000xg for 1 minute, then placed in a clean 2ml collection tube, discarding the filtrate

vi) 500µl Buffer AW2 was added and centrifuged at 20 000xg for 3 minutes, then placed in a new 2ml collection tube, discarding the filtrate.

vii) The tube was centrifuged at 20 000xg for 1 minute, then the mini spin column was placed in a clean 1.5ml microcentrifuge tube, discarding the filtrate

viii) To elute the DNA, 100µl buffer AE was added and incubated at room temperature for at least 5 minutes, then centrifuged at 6000xg for 1 minute

2.2.4.1.2. DNA Concentration Analysis

The two most commonly utilised methods for quantification of DNA are based on spectrophotometric and fluorescence technologies and both methods were used in this analysis.

The concentration of the DNA was checked using the NanoDrop spectrometer (Thermo Scientific, Wilmington, DE) alone for the first 10 specimens, then both the NanoDrop spectrometer and Qubit 2.0 Fluorometer, when this was introduced to the lab.

NanoDrop is a spectrophotometric instrument with a patented sample retention system that uses fibre optic technology and surface tension to hold a 1µl sample in place between two optical surfaces that define the pathlength [242]. During each measurement, the sample is evaluated at a 1mm and 0.2mm path to measure DNA concentration across a wide range from 2ng/µl to 3700ng/µl.

The quantification of DNA by NanoDrop relies on the fact that its peak light absorption is 260nm. Protein absorbs at 280nm, thus to determine the purity of the DNA, a ratio of the absorbance values is calculated. The 260/280nm ratio should be 1.7-1.8 for pure DNA, whereas lower values may reflect protein contamination. Higher ratios usually indicate the presence of RNA, which has a 260/280 ratio of approximately 2.2-2.3. The 260/230nm ratio is also used as peptide bonds absorb at 228nm and may be a more reliable indicator of proteins or peptides present in nucleic acid samples. The 260/230 ratio should be >1.5. Buffer salts can also contribute to absorbance readings below 260nm, and it is important to blank the spectrophotometer prior to use in the same buffer in which your DNA sample is diluted.

Fluorescence methods of DNA quantification on the other hand, rely on the use of sensitive dyes that fluoresce in proportion to the component of interest, whether that is protein, RNA or DNA.

The Qubit 2.0 fluorometer (Life Technologies, Paisley, UK; Cat No Q32866) utilises fluorometric technology and Molecular Probes dyes to quantify biomolecules of interest, whether DNA, RNA or protein. The fluorescent dye emits signals only when bound to specific target molecules, even at low concentrations and have extremely low fluorescence when unbound. The Qubit DNA dye binds to DNA by intercalation between the bases, where it assumes a more rigid shape and becomes fluorescent. It binds within seconds and reaches equilibrium in less than 2 minutes. At a specific dye concentration, the amount of fluorescence is proportional to the concentration of DNA in the sample, and the Qubit fluorometer uses DNA standards of known concentration to convert the fluorescence signal of a sample into a DNA concentration.

The protocol of sample assessment on the Qubit Fluorometer is as follows:

i) The number of 0.5ml Qubit assay tubes (Life Technologies, Paisley, UK; Cat No Q32856) needed for the samples and two standards were prepared

ii) The Qubit working solution was made by diluting the Qubit dsDNA reagent 1:200 in Qubit dsDNA buffer

iii) For preparation of the two standards, 190µl of Qubit working solution was added to 10µl of each Qubit standard and mixed by vortexing for 2-3 seconds

iv) For sample preparation, 199µl of Qubit working solution was added to 1µl of each sample, to make a total volume of 200µl and mixed by vortexing for 2-3 seconds.

v) All the tubes were incubated for 2 minutes then run on the Qubit fluorometer, following the screen instructions.

vi) Using the Calculate Stock Concentration option, the concentration given in µl/ml was converted to ng/µl.

The Qubit dsDNA broad range (BR) assay kit (Life Technologies, Paisley, UK; Cat No Q32853) detects concentrations between 100pg/µl and 1000ng/µl, while the Qubit dsDNA high sensitivity (HS) assay kit (Life Technologies, Paisley, UK; Cat No Q32854) detects concentrations between 10pg/µl and 100ng/µl. The concentration measured on NanoDrop was used as an estimate for the appropriate Qubit assay kit. If the concentration could not be detected on initial assessment and was out of range, the alternate kit was used.
Once the concentration was measured, specimens were stored at -20°C for further modification and analysis.

2.2.4.1.3. FFPE Quality Check

After the first run on Illumina HumanMethylation 450K array was performed, a further quality control step was introduced to assess the integrity of the DNA that was extracted. This step was designed to reduce the likelihood of performing an array on DNA of poor quality. The Applied Biosystems 7300 Real-Time (RT) PCR System was used for this analysis in 7300 mode. This step was performed using the Illumina Infinium HD FFPE QC Kit (Cat number WG-321-1001), as per the manufacturer's instructions and summarised below:

i) 1µl of extracted DNA was diluted to concentration of 1ng/µl using dd H_20

ii) SybrGreen, QC Primer Reagent and QC Template (QCT) reagent were thawed to room temperature

iii) 1µl QC Template Reagent was added to 990µl ddH_20, vortexed and centrifuged briefly

iv) A mastermix for 20µl reaction volumes was prepared including 10µl SybrGreen, 2µl QC Primer Reagent and 4µl ddH20 for each DNA sample to be run in triplicate as well as for the QCT reagent and dd H_20 controls and for approximately 33% overage. It was mixed by inverting 10 times and tapped onto the bench to collect droplets. 16µl of this mastermix was pipetted into each well of a 96-well PCR plate

v) To the designated triplicate wells, 4µl of the diluted QCT reagent was added

vi) To the further designated wells, 4µl of dd H_20 in triplicate was added

vii) 4µl of genomic DNA in triplicate into the other designated wells was added

viii) The plate was sealed and briefly centrifuged at 280xg

ix) The plate was placed in the 7300 RT-PCR machine and programmed for the following thermal profile: 95°C for 10 minutes, 95°C for 30 seconds x40, 57°C for 30 seconds x 40, 72°C for 30 seconds x40, followed by a dissociation stage (where dsDNA breaks into ssDNA)

For data analysis:

i) The H_20 wells were checked for amplification as this should be 0.

ii) Other samples were checked for good amplification across all triplicates and that the threshold/quantification cycle (Cq) values, generated by the 7300 RT-PCR, were consistent within half a unit

iii) The average Cq values were obtained for the triplicate wells for each FFPE and QCT sample

iv) The average Cq value for the QCT sample was subtracted from the average Cq value for each sample to compute the Delta Cq value for each sample.

2.2.4.1.4. DNA Ligation

Using the REPLI-g FFPE kit (Qiagen GmbH D-40724 Hilden, Cat No 150245), ligation of the FFPE-extracted DNA was performed. 2.5µg DNA in a total volume of 10µl (concentration 250ng/µl) was required for this step and samples were either concentrated or diluted with dd H_20 to make up the appropriate concentration. If the DNA concentration was too low to provide 2.5µg DNA, 1.5-2µg was used.

A mastermix was prepared on ice from contents of the REPLI-g FFPE kit, including 8µl FFPE buffer, 1µl of ligation enzyme and 1 µl of FFPE enzyme per sample to be processed. This was added to the 10µl of DNA, vortexed and centrifuged. Using a thermocycler, the specimen was then incubated at 24°C for 90 minutes for ligation to occur, followed by a denaturing step at 95°C for 5 minutes, then cooled to 4°C.

2.2.4.1.5. Bisulfite Conversion

The EZ DNA Methylation kit (Zymo Research Cop, Orange, CA 92867, USA, Cat No D5001) was used for bisulfite conversion and is summarised as follows:

i) The CT conversion reagent was prepared by adding 750µl autoclaved H_20 and 210µl M-dilution buffer to the CT tube, mixing at room temperature for 10 minutes with frequent vortexing. This prepared reagent could be store for up to 1 month at -20°C but needed to be warmed to 37°C and vortexed prior to use

ii) M-Wash buffer was prepared by adding 96ml of 100% ethanol to the 24ml M-Wash buffer concentrate

iii) In a 12-PCR tube row, 20µl of ligated DNA was added to 5µl of M-dilution buffer and 25µl H_20 to make a total volume of 50µl, and incubated at 37°C for 15 minutes to denature the DNA.

iv) 100µl of CT conversion reagent was added to each sample and mixed with minimal exposure to light

v) Bisulfite conversion was performed by incubating the samples in the dark at 50°C for 16 hours. Every 60 minutes, the DNA was denatured through thermocycling at

95°C for 30 seconds before returning to 50°C for further incubation. The hourly denaturing step improved bisulfite conversion efficiency. Once completed, the sample was returned to 0-4°C for 10 minutes or longer if required.

vi) 400µl of M-Binding Buffer was then added to the sample in a Zymo-Spin IC column, which was then inverted several times to mix, placed in a collection tube and centrifuged at full speed ≥10 000xg for 30 seconds, discarding the flow-through

vii) 100µl of M-Wash Buffer was added to the column and centrifuged at full speed for 30 seconds, discarding the flow-through.

viii) 200µl of M-Desulphonation Buffer was added to the column and left to stand at room temperature (20-30°C) for 15-20 minutes, then centrifuged at full speed for 30 seconds.

ix) 200µl of M-Wash Buffer was added to the column, centrifuged at full speed for 30 seconds, followed by a further addition of 200µl of M-Wash Buffer and centrifuged for an additional 30 seconds.

x) The column was then placed in a 1.5ml microcentrifuge tube and 10µl of M-Elution Buffer was added directly to the column matrix. The sample was incubated for 5 minutes then centrifuged for 30 seconds at full speed to elute the DNA.

The DNA could then be stored at -20°C prior to analysis.

2.2.4.1.6. Quality Control of Bisulfite Conversion

This quality control step was performed to check the percentage success of the bisulfite conversion, through the use of PCR reaction and primer sets specific for bisulfite converted and unconverted DNA. The Applied Biosystems 7300 Real-Time PCR System was used for this analysis in 7300 mode.

From each specimen, 1µl of bisulfite converted DNA was added to 9µl autoclaved water. At the same time, a mastermix was prepared containing SybrGreen 6.25µl, primers 0.625µl, autoclaved H_20 4.375µl, calculated for each specimen to be run in triplicate. The primers used were actin positive and actin negative (Sigma-Aldrich, Dorset, England; Batch No HA05224892-5) and a separate mastermix was prepared for each. The actin positive primer would bind to bisulfite converted DNA and the actin negative primer would bind unconverted DNA, such that comparison of the two, would generate the percentage bisulfite conversion for each sample.

Both the actin positive and negative primers consisted of forward and reverse oligo sequences (5'-3') as follows:

Actin BisConv Pos Forward: TGGTGATGGAGGAGGTTTAGTAAGT

Actin BisConv Pos Reverse: AACCAATAAAACCTACTCCTCCCTTAA

Actin BisConv Neg Forward: TGGTGATGGAGGAGGCTCAGCAAGT

Actin BisConv Neg Reverse: AGCCAATGGGACCTGCTCCTCCCTTGA

A water sample and template control were also run in triplicate as a comparator for the specimens.

11.25µl of mastermix was pipetted into the designated wells of a 96-well PCR plate (an example is illustrated in Table 2.2). 1.25µl either of the water, control or sample were added, placing the pipette to the bottom of the well to ensure adequate mixing and to avoid bubble formation.

An optical seal was then placed on top of the PCR plate without touching the sides. The plate was flat vortexed to mix for 10 seconds, then centrifuged for 30-60 seconds.

The plate was then analysed on the RT-PCR machine as follows:

- Stage 1: 95°C for 10 min
- Stage 2: 95°C for 15sec x40, 60°C for 1min
- Dissociation stage: 95°C for 15sec, 60°C for 1min, 95°C for 15sec, 60°C for 1min
- hold at 4°C

Table 2.2: Example of a plate prepared for Bisulfite Conversion QC

	Wells 1-3	Wells 4-6	Wells 7-9	Wells 10-12
A	S1 A+	S1 A-	S2 A+	S2 A-
B	S3 A+	S3 A-	S4 A+	S4 A-
C	S5 A+	S5 A-	S6 A+	S6 A-
D	S7 A+	S7 A-	S8 A+	S8 A-
E	S9 A+	S9 A-	S10 A+	S10 A-
F	S11 A+	S11 A-	S12 A+	S12 A-
G	S13 A+	S13 A-	S14 A+	S14 A-
H	H_20 A+	H_20 A-	Control A+	Control A-

S: sample, A: actin

Data was then analysed both visually and from quantitative data that was extracted and analysed in Microsoft Excel. If bisulfite conversion was successful, the samples were then ready to be analysed on the Illumina HumanMethylation 450K BeadChip array.

2.2.4.2. Fresh Frozen Tissue

2.2.4.2.1. DNA Extraction

For DNA extraction, tissue was transferred from the cryovials onto a petri dish placed on dry ice, to minimise thawing of unused tissue. In a fume hood, a 25-50mg section of tissue was dissected using a sterile scalpel and placed in a 1.5ml microcentrifuge tube.

DNA was then extracted using the DNeasy Blood and Tissue Kit (Qiagen GmbH D-40724 Hilden, Category number 69504 and 69506) and the protocol below:

i) 180µl buffer ATL and 20µl proteinase K were added to the tube, then mixed by vortexing and incubated at 56°C until completely lysed (up to 24-48 hours).

ii) 200µl buffer AL was added and mixed by vortexing then incubated at 56°C for 10 minutes

iii) 200µl 100% ethanol was added and mixed by vortexing

iv) The mixture was pipetted into a DNeasy Mini spin column placed in a 2ml collection tube, centrifuged at ≥6000 xg (8000rpm) for 1 minute, discarding the flow-through and the collection tube.

v) The spin column was placed in a new 2ml collection tube, 500µl buffer AW1 was added and then centrifuged for 1 minute at ≥6000 xg (8000rpm), discarding the flow-through and the collection tube.

vi) The spin column was placed in a new 2ml collection tube, 500µl buffer AW2 was added and then centrifuged for 3 minutes at 20 000 xg (14 000rpm), discarding the flow-through and the collection tube.

vii) The spin column was placed in a new 2ml collection tube and centrifuged at 20000xg for 1 minute, discarding the filtrate and tube and then placed in a clean 1.5ml microcentrifuge tube.

viii) The DNA was eluted by adding 100µl buffer AE to the centre of the spin column membrane, incubated for 5-10 minutes at room temperature and then centrifuged for 1 minute at ≥6000 xg (8000rpm).

ix) The eluted DNA was then stored at -20°C for bisulfite conversion and further analysis.

2.2.4.2.2. DNA Concentration Analysis

As for the FFPE samples, DNA concentration for the FF samples was measured using both the NanoDrop spectrophotometer and the Qubit fluorometer.

2.2.4.2.3. Bisulfite Conversion

As for the FFPE samples, the same method was used for bisulfite conversion; however, the ligation step was not required.

2.2.5. The Illumina HumanMethylation 450K BeadChip array

2.2.5.1. Background to the Illumina HumanMethylation 450K BeadChip array

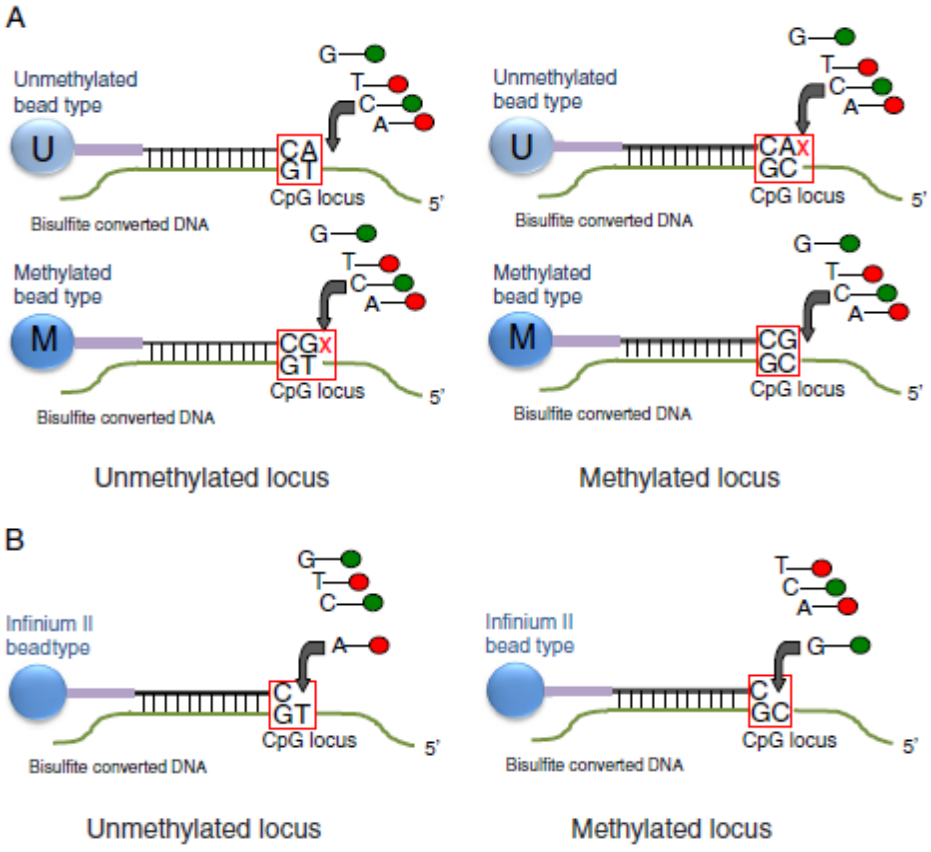
The Illumina bead array technology is a more recent addition to the methylation technologies described in section 1.4.2.3, and started with the GoldenGate and Infinium HumanMethylation 27 array platforms. These technologies used shorter stretches of target DNA including fragmented DNA extracted from FFPE tissue. The Infinium HumanMethylation 27K platform was a high throughput single nucleotide extension or methylation sensitive single nucleotide primer extension that analysed over 27 000 CpG dinucleotide sites, selected predominantly from the promoter regions of 14 000 annotated genes. Data was expressed as ß, continuous variables between 0 and 1, and represented the ratio of the intensity of the methylated bead type to the combined locus intensity. The BeadChip technology allowed detection of methylation levels down to as little as 2.5% [230].

The Infinium methylation array uses beads with long target-specific probes designed to interrogate individual CpG sites, with a span of 50 bases. In the Infinium HumanMethylation 27K array, the Infinium type I primer extension assay was used where there were two probes per CpG locus, one as an unmethylated and one as a methylated query probe. The 3' prime terminus of the probe matched either the protected cytosine (C, methylated) or the thymine base (T, unmethylated) resulting from bisulfite conversion and whole genome amplification. As methylation changes are regionally correlated and 90% of CpG sites within 50 bases have the same methylation status [243][244], the use of probes with a span of 50 bases is thought to give highly correlated results.

The Infinium type II assay design has one probe per locus to detect CpG sites located in regions of low CpG density, as illustrated in Figure 2.3. The 3' terminus of the probe complements the base directly upstream of the query site while a single base extension results in the addition of a labelled G or A base, complementary to either the methylated C or unmethylated T. The Infinium II probes can have up to 3 underlying CpG Sites within the probe sequence without compromising data quality. This allows the methylation status at a query site to be assessed independently of neighbouring CpG sites. The Illumina HumanMethylation 450K BeadChip

array (450K array) is made up of both Infinium type I (28%) and II (72%) assays and allows rapid genome-wide methylation analysis of hundreds of thousands of CpG sites across a large number of samples.

Figure 2.3: Infinium Methylation Type I versus Type II Assay [231]
2.3A shows the Infinium I assay design with 2 bead types to bind methylated (C) and unmethylated (T) CpG loci respectively.
2.3B shows the Infinium I assay with a single bead type for each CpG locus. Methylation state is detected by a single-base extension with 'A' incorporated at an unmethylated query site (T) and 'G' incorporated at a methylated query site (C).



2.2.5.2. Processing of the Illumina HumanMethylation 450K BeadChip array

The array analysis was carried out by UCL genomics, ICH microarray centre, Institute for Child Health, according to standard, quality assured techniques.

The 450K BeadChip protocol comprised a number of steps including whole genome amplification, fragmentation, precipitation, hybridisation, washing, counterstaining and scanning, as per the standard Infinium protocol and took approximately 4 days.

In the whole genome amplification reaction, 4µl of bisulfite-converted DNA was mixed with MA1 and NaOH, vortexed, centrifuged, then incubated at 22°C for 10 minutes. Further MA1 and MSM was then added, mixed, centrifuged and incubated in the Illumina hybridisation oven

at 37°C for 20-24 hours. Enzymatic fragmentation was then required to produce bisulfite converted DNA at a length that would hybridise to the oligonucleotides on the array. The sample was removed from the oven and FMS was added to each sample, vortexed, centrifuged and incubated at 37°C for 1 hour. Isopropranol was then added to cause precipitation of fragmented DNA and incubated at 37°C for 5 min, then 4°C for 20 minutes. Precipitated DNA was then resuspended prior to hybridisation in RA1 buffer and placed in the hybridisation oven at 48°C for 1 hour. The samples were then denatured by heating the plate to 95°C for 20 minutes. Hybridisation to the BeadChips then occurred at 48°C for 16-24 hours. After hybridisation, the BeadChips were washed in solutions WB1 and PB1 to remove any unhybridised DNA. Labelled nucleotides were added to extend the oliogonucleotides to which the template DNA had hybridised. The signal from the stain was amplified via a multi-layer IHC staining process and the BeadChip were then scanned in the Illumina iScan.

2.2.6. Data Analysis

Data analysis was carried out using R statistical software (version 3.03) and validated algorithms developed at UCL Cancer Institute. Data for all epigenetic and genetic analysis was processed by a trained staff member and stored on a password secured database, accessible only to relevant members of the research team.

2.2.6.1. Methylation Data

For methylation data analysis, R statistical software and ChAMP packages were used, as these are validated and in use at the UCL Cancer Institute [245] and provide an integrated pipeline of the most commonly used normalization methods with novel techniques to detect differentially methylated regions (DMRs) and copy number(CN) aberrations.

The R console is a language and environment for statistical computing and graphics with a suite of software facilities for data manipulation, calculation and graphical display.

The ChAMP package is a pipeline that integrates both available 450K analysis methods as well as several analysis options not available in other packages [245]. It uses the data import, quality control and SWAN (subset-quantile within array normalisation) functions as part of the *minfi* package. The PBC (Peak Based Correction) method and the BMIQ normalization method are also utilised.

It includes SVD (singular valve decomposition) methods to assess batch effects and allows correction for this using the ComBat method. The new Probe Lasso method also allows identification of DMRs and there is additional functionality to analyse 450K for CN variation.

After R was installed, the required packages were downloaded including *minfi, DNAcopy, impute, marray, limma, preprocessCore, RPMM, sva* and *IlluminaHumanMethylation450kmanivest*., followed by the *ChAMP* package.

Script 1 (appendix 2A) contains the instructions to enter into R as part of the ChAMP pipeline to load the data and run the quality control steps described.

The data was processed through a number of quality control and normalisation packages, as demonstrated in Figure 2.4. The raw data generated from the 450K array was loaded from .idat files using the *minfi* package and ChAMP filtered the data to detect probes with a detection p-value greater than 0.01 in at least one sample. A sample list was generated showing each sample and the fraction of failed probes per sample. Samples with a failure rate of >5%-10% were then removed manually and the data was reloaded into R to increase the number of probes that could be analysed.

The load function then generated 3 images, a density plot, MDS (multidimensional scaling) plot and sample cluster plot. The density plot showed unnormalised beta values and could identify any samples with significantly different profiles. The beta MDS plot allowed a visualisation of the similarity of the samples based on the top 1000 most variable probes. The cluster plot also visualised the similarity of samples based using hierarchical clustering on all probes.

A second control step was the intra-array normalisation of data using BMIQ to adjust the data for bias introduced by the Infinium type II probe design. As outlined above, Infinium type I and II probes are used on the 450K array and have two different designs with different DNA methylation distributions that may bias downstream analysis. The Infinium type II probes are generally less accurate and reproducible than the Infinium I probes [246] and thus, BMIQ normalisation was used to reduce this technical variation and eliminate bias caused by the type II probes. Poorly performing probes were automatically removed and three further images were then generated of the normalised beta MDS plot, cluster plot and density plot.

A third control was performed via Singular Value Decomposition (SVD) that assessed the most significant components of variation and batch effects, associated with both biological aspects and technical factors of the array. SVD was applied to the data matrix to compute p-values based on the significance of each component and generated a heatmap of 18 internal controls as well as the top 6 principle components correlated to the provided information.

The ComBat function was then used to correct for batch effects.

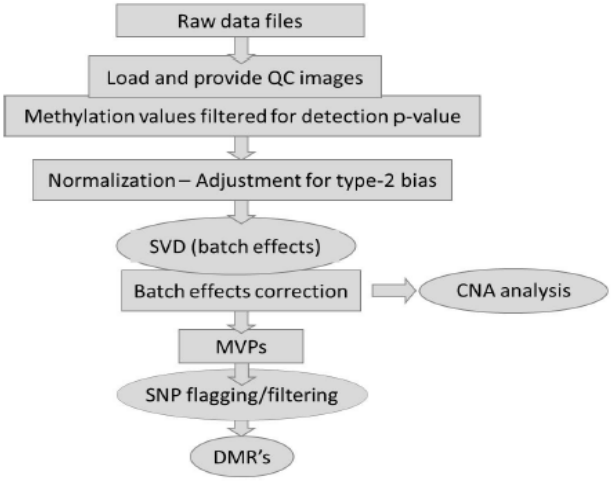


Figure 2.4: An Overview of the ChAMP Pipeline
SVD: singular valve decomposition, CNA: copy number alteration, MVP: methylation variable positions, SNP: single nucleotide polymorphism, DMR: differentially methylated regions.

The *limma* package was then utilised to calculate the p-value for differential methylation using a linear model. It generated a file including all probes with methylation variable positions (MVPs) between the two groups of interest, with a given p-value. A file was also saved with the significant MVPs that could be used for downstream pathway analysis.
The DMR hunter-probe lasso then computed a data frame of probes divided into discrete differentially methylated regions (DMRs), with accompanying p-values. DMRs are segments of the genome that show a quantitative alteration in DNA methylation levels between 2 groups. The DMR and MVP files contained information on the probes, associated genes, hypomethylation and hypermethylation, genomic distribution and CpG content. This was performed for NvC, NvA and NvC comparisons, as only 2 groups could be compared per pipeline analysis for MVPs and DMRs.

There were a number of other features in addition to the ChAMP pipeline that were performed for further data analysis.
Firstly, cluster dendrograms and heatmaps showing either the 500 or 1000 most variable probes between two or three comparator groups could be generated using script 2 (appendix 2B). These figures demonstrated which samples clustered together, regardless of whether significant MVPs or DMRs were detected, and could also incorporate clinical data and its relationship to sample clustering. This was performed for NvC, NvA, AvC and NvAvC comparisons. For the NvAvC comparison, the raw data was loaded as above but the limma package and hunter probe lasso steps were not performed.

The heatmaps demonstrate differential methylation changes based on a beta value between 0 and 1. The beta value is the fluorescence intensity ratio between methylated and unmethylated probes. A ratio value of 0 (yellow) equals non-methylation of the locus, 1 equals total methylation (blue) and 0.5 means that one copy is methylated and the other is not.

Study data could also be compared to publicly available TCGA data through Marmal-aid [247], using script 3 (appendix 2C). This is a standardised, freely available database for DNA methylation that incorporates the majority of publicly available Illumina HumanMethylation450 data into a single repository, including data from TCGA.

Secondly, when MVPs and DMRs were present, a gene list was generated based on the associated probes and could be further analysed. The Web-based Gene Set Analysis Toolkit (Web Gestalt) was used to perform enrichment analysis via KEGG analysis or Pathway Commons analysis to determine the gene associations and pathways involved [248]. WebGestalt is designed for functional genomic, proteomic and large-scale genetic studies from which a large number of gene lists are generated. WebGestalt uses information from different public resources to inform these gene associations. To initiate the search, an organism (hsapiens) and a gene ID type (hsapiens_gene_symbol) were defined. The gene list was then uploaded and could be compared to a predefined reference set (hsapienes_genome). A hypergeometric statistical method was used and a significance level was selected of $p<0.001$. Once the enrichment analysis was performed through WebGestalt, the gene list was reviewed within the MVP and DMR tables to assess the location of the gene, whether in the promoter region (TSS200, TSS1500 or 1xt exon), body or other location.

Thirdly, genes of interest were analysed through the R console using script 4 (appendix 2D), and differential methylation across normal endometrium, AEH and EC tissue and within a specific locality of the gene could be identified and assessed for statistical significance using the Welch Two Sample t-test.

2.2.6.2. Copy Number Variation analysis

The raw data from the 450K array could also be used for analysis of CN variation (CNV). After the normalisation, SVD and batch correction steps were performed, CNV was analysed using script 5 (appendix 2E). The CNV analysis only allowed 2 groups to be compared at one time and as such, the following comparisons were performed: NvC, NvA, AvC. The difference in CNV between normal, AEH and EC tissue could then be analysed to see where the change in

CNV occurred, whether between the normal and atypical transition or between atypical and cancer.

This step also generated raw data per sample demonstrating CN amplification and deletion per chromosome. This raw data was analysed through the GISTIC2 module Version 6.2 on the Broad Institute GenePattern web-based server [249], using the following parameters.

- Refgene file: Human Hg18 (contains information about the location of genes and cytobands on a given build of the genome)
- Seg file: segmented data for all the samples. The file includes sample name, chromosome number, start position (in bases), end position (in bases), number of markers in the segment and the SegCN [log2()-1 of CN].
- Markers file: contains the marker names and positions based on the markers in the original dataset. This file includes the marker name (probe number for 450K array), chromosome location and marker position.
- Gene GISTIC: yes (flag indicating that the gene GISTIC algorithm should be used to calculate the significance of deletions at a gene level)
- Amplifications threshold: 0.1 (threshold for copy amplifications. Regions with a log2 ratio above this value are considered amplified)
- Deletions threshold: 0.1 (threshold for copy deletions. Regions with a log2 ratio below the negative of this value are considered deletions)
- Join segment size: 4 (smallest number of markers to allow in segments from the segmented data. Segments that contain a number of markers less than or equal to this number are joined to the neighbouring segment that is closest in CN)
- Qv threshold: 0.25 (threshholding value for q-values)
- Remove X: yes (flag indicating whether to remove data from the x-chromosome before analysis)
- Cap val: 1.5 (minimum and maximum cap values on analysed data. Regions with a log2 ratio greater than the cap are set to the cap value; regions with a log2 ratio less than –cap are set to –cap)
- Confidence level: 0.75 (confidence level used to calculate the region containing a driver)

The GISTIC2 module identified regions of the genome that were significantly amplified or deleted across the samples for comparison. A number of parameters were assessed including the G-score, Q-values, peak region, wide peak and broad versus focal events. The G-score is based on the amplitude of the aberration as well as the frequency of its occurrence across samples. The Q-value indicates the false discovery rate for the aberrant regions and

probability that the event is due to a chance fluctuation. This defines the statistical significance of each aberration based on comparing the observed statistic to that expected by chance. Regions with a Q-value below a defined threshold are considered significant.
For each significant region, a 'peak region' is identified with the greatest amplitude and frequency of alteration and a 'wide peak' is defined for identifying the most likely gene targets in the region. Each region is also labelled as resulting from a broad event (longer than half a chromosome arm) or from a focal event or both.
Thus, the GISTIC module reported the genomic location and Q-values for each of the regions with CNV, along with the genes found in regions of significant amplification or deletion.

Details of the output files from GISTIC2 module are outlined as follows:

- Segmented CNV heat map per comparison
- Amplification plot for each comparison
- CN amplification loci and associated genes
- Deletion plot for each comparison
- CN deletion loci and associated genes

The loci and gene data were then analysed similar to the methylation data through Web Gestalt and the KEGG and Pathway Commons enrichment analyses to identify the significant cancer related genes and associated pathways. A p-value ≤0.1 was used for this analysis.

CHAPTER 3: Enumeration and Clinical Correlation of CTCs in Endometrial Cancer

3.1. Introduction

The currently used prognostic and predictive parameters that guide patient management in EC are based on historical clinicopathologic factors, without correlation with molecular profiling that is standard in breast, lung and colorectal cancers for example. There is a need to develop validated prognostic and predictive biomarkers to better guide therapeutic management and streamline the drug development process in EC.

CTCs are epithelial cancer cells known to circulate in the peripheral blood of patients, including in breast, prostate, ovarian, colorectal and lung amongst others, and are not detected in healthy individuals [88]. As CTCs can be analysed from a simple blood draw, their use for predictive purposes and to assess drug response and resistance is appealing, as it can spare patients undergoing more invasive procedures. Novel technologies also allow molecular profiling on CTCs that could assist with treatment selection and identify changes in the molecular profile of an individual's tumour over time.

Numerous techniques for the isolation and enumeration of CTCs have been reported over the years, but only the Veridex CellSearch method is analytically validated and FDA approved for use for prostate, breast and colorectal cancers as a predictive and prognostic biomarker. In prostate and breast cancer, a CTC count ≥5 [102, 107] is used to define a poor prognostic group, while in colorectal cancer, a CTC count ≥3 [111] is used.

The first step in CellSearch CTC detection is based on an anti-EpCAM ferrofluid capture reagent. There are arguments for and against the use of EpCAM as a selection tool for CTC detection; however, as the CellSearch platform is the only FDA approved and standardized system to capture and quantify CTCs and EC has a reported high expression of EpCAM [250], it was used for this feasibility study.

As detailed in the introduction section 1.4.1.3, stathmin is a cytosolic phosphoprotein and regulator of the microtubule cytoskeleton and cell cycle [148] that has also been shown to be a highly relevant biomarker in EC [163]. Not only is it overexpressed in EC with evidence of predictive value for metastases and survival, it may also predict response to chemotherapy and act as a surrogate marker of PI3K pathway activation [122], one of the most commonly aberrant pathways in EC. At the UCL Cancer Institute, stathmin has been validated for IHC assessment on FFPE tissue, as well as for use in cell lines and CTCs, as outlined in Methods section 2.1.7.

Here, a feasibility study was conducted for the enumeration and molecular analysis of CTCs as a prognostic and predictive biomarker in EC. Peripheral blood samples were taken from

patients with locally advanced or metastatic EC at any stage of their treatment and patients were then followed over time with repeat CTC assessment. In conjunction with the CTC collection, clinical data was collected on clinicopathologic parameters and treatment history.

The aims of the study are summarised in Figure 3.1 below.

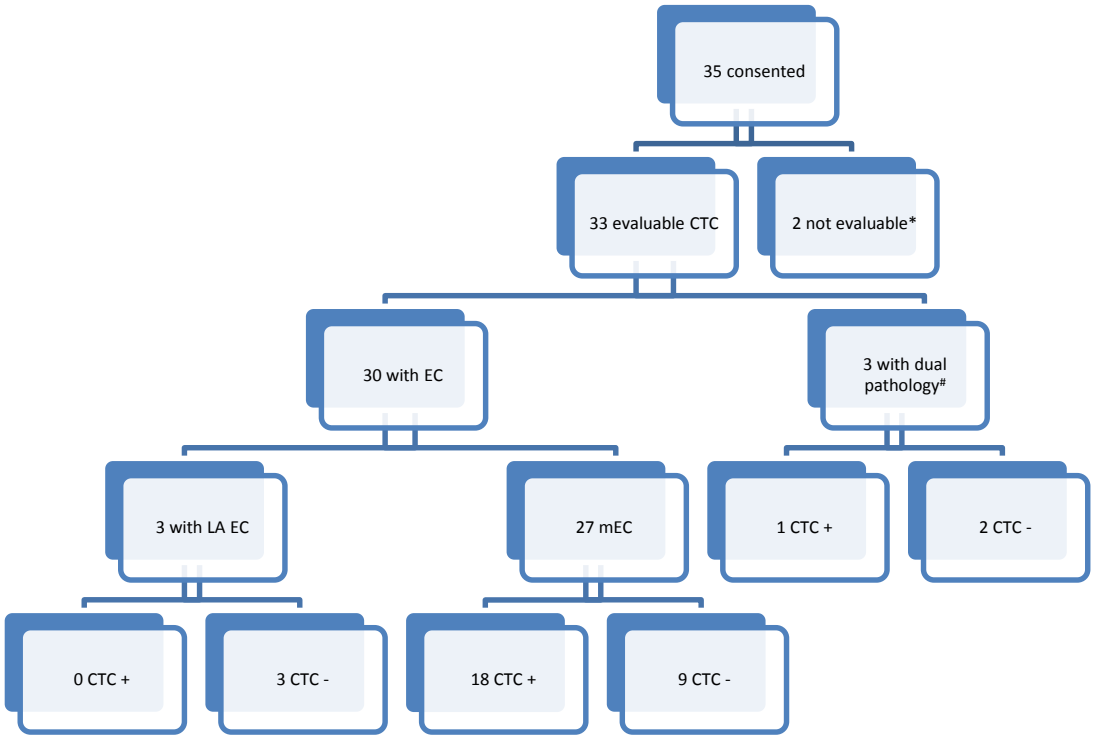
Figure 3.1. CTC Enumeration and Molecular Profiling in EC (EEC and NEEC)

- Feasibility of CTC enumeration and molecular profiling in EC on the Veridex CellSearch platform
- Correlation with clinical data and treatment
- Correlation of tumour tissue IHC with CTC profiling

3.2. Summary of Patient Recruitment and CTC Enumeration

35 patients were consented for involvement in this study. Figure 3.2 summarises the CTC testing and results for patients on study, related to whether they had dual pathology with endometrial and ovarian/primary peritoneal cancers, locally advanced or metastatic disease.

Figure 3.2: CTC Enumeration in the 35 Consented Study Patients



EC: endometrial cancer; m: metastatic; LA: locally advanced, +: positive, -: negative
* 1 consented but did not have a further blood test for CTC collection, and the other had a sample collected, but the CellSearch system malfunctioned at the time of analysis.
# 3 patients (10, 16, 18) had pre-operative CTC assessment performed based on staging as locally advanced endometrial cancer, and were then confirmed at surgery to have synchronous stage I endometrial and stage III or IV ovarian/primary peritoneal cancers.

3.3. CTC Detection on CellTracks Analyser II

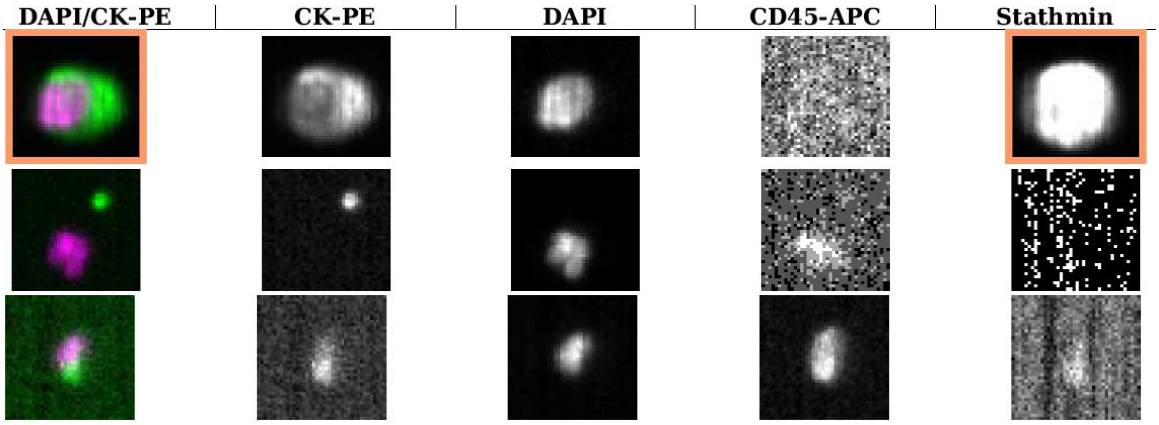
As outlined in the methods section 2.1.5, the CellTracks Analyser II displays CTC candidate images generated after a blood specimen has been processed on the CellTracks Autoprep system. These images were reviewed by the operator and researcher to assess if any fulfilled the criteria for a CTC. A CTC was defined as being positive for cytokeratins (CK-PE) and nuclear staining (DAPI), negative for leucocyte staining (CD45-APC) and as having the correct morphology and CK-PE and DAPI overlay to be characterised as a tumour cell.

Figure 3.3 demonstrates examples of patients with positive and negative CTCs detected, with or without Stathmin positivity. Figure 3.4 illustrates the total number of patients with positive CTC counts and those with positive Stathmin.

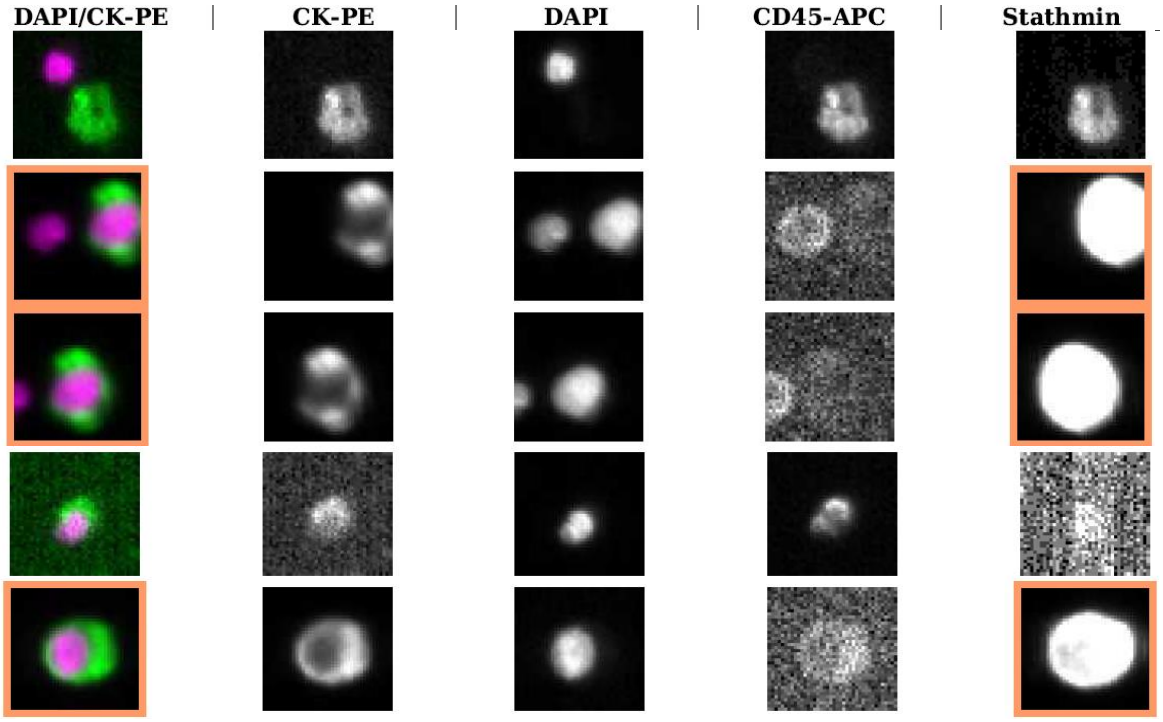
Figure 3.3: CellTracks Analyser CTC Candidate Images and Interpreter Detection

A: Patient 9, B: Patient 10, C: Patient 22, D: Patient 26, E: Patient 27
Orange boxes mark those images which met the criteria for a positive CTC.

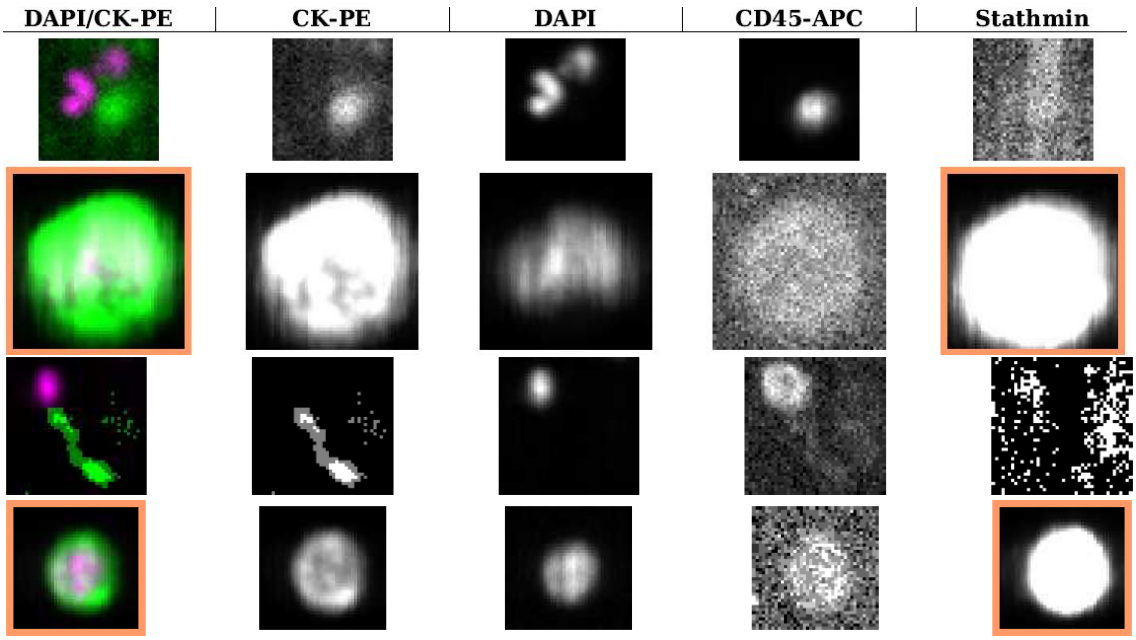
A: Patient 9. There is one CTC positive and Stathmin positive image with appropriate overlay of positive CK-PE and DAPI images. Of the 2 negative images, one has incorrect overlay of CK-PE and DAPI images and the second is CD45-APC positive.



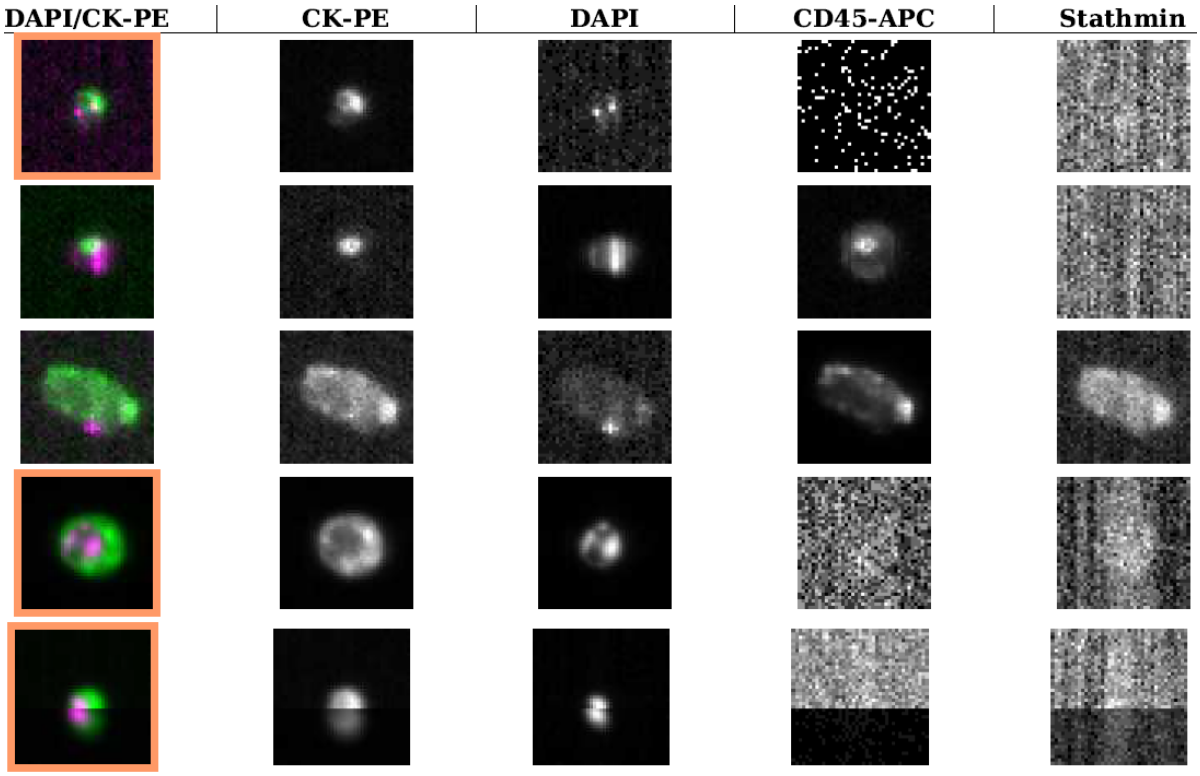
B: Patient 10. There are 3 positive CTC images that are also Stathmin positive, with appropriate overlay of positive CK-PE and DAPI images. Of the 2 negative images, one has incorrect overlay of positive CK-PE and DAPI images and the second is a blurred image that is CD45-APC positive.



C: Patient 22. There are 2 positive CTCs with appropriate CK-PE and DAPI overlay and negative/non-specific CD45-APC staining. Both are also stathmin positive. Of the 2 negative images, both have incorrect CK-PE and DAPI overlay and the top image also has positive CD45-APC staining.



D: Patient 26. There are 3 positive CTC images, with appropriate overlay of positive CK-PE and DAPI images, but none are stathmin positive. Of the 2 negative images, both are CD45-APC positive.



E: Patient 27. There are 3 positive CTC images with appropriate overlay of positive CK-PE and DAPI images and one of which is stathmin positive. The negative image does not have correct overlay of the CK-PE and DAPI images and is CD45-APC positive.

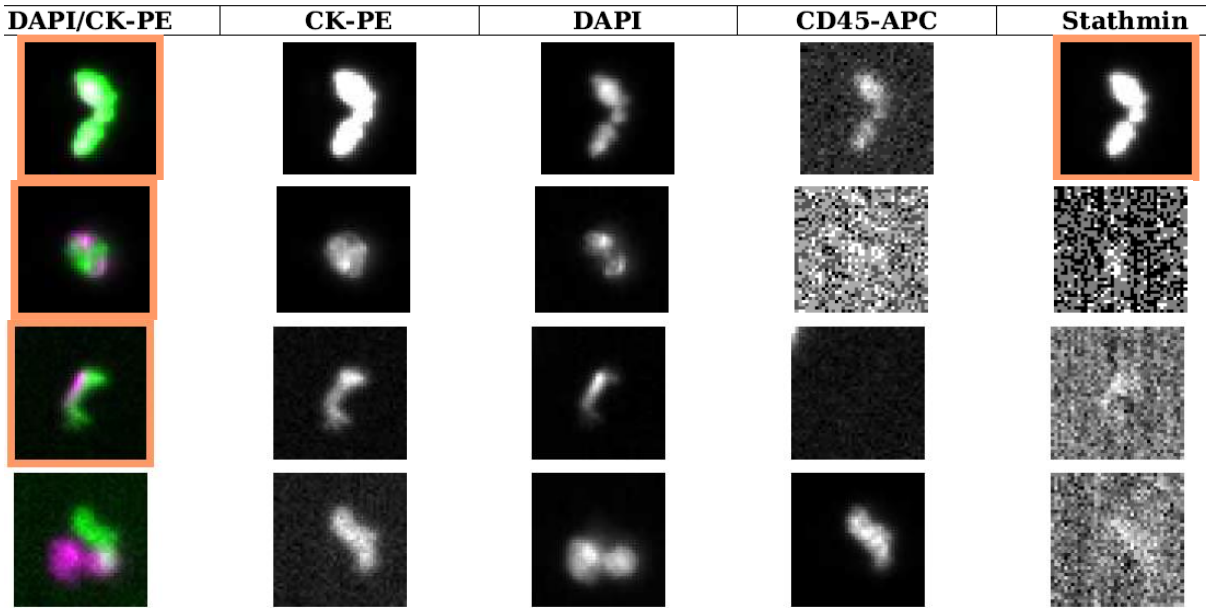
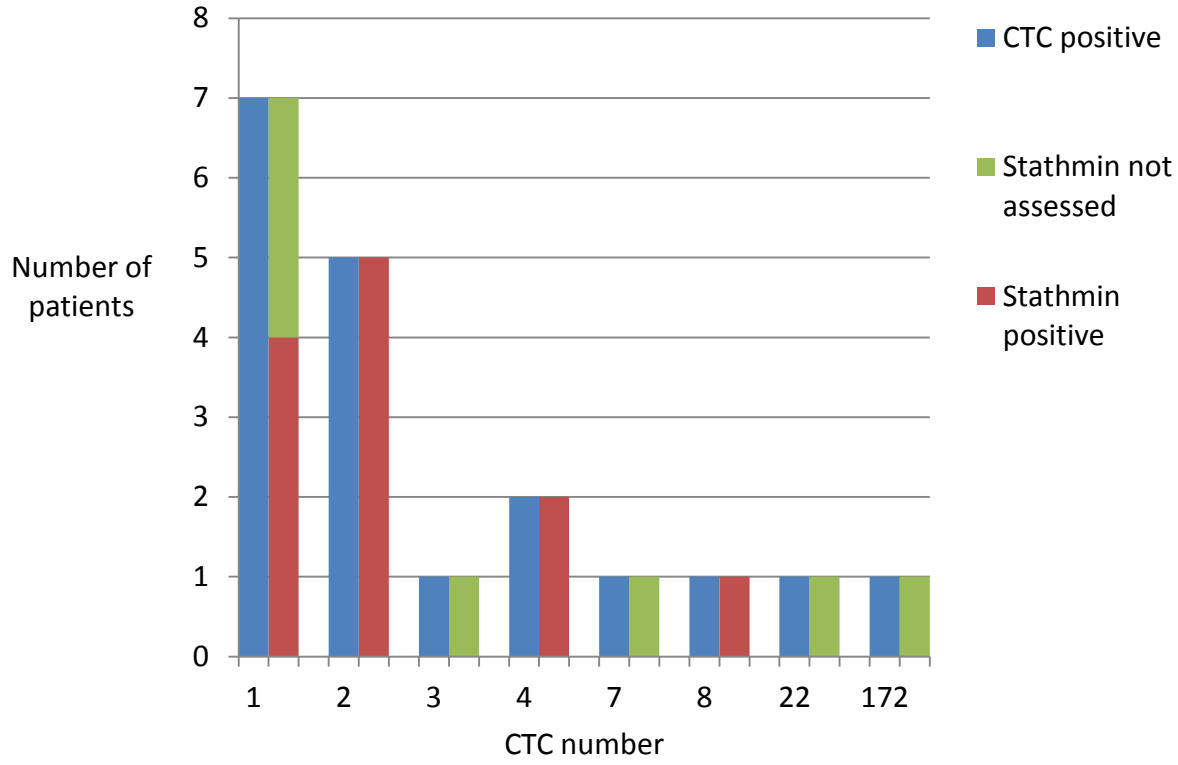


Figure 3.4: Number of Patients with Positive CTC counts and Stathmin Staining at Baseline
The majority of patients that are CTC positive have counts of 1-2, though there are 7 patients with CTC counts of ≥3 and 4 patients with CTC counts ≥5.



1 patient with CTC 2 had dual endometrioid endometrial cancer and primary peritoneal cancer. Stathmin was not assessed for 3 patients with CTC 1, 1 patient with CTC 3 and patients with CTC 7, 22 & 172.

From the 19 patients with positive CTC count, 7 patients did not have stathmin assessed due to supplier restrictions. Of the 12 remaining patients, all demonstrated positive stathmin staining, though the proportion of positive cells varied between patients.

From the 12 patients, 9 had all CTCs stain positive with stathmin while the remaining 3 patients had 1-2 cells stain positive for stathmin, from a total count of 4-8. This is detailed in Table 3.1.

3.4. Clinical Data, Treatment Course and Observations

The clinical course at the time of CTC collection for the 19 (58%) CTC positive (+) patients and the 14 (46%) CTC negative (-) patients are outlined in Tables 3.1 and 3.2 respectively including stage and treatment at the time of CTC collection, results of CTC assessment, time to recurrence for patients who had early stage disease at initial diagnosis and survival.
When patients with dual pathology are excluded, 18 from 30 patients (60%) were CTC+ and 12 from 30 patients (40%) were CTC-.
Further details of the initial treatment and treatment at relapse is included in the Appendix Sections 3A & 3B. A summary of the key findings from the comparison of these clinical factors in the CTC+ and CTC- patients are presented in Table 3.3.

Pathologic features from primary surgery, haemoglobin and albumin results are included in the Appendix in Table 3C. A summary of the key findings from the comparison of these pathological factors in the CTC+ and CTC- patients are presented in Table 3.4.

Table 3.1: Clinical Course of CTC positive (+) Patients' with Endometrial Cancer

Pt	Disease status at CTC collection	Time to recurrence (m)	Treatment at CTC collection	CTC count	Stathmin count	Survival at Feb 2014 (m)
3	RM	38	Ch C1	7	ne	2
7	IV	ne	1m post chemo	1	ne	21
8	RM	11	At relapse, no treatment	22	ne	1
17	RM	19	Ch C6	8	1	9
14	RM	9	1m post relapse 4m post relapse	0 1	0 1	12 (alive)
6	RM	27	Ch C2 C4	172 3878	ne 0	4
9	IV	ne	Ch C1 C4 4m post ch 6m post ch	2 0 2 2	2 0 2 ne	12
12	RM	24	Ch C1 C4 1m post ch 4m post ch	0 0 0 3	0 0 0 ne	12 (alive)
13	RM	26	Ch C2 C4	1 0	1 0	26 (alive)
24	RM	22	Ch C1 C3 C6	2 1 H	2 1 H	10 (alive)
26	IV	ne	Ch C1 C3 C6	4 1 H	2 1 H	5 (alive)
27	IV	ne	Ch C1	4	2	7 (alive)

			C4 C6	0 0	0 0	
31	RM	8	1m post relapse, 2m pre-ch C1 C2	0 1 0	0 1 0	5 (alive)
11	IV	ne	Ch C1 C6 1m post ch 3m post ch, at relapse	0 2 0 0	0 2 0 0	15
15	RM	29	Ch2 C1 C6 5m post ch2, 1m pre ch3	0 1 0	0 1 0	10 (alive)
22	IV	ne	Ch C1 C4	2 0	2 0	9
25	RM	10	Pre-S Post-S Ch C3	0 0 1	0 0 ne	7 (alive)
28	IV	ne	Ch C2 C4 C6	0 1 0	0 ne 0	5 (alive)
10	**RM**	**3**	**Pre-S** **Ch2 C2** **Ch2 C3**	**2** **0** **1**	**2** **0** **1**	**na**

Patients 3, 7, 8 and 17 had only 1 sample taken for CTC enumeration. Patient 14 had 2 CTC samples taken during follow-up but received no active treatment.
Patients 6, 9, 12, 13, 24, 26, 27, 31 had multiple samples taken for CTC assessment and the CTC results were concordant with clinical outcome. Patients 11, 15, 22, 25, 28 had multiple samples taken for CTC assessment and the CTC results were discordant with clinical outcome
RM: recurrent, metastatic disease, IV: FIGO stage IV, S: surgery, Ch: chemo, Ch2: second-line chemotherapy, Ch3: third-line chemotherapy, C: cycle number, H: haemolysed, ne: not evaluable, na: not available
Survival is based on time from diagnosis with either stage IV or recurrent metastatic disease. All chemotherapy cycles were given for stage IV or for recurrent metastatic disease, unless stated otherwise.
Bolded row (patient 10) had dual IA endometrioid endometrial and IIIC primary peritoneal cancer at diagnosis. First CTC assessment was pre-operative. The 2nd and 3rd CTC counts were taken after diagnosis with recurrent disease. This patient is excluded from subsequent analysis due to the dual pathology.

Table 3.2: Clinical Course of CTC negative (-) patients' with Endometrial Cancer

Pt	Disease status at CTC collection	Time to recurrence (m)	Treatment at CTC collection	CTC count	Survival at Feb 2014 (m)
1	IV	ne	Ch C1 C6	0 H	13
2	RM	51	10m post relapse 1m post ch & pre RT brain	0	13
34	IV	ne	44m post diagnosis, pre-ch3	0	45 (alive)
35	RM	11	Ch C2	0	3 (alive)
4	*IIIb*	*ne*	*3m post diagnosis, pre-RT* *14m post diagnosis* *17m post diagnosis*	*0* *0* *0*	*18 (alive)*
20	RM	5	28m post relapse (on H) 32m post relapse (on H)	0 0	36 (alive)
21	RM	18	Ch C1 C4	0 0	9 (alive)
23	*IB*	*ne*	*9m post diagnosis, 1m post Ch & RT* *3m post Ch & RT* *6m post Ch & RT*	*0* *0* *0*	*15 (alive)*
29	*IIIC*	*ne*	*Ch C2* *C5*	*H* *0*	*5 (alive)*
30	RM	95	Ch C2 C6	0 0	5 (alive)
33	IV	ne	Ch C1 C3	0 0	4 (alive)
19	RM	5	Post-S Post RT At relapse	0 0 0	3 (alive)
16	**RM**	**7**	**Pre-S** **Ch C4** **C6** **Post relapse, Ch2 C2**	**0** **0** **0** **0**	**na**
18	**RM**	**28**	**Post relapse, Ch C4** **Refused repeat**	**0**	**na**

Patients 1, 2, 34, 35 had only 1 sample taken for CTC enumeration.
Patients 4, 20, 21, 23, 29, 30, 33 had multiple samples taken for CTC assessment and results were concordant with clinical outcome. Patient 19 had multiple samples taken for CTC assessment and results were discordant with clinical outcome.
Patients indicated in italicised rows (patients 4, 23, 29) had stage I-III disease at time of CTC collection.
Patients indicated in bolded rows (patients 16, 18) had dual IA endometrial with IIIB serous ovarian and IIIA primary peritoneal cancers respectively at diagnosis.
Patients 16 and 18 were excluded in further analysis and survival data was not collected.
RM: recurrent, metastatic disease, IV: FIGO stage IV, S: surgery, RT: radiotherapy, Ch: chemo, Ch2: second-line chemotherapy, Ch3: third-line chemotherapy, H: hormones, C: cycle number, H: haemolysed, ne: not evaluable, na: not available
Survival is based on time from diagnosis with either stage IV or recurrent metastatic disease, except for patients with stage I-III disease who did not relapse.
All chemotherapy cycles were given for stage IV or for recurrent metastatic disease, unless stated otherwise.

Table 3.3: Differences in Treatment and Clinical Course between CTC+ and CTC- Patients

	CTC - (%), n=12 EC	CTC + (%), n=18 EC
Age (years)	66	66
Stage at diagnosis for EC pts	I/II: 5 (42), III/IV: 7 (58)	I/II: 6 (33) , III/IV: 12 (67)
Stage at collection for EC pts	I: 1 (8), III: 2 (17), IV/RM: 9 (75)	IV/RM: 18 (100)
Sites of metastases Peritoneal, LN, Liver	4 (44), 7 (77), 1 (8)	13 (72), 15 (83), 5 (27)
Grade 3 disease	7 (58)	15 (83)
Surgery alone*	0 from 9	4 from 11 (2 each of IA & IB)
Adjuvant chemo/RT/ both*	0/5/4	1/2/4
Time to recurrence (months)	Based on IA,B,II,IIIB,IIIC (n=6) 30.8	Based on IA,IB,IIIA,IIIC (n=11) 20.3
Survival (pts alive with stage IV disease at Feb 2014)	7 from 9 (78)	10 from 18 (55)
Survival status with stage IV disease based on f/u time	2 from 9 died at >12m (22) 5 alive at ≤12m f/u (56) 2 alive >12m f/u (22)	6 from 18 died ≤12m (33) 2 died >12m (11) 9 alive ≤12m f/u (50) 1 alive >12m f/u (6)

*based on patients that had stage I-III disease at diagnosis, f/u: follow-up
RM: recurrent metastatic, pts: patients

There were more patients with FIGO stage IV or recurrent metastatic disease at first CTC collection in the CTC+ group with 100% of patients having stage IV or recurrent metastatic disease when CTC+ compared to 75% when CTC -. Similarly, stage at initial diagnosis, which may have been years prior to CTC collection, was also higher when CTC+ with 67% having stage III/IV disease compared to 58% when CTC-. Grade 3 disease and the presence of peritoneal and liver metastases were also more common in CTC+ patients.

For patients who were sampled after disease recurrence, average time to recurrence was shorter in CTC+ patients at 20.3 months (range 9-38m) compared to 30.8 months (range 5-95m) in CTC- patients. The hazard ratio for recurrence was 1.7 in favour of CTC- patients, though the 95% confidence interval was wide, 0.5-5.2. The Kaplan-Meier plot for time to recurrence shows no significant difference between the groups, as demonstrated in Figure 3.5

In terms of the treatment received at initial diagnosis, all patients that were CTC- had received adjuvant therapy, including stage I patients. In contrast, for patients that were CTC+, 4 patients received surgery alone for stage IA/IB disease. This suggests that in these patients with early stage disease who recurred, the traditional clinicopathological features used to guide primary treatment, may not have adequately reflected the likelihood of recurrence.

In addition, more CTC- patients were alive at last follow-up compared to CTC+ patients. For patients who had died during follow-up, more CTC+ patients were deceased within 12 months after diagnosis with stage IV or recurrent metastatic disease compared to CTC- patients. The

hazard ratio for survival was 3.3, but the confidence interval was wide, from 0.7 to 16.2. The small sample size and limited follow-up time for some patients limit the statistical analysis of this data.

Figure 3.5: Kaplan-Meier plot for time to recurrence for CTC positive and CTC negative patients

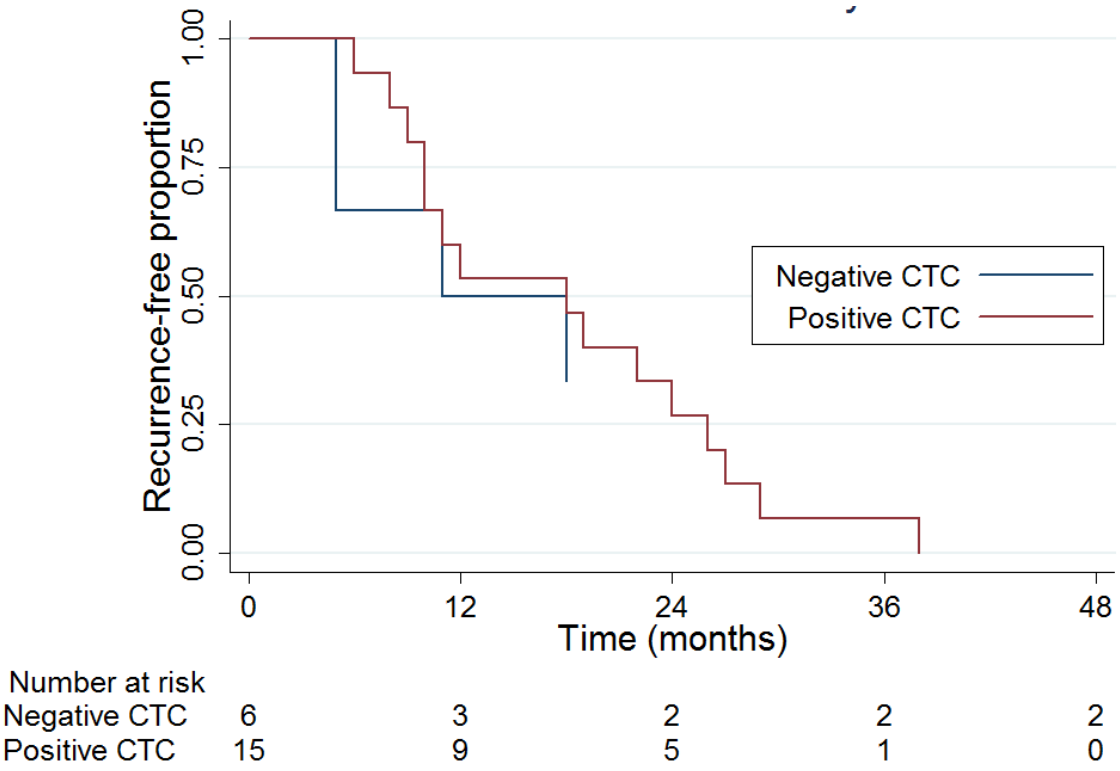


Table 3.4: Differences in Histology and Pathology Results between CTC Positive (+) and Negative (-) Patients

	CTC negative (%), n=12	**CTC positive (%), n=18**
Type II EC	6 (50)	14 (78)
Tumour size 2-5cm ≥ 5cm	2 na (17) 4 (33) 6 (50)	3 na (17) 4 (22) 11 (61)
MMI>50%	8 (67)	10 (56)
LVSI	6 (50)	8 (44)
Cervical invasion	3 (25)	5 (28)
Mean Haemoglobin	11.5	11.2
Mean Albumin	41.8	40.3

EC: endometrial cancer, MMI: myometrial invasion, LVSI: lymphovascular space invasion

There were more type II EC tumours and tumours larger than 5cm in the CTC+ patients compared to the CTC- patients. There were more CTC- patients with evidence of MMI greater than 50%. Other pathological features were similar between the two groups, and there was no difference in mean haemoglobin nor albumin.

3.5. Correlation of Clinical Data with CTC Status

Of the 18 patients with EC only and CTC+, 12 patients had follow-up samples taken during first-line chemotherapy for metastatic disease and 1 patient had follow-up samples taken during clinical trial involvement. 4 patients had only 1 sample taken before they clinically deteriorated and died. 1 CTC+ patient was too unwell for systemic therapy but was monitored in follow-up and had a rise in CTC count and clinical deterioration.

From the 13 CTC+ patients during first-line treatment, 8 had CTC results that were concordant with their clinical and radiological progress and 5 results were less clearly correlated, as demonstrated in Figure 3.6.

For the patients with concordant CTC trend and clinical course:

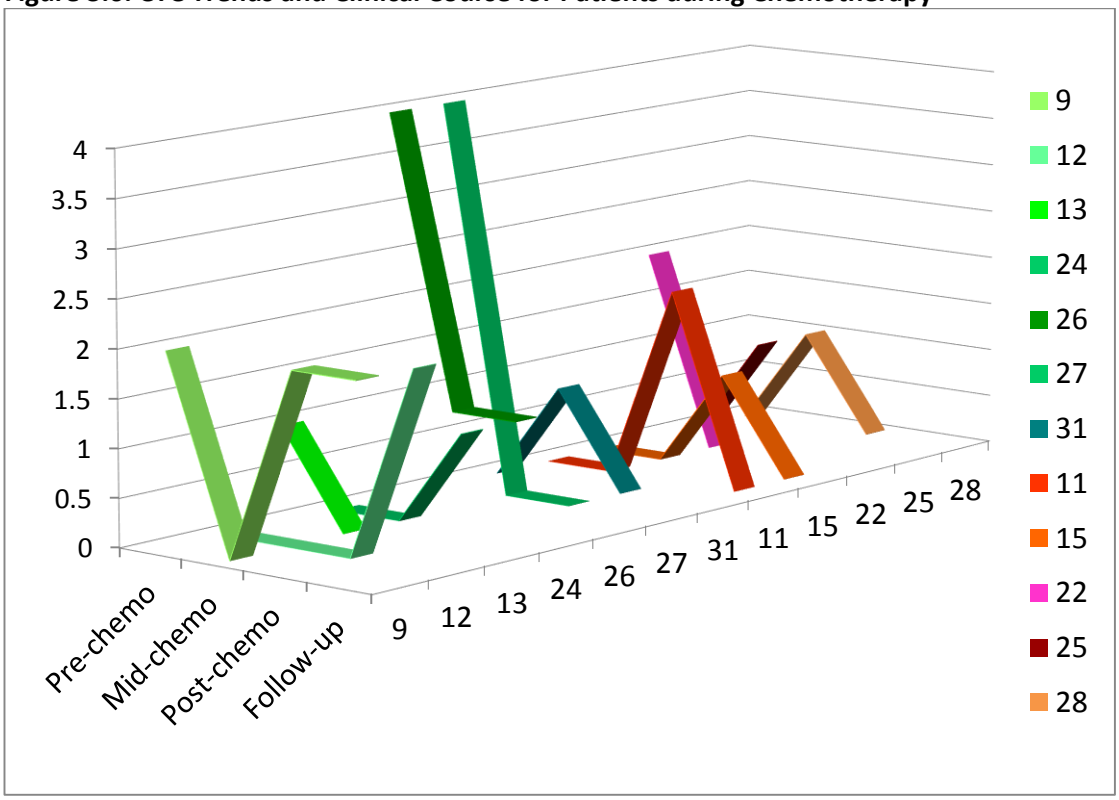
- Patient 6's CTC count rose markedly while on chemotherapy and the patient deteriorated shortly after.
- Patient 9 was baseline CTC+ and had a decrease in CTC count during chemotherapy with a radiographic response. At disease progression post-chemotherapy, the CTC count rose.
- Patient 12 was CTC- while on chemotherapy but CTC+ post-chemotherapy with evidence of disease progression.
- Patients 13, 24, 26 and 27 all were baseline CTC+ and had a decrease in CTC count during chemotherapy that correlated with a radiographic response.
- Patient 31 was CTC- on initial documentation of disease progression but was CTC+ at cycle 1 of the clinical trial. By cycle 2, the CTC count had again decreased with a corresponding radiographic response.

From the 5 patients with discordant CTC trend and clinical course, as illustrated in Figure 3.5:

- Patient 11 was CTC- at the beginning of chemotherapy and CTC+ at cycle 6, though a CT scan was stable at this time. Over the next 3 months, she had two further negative CTC counts, the last count proceeding disease progression by only one month.
- Patient 15 was CTC- at baseline and CTC+ at the end of chemotherapy, though the CT scan was stable at this time. The patient had new brain metastases 3 months later. A further 2 months later, the repeat CTC count was 0 but the CT demonstrated disease progression.
- Patient 22 was CTC+ at baseline and the count decreased at cycle 4. However, the patient had radiographic disease progression at that time and died.

- Patient 25 was CTC- at baseline and CTC+ at cycle 3, when the patient was clinically stable. A CT scan 2 months later showed her to be disease-free.
- Patient 28 was CTC- at baseline and CTC+ at cycle 4, when a progress CT showed a partial response to treatment. At cycle 6, an end of treatment scan confirmed this response and the CTC count also returned to 0 at this point.

Figure 3.6: CTC Trends and Clinical Course for Patients during Chemotherapy



Patients marked in a green shade (9, 12, 13, 24, 26, 27, 31): CTC trend concordant with clinical course. Patient 6 was not included as the CTC count was too high to depict here.
Patients marked in a red shade (11, 15, 22, 25, 28): CTC trend discordant with clinical course

For the 12 patients with EC alone who were CTC- at baseline and in follow-up, 7 patients had a clinical and/or radiological picture that correlated with the CTC result, 1 had a discordant picture and 4 had only a single CTC assessment and evaluation of concordance was limited.

Of the 7 patients with concordant CTC results and clinical outcome:

- Patients 4, 23 and 29 had localised or locally advanced disease at first CTC assessment and remained disease-free or stable during follow-up.
- Patients 21, 30 and 33 had metastatic disease at diagnosis and had CTC assessment performed during first-line chemotherapy. All had CTC counts of 0 pre-chemotherapy and either mid or post chemotherapy, corresponding with radiological response at these time points.

- Patient 20 had metastatic lung disease, managed with hormonal therapy and was clinically stable throughout the follow-up period.

For the patients with discordant results or limited follow-up:

- Patient 19 had 3 CTC samples taken, of which the first was post-operatively and the 2nd and 3rd at the time of disease recurrence. All were 0 despite the development of ascites and omental disease at the time of the last assessment.
- Patients 2 and 34 were CTC- though the samples were taken at the time of progression, one in the brain and one in the lung.
- Patients 1 and 35 had a pre-chemotherapy sample taken but further samples either haemolysed or were not taken so further evaluation was limited.

3.6. Evaluation of a CTC Cut-off

As detailed in the introduction section 1.3.2 and discussed further in section 3.7, CTC values ≥5 for breast and prostate cancer and ≥3 for colorectal cancer have been used for prognostic and predictive purposes.

In this patient group, 7 from 18 EC patients had CTC ≥3 and 4 patients had CTC ≥5. In section 3.3, it is outlined that higher stage at diagnosis and CTC collection, the presence of type II disease, grade 3 disease, liver metastases, shorter time to recurrence and decreased survival were associated with CTC positivity. These parameters are reviewed in further detail in Table 3.5, for patients with CTC 1-2, CTC ≥3 and CTC <3, as well as for CTC ≥5 and CTC <5, though there are limitations to this with the smaller patient numbers.

Table 3.5: Clinicopathological Features of Patients with CTC ≥3 vs <3 and CTC≥5 vs <5.

	CTC- (%) n=12	**CTC+ (%) n=18**	**CTC 1-2 (%) n=11**	CTC<3 (%) n=23	**CTC≥3 (%) n=7**	CTC<5 (%) n=26	**CTC≥5 (%) n=4**
Stage at diagnosis I/II III/IV	 5 (42) 7 (58)	 **6 (33)** **12 (67)**	 **3 (27)** **8 (73)**	 8 (35) 15 (65)	 **3 (43)** **4 (57)**	 8 (31) 18 (69)	 **3 (75)** **1 (25)**
Sites of metastases Peritoneal Liver	 4 (44) 1 (8)	 **13 (72)** **5 (27)**	 **8 (73)** **0**	 11 (48) 1 (4)	 **6 (86)** **5 (71)**	 14 (54) 2 (8)	 **3 (75)** **4 (100)**
Grade 3	7 (58)	**15 (83)**	**9 (82)**	16 (70)	**6 (86)**	19 (73)	**3 (75)**
Type II EC	6 (50)	**14 (78)**	**9 (82)**	15 (65)	**5 (71)**	18 (69)	**2 (50)**
Time to recurrence (m)	30.8	**20.3**	**20.5**	24.1	**23.8**	24.6	**23.8**
Pts alive with stage IV/RM disease, based on f/u time	5 ≤12m (56) 2 >12m (22)	**9 ≤12m (50)** **1>12m (6)**	**6 ≤12m (55)** **1 >12m (9)**	11 ≤12m (55) 3 >12m (15)	**3 ≤12m (43)**	14 ≤12m (61) 3 >12m (13)	**0**
Pts deceased with stage IV/RM disease, based on f/u time	0 ≤12m 2 >12m (22)	**6 ≤12m (33)** **2 >12m (11)**	**2 ≤12m (18)** **2 >12m (18)**	2 ≤12m (10) 4 >12m (20)	**4 ≤12m (57)**	2 ≤12m (9) 4 >12m (17)	**4 ≤ 12m (100)**

RM: recurrent metastatic, m: months, f/u: follow-up
For survival figures, all CTC+ patients are included, whereas 3 CTC- patients had stage I-III disease and are not included in the survival analysis.

There is a progressive decrease in survival for patients with CTC 1-2, CTC ≥3 and CTC ≥5. In both the CTC ≥3 and CTC ≥5 patient groups, there were a higher percentage of patients with peritoneal and liver metastases and poorer survival, compared to those patients with CTC <3 and CTC <5 respectively.

In the CTC ≥3 patient group, grade 3 disease and type II EC were more common than in the CTC <3 patient group, as with the CTC+ patients overall, but this difference was not maintained in the CTC ≥5 patient group.

There were fewer patients with stage IV disease in the CTC ≥3 and CTC ≥5 groups compared to the CTC <3 and CTC <5 groups respectively. The difference in time to recurrence seen in the CTC+ group compared to the CTC- group was also less prominent in the CTC ≥3 and CTC ≥5 patient groups compared to the CTC <3 and CTC <5 groups respectively.

Thus, although the use of a higher CTC cut-off continues to show an association with survival, the smaller patient number in the subgroups limits interpretation of other parameters.

3.7. Correlation of EpCAM and Stathmin IHC with CTC status

EpCAM and Stathmin IHC were performed on those blocks available at UCLH Histopathology Department. The Tumour Immunostaining Score (TIS) was used to evaluate EpCAM and stathmin expression, based on the product of staining intensity multiplied by the proportion of immunoreactive cells in the areas of interest, as outlined in the Methods section 2.1.6 [97, 163]. For both EpCAM and stathmin, an intensity score was assessed as 0 (no staining), 1 (weak), 2 (moderate) and 3 (strong). For the EpCAM TIS, the fraction of positively stained tumour cells was defined as 0 (none), 1 (<10%), 2 (10-50%), 3 (51-80%) and 4 (>80%). For the stathmin TIS, the fraction of positively stained tumour cells was defined as 0 (none), 1 (<10%), 2 (10-50%) and 3 (>50%). The CTC results, EpCAM and stathmin IHC are outlined in Table 3.6. Table 3.7 summarises the staining intensity of EpCAM and stathmin in the CTC+ and CTC- patient groups.

Another research antibody for the PI3K pathway, p70S6 kinase, was also investigated to correlate with the stathmin findings. However, even with the optimisation techniques described in the methods to improve staining, consistency was not seen across the tested controls, nor in the small number of samples evaluated. As such, it was not reviewed on the blocks listed. In comparison, stathmin produced far more consistent staining.

Table 3.6: CTC Enumeration and Stathmin Analysis and FFPE tissue EpCAM and Stathmin IHC

Patient number	CTC positive	CTC Stathmin	EC	FFPE EpCAM TIS	FFPE Stathmin TIS
11	Y	Y	NE	No (0) – sarc	0
31	Y	Y	NE	Weak (4)	na
7	Y	na	NE	Moderate (6)	Moderate (6)
28	Y	na	NE	Intense (12)	na
3	Y	na	E	Intense (12)	Moderate (4)
8	Y	na	NE	Intense (12)	Moderate (4)
9	Y	Y	NE	Intense (12)	Moderate (4)
17	Y	Y	NE	Intense (12)	Moderate (4)
25	Y	na	E	Intense (12)	Moderate (4)
11	Y	Y	NE	Intense (12)	Moderate (6)
12	Y	na	NE	Intense (12)	Moderate (6)
13	Y	Y	NE	Intense (12)	Moderate (6)
14	Y	Y	E	Intense (12)	Moderate (6)
15	Y	Y	NE	Intense (12)	Moderate (6)
26	Y	Y	NE	Intense (12)	Intense (9)
23	N	na	NE	Weak (3)	Weak (2)
19	N	na	NE	Moderate (8)	Moderate (6)
4	N	na	E	Intense (12)	na
20	N	na	E	Intense (12)	Moderate (4)
34	N	na	E	Intense (12)	Moderate (4)
35	N	na	E	Intense (12)	Moderate (4)

Patients 4, 28, 31 did not have adequate tissue for Stathmin assessment or the blocks were recalled for trial testing. Patients in grey rows are CTC positive.
E: endometrioid, NE: non-endometrioid, sarc: sarcoma, B: biopsy, S: surgical resection sample, S I-III: stage I-III, S IV: stage IV, na: not available

Table 3.7: EpCAM and Stathmin IHC results relative to CTC Status

	CTC positive; n=15 (%)	CTC negative; n=6 (%)
EpCAM TIS		
Intense	12 (79.9)	4 (66)
Moderate	1 (6.7)	1 (17)
Weak	1 (6.7)	1 (17)
No staining	1 (6.7)	0
Stathmin TIS		
Intense	1 (6.7)	0
Moderate all [4,6]	11 [5, 6] (73.3 [33.3, 40])	4 [3,1] (66 [50, 16])
None/Weak	1 (6.7)	1 (17)
Not available	2 (13.3)	1 (17)

Overall, tissue samples of those patients that were CTC + demonstrated stronger staining for both EpCAM and Stathmin than the tissue from CTC- patients, though the degree of EpCAM staining could not be used to predict CTC status.

More intense stathmin IHC in the CTC+ patients may be consistent with the prognostic association with stathmin overexpression.

The majority of specimens assessed for staining were NEEC. Interestingly, all the EEC samples demonstrated intense EpCAM and moderate Stathmin overexpression. For the NEEC samples, there was greater variation in both EpCAM and Stathmin overexpression, ranging from weak to intense, though most of the stathmin staining had moderate intensity.
From the 7 patients that had Stathmin positive CTCs and tissue available, 6 demonstrated a moderate Stathmin TIS and 1 demonstrated an intense TIS.
As all the CTC samples that were evaluated for stathmin stained positive, correlation between CTC stathmin status with and stathmin IHC TIS was limited. As such, any difference in CTC status and in expression between tissue taken at initial diagnosis compared to that from metastatic disease, or that between surgical and biopsy specimens could not be assessed.

3.8. Discussion

3.8.1. CTC Enumeration

CTC enumeration and molecular analysis with stathmin on the Veridex CellSearch platform is feasible in advanced EEC and NEEC, with CTCs detected in 60% of patients with EC alone. Interestingly, there was a trend between CTC positivity and higher stage, grade 3 disease, type II EC, shorter time to recurrence and reduced survival, though the interpretation of these results is limited by the small sample size and small number of events in follow-up.
There were a number of parameters that did not correlate with CTC positivity, including MMI, LVSI and cervical invasion, all being traditional parameters currently used to guide primary treatment post-surgery. There was also no correlation with haemoglobin nor albumin. This may reflect the limitations of these traditional pathologic factors in guiding primary treatment or limitations in CTC evaluation, and thus requires further validation in a larger patient cohort.
There was also a group of CTC+ patients, for whom there was concordance between CTC trends and their clinical outcome while on first-line chemotherapy for metastatic disease, suggesting that CTCs may act as a predictive biomarker in this setting. At the same time however, there were patients whose CTC follow-up did not correlate with their clinical course and there are a number of factors that need further investigation that might explain this.

EpCAM may be down-regulated in CTCs, as they undergo EMT to escape their local microenvironment and enter the circulation. In a comparison of EpCAM expression on CTCs versus FFPE tissue and other benign disease, EpCAM expression was demonstrated to be 10-fold lower on CTCs compared to primary or metastatic tissue [251], such that the Veridex CellSearch platform may not be sensitive enough to detect CTCs reliably and reproducibly.
When cancers and CTCs undergo EMT, they develop a more invasive phenotype with metastatic potential. EMT is a complex process whereby cells switch their epithelial

phenotype into a mesenchymal one, resulting in a down-regulation of epithelial markers including EpCAM and an up-regulation of mesenchymal markers [252]. Other factors that may drive EMT include EGFR, TGFB, transcription factors and small non-coding RNAs. The right combination of markers or other detection techniques may thus be required to better enumerate and characterise CTCs in EC.

3.8.2. Correlation of CTC Analysis and FFPE IHC for EpCAM and Stathmin

Correlation of EpCAM and stathmin IHC with CTC positivity demonstrated more intense staining intensity of FFPE tissue from CTC+ compared to CTC- patients, though the overall presence/absence of staining was similar between the two groups. Even with this difference in intensity in EpCAM and stathmin IHC, it was not possible to define a level of overexpression for predicting CTC results, either for CTC positivity, or for stathmin positivity.

Consistent with the published literature which reports moderate to intense EpCAM overexpression in 80-88% [95, 250] of EC specimens, overexpression at this level was reported in 86% of samples here, with more intense expression in EEC compared to NEEC. Stathmin overexpression on IHC is reported in between 27% and 57% of primary tumours [164, 165]. Stathmin overexpression on FFPE IHC was higher here at 80%, though a higher antibody concentration was used and scoring was based on the TIS versus the upper quartile of scoring reported elsewhere [164, 165]. Stathmin overexpression is also more common in NEEC [164] and interestingly, the majority of the tested blocks here were NEEC rather than EEC, which may also have contributed to the higher reported overexpression. Although other markers of the PI3K pathway were not analysed here, there is already evidence that stathmin is a more reliable marker of PI3K activation than AKT and p-AKT for example [164]. Thus for further study, the ability to detect stathmin on CTCs, the reproducibility of staining and evidence as a marker of PI3K activation [162] make stathmin an attractive choice of antibody.

Further investigation with a greater availability of tumour tissue in a larger patient cohort would be required to better a assess changes in these markers over time and correlation between IHC and CTC status. Longitudinal tissue review would also be of interest to better delineate the change in EpCAM and stathmin that may occur in EC between primary diagnosis and subsequent relapse. If the tissue analysis could then be mapped through changes in CTC molecular analysis, this could provide evidence for using CTCs both to predict response to standard chemotherapy as well as molecularly targeted agents for each patient.

3.8.3. Challenges in Defining a CTC Cut-off

Although the feasibility of CTC enumeration in EC was demonstrated here, the overall sample size was small for drawing conclusions on the prognostic and predictive nature of CTCs in EC. While it was possible to generate hypotheses, with the suggestion that CTCs may act as a prognostic and predictive biomarker in some EC patients, longitudinal data is required over a longer time period and from a larger patient cohort to validate these early findings.

While this study demonstrated a trend between higher CTC count and survival, this was too small a study to definitively address the question of a significant CTC cut-off. The CTC cut-off of 5 as a prognostic factor was first defined in breast cancer [100] by correlating CTC thresholds with PFS in a training set of 102 patients. Median PFS and a difference between slow and rapid progression was best defined by 5 CTCs per 7.5ml blood. Further evaluation in a validation set of 75 patients then confirmed that this CTC cut-off was an independent predictor of PFS and OS in metastatic breast cancer. In prostate cancer, the cut-off derived from the breast cancer setting was used and a correlation was then demonstrated between CTC count at baseline and during follow-up with prognosis in 231 patients [107]. The colorectal cancer CTC cut-off was defined based on correlation of CTC count with response on first imaging in a training set of 109 patients and then taken forward to a validation set of 321 patients, where its prognostic and predictive utility was demonstrated [111]. Thus, although there is a trend between higher CTC count and poorer survival in EC, greater patient numbers are required to validate this finding.

3.8.4. Key Findings and Future Directions

Key findings for the analysis of CTCs in EC are outlined in Figure 3.7. CTCs in EC are worthy of further study in view of their prognostic and predictive potential. This may involve further evaluation with the current Veridex CellSearch methodology as well as expanding assessment into novel CTC markers and detection techniques, which will be discussed further in Chapter 6. Although, there were limitations with sample size here, the data was hypothesis generating in terms of the clinical and pathologic associations, as well as correlation with patient outcomes over time. Further longitudinal molecular analysis on blood and tissue samples in this setting is also important for further studies and again, with these early signs of its feasibility, expansion of this type of study with incorporation of novel techniques would be of great interest.

Figure 3.7: Key Findings of CTC enumeration in EC

- Detection of CTCs in EC patients and molecular analysis with Stathmin performed on the CellSearch platform is feasible
- CTC+ in EC is associated with stage IV and recurrent metastatic disease, type II EC, higher grade and reduced survival
- CTC count correlates with clinical course in some EC patients and could be used to predict outcome on first-line chemotherapy
- A CTC cut-off for predictive and prognostic purposes is still to be defined
- EpCAM IHC does not reliably predict CTC detection, though correlation with metastatic tissue is required
- Stathmin IHC overexpression was associated with CTC positivity

CHAPTER 4: Results from DNA Extraction, Modification and Quality Assessment

4.1. Introduction

Historically, oncology drug development has been hampered by clinical trials being performed in unselected patient groups, even when testing newer targeted agents. Improved understanding of tumour biology and incorporation of biomarkers into as early a setting as possible however can lead to more rapid drug approval and improved patient outcomes. Compared to other solid tumours, there has been little correlation between clinical outcomes, genetic aberrations and drug development in EC. For example, despite substantial literature on PI3K pathway aberrations, studies of temsirolimus and everolimus were developed in an unselected patient population with limited biomarker assessment and no current predictive markers to guide treatment [81]. Similarly, although FGFR mutations have been well documented in EC, studies to correlate mutation status and response are only recently underway [253]. The molecular basis for distinctions between and within Type I and Type II EC are only partly characterised and the transition of normal endometrium to AEH and EC is poorly understood. Thus, further research into genetic and epigenetic abnormalities and profiling on FFPE endometrial tissue may be better able to identify genes and molecular pathways that drive carcinogenesis and move treatment from a "one-size-fits-all" strategy to a targeted approach that improves clinical outcomes for this disease.

There are a number of methods used for epigenetic analysis as outlined in section 1.4.2.3, one of which has been developed and validated within the UCL Cancer Institute for use in both FF and FFPE tissue. This method utilises bisulfite-ligated DNA to assess methylation signatures in FFPE tissue on the Illumina Human Methylation 450K BeadChip array [231][232]. This platform promoted that FFPE tissue could be used, but there was limited published data on this until Thirlwell et al [232] published their protocol to analyse FFPE-derived DNA on the Infinium platform using the HumanMethylation27 BeadChip. The Illumina Infinium assay requires the template DNA to be >1kb in length for the amplification to work. As such, the protocol included a DNA ligation step to increase DNA fragment length for subsequent whole-genome amplification and analysis. This method was demonstrated to be feasible with ligated FFPE samples performing as well as FF samples on the 27K array, including FFPE DNA from tissue up to 10 years of age [232].

The Illumina 450K array can also be used to detect CNV, with comparable sensitivity to SNP platforms [254], including the HumanCytoSNP-12 BeadChip and publicly available Affymetrix SNP6.0 segmented data. To date, interrogation of methylation and CNV data have been performed on separate array platforms which increases cost and the amount of tissue

required. Thus, developing methodology that utilises methylation arrays to detect CN and methylation changes in a single experiment may decrease both these components. In view of the limited tissue available in some of the endometrial samples, this was particularly relevant.

Interpretation of the methylation and CNV data utilised the validated ChAMP pipeline on the R console [245]. This has been outlined in Chapter 2, but for ease of reference, is summarised again here. The ChAMP package is a pipeline that integrates available 450K analysis methods with several analysis options, including CNV. It uses the data import, quality control and SWAN (subset-quantile within array normalisation) functions as part of the *minfi* package, as well as PBC (Peak Based Correction), BMIQ normalization, SVD (singular valve decomposition) and ComBat methods to select the best performing samples, correct for probe bias and correct for batch effects respectively.

In this study, DNA methylation and CNV analysis on FFPE endometrial tissue was performed, with FF tissue as a control, using the Illumina 450K array. Samples for comparison included normal endometrium, AEH and EEC. EEC was chosen in this study rather than both EEC and NEEC to maintain as molecularly homogenous a sample group as possible for testing.

Figure 4.1 outlines the components for assessing sample selection, DNA quality and preparation prior to epigenetic and CNV analysis. These processes are detailed in this chapter to demonstrate methodological quality control and validation.

Figure 4.1: Key Components of Sample Management for Epigenetic and CNV analysis

- Accurate sample selection for tissue type
- Optimisation of DNA extraction
- Accurate DNA concentration analysis
- Assessment of DNA quality
- Assessment of bisulfite conversion
- Quality control on the ChAMP pipeline on the R console

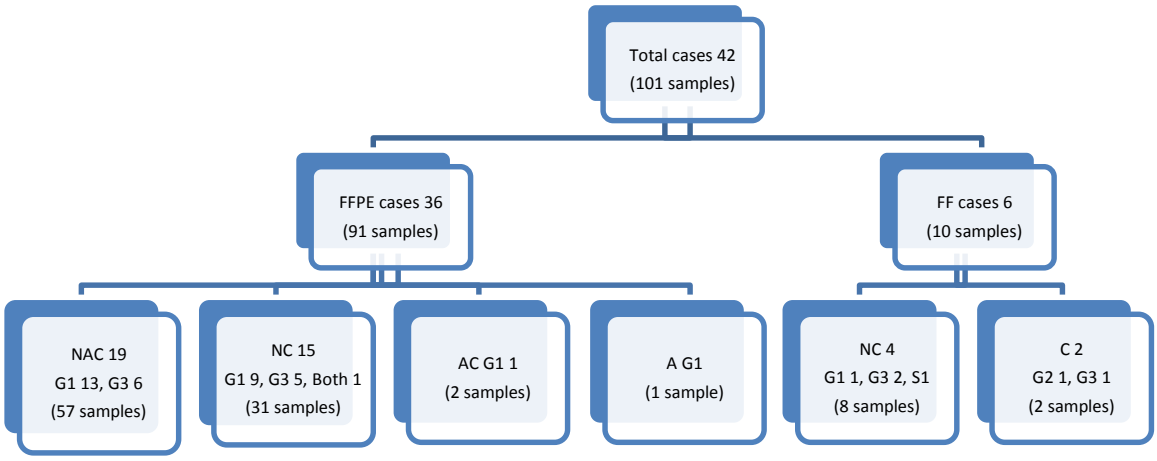
4.2. Sample Selection

From the UCLH histopathology database search, 36 archival FFPE cases were identified from between 2009 and 2013 and 6 fresh frozen cases between 2013 and 2014.

Selected cases had the required tissue type (normal +/- AEH +/- EEC) and were deemed amenable to macrodissection based on review with a Gynae-oncology consultant histopathologist.

A total of 42 patient cases identified equated to dissection of 91 FFPE normal, AEH or EEC samples and 10 FF normal or EEC samples. The number of samples with the dissected tissue type and grade is outlined in Figure 4.2. Note that the 91 FFPE samples does not include duplicates that may have been collected if there was adequate tumour tissue or if analysis was repeated. The FF cases that demonstrated grade 2 and sarcomatous disease post-operatively were also not used in the final analysis through the ChAMP pipeline.

Figure 4.2: Samples available for DNA Extraction of Normal Endometrium, Atypical Endometrial Hyperplasia and Endometrioid Endometrial Cancer



N: normal endometrium, A: atypical endometrial hyperplasia, C: endometrial cancer.
G1: grade 1, G3: grade 3, S: sarcoma

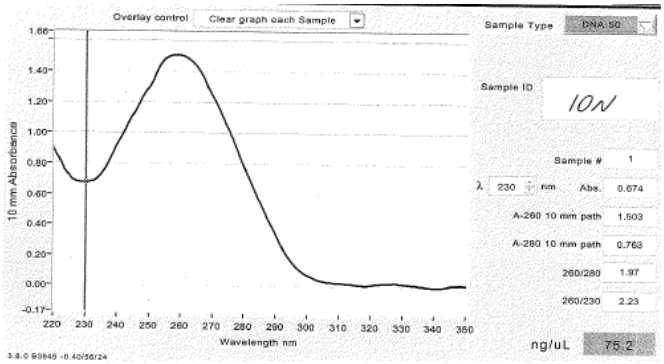
4.3. DNA Extraction and Concentration

The first 10 samples dissected were analysed on the Illumina 450K array prior to the dissection of any further samples for review of methodological quality control. These 10 samples consisted of normal, AEH and EC tissue from 6 patient cases and DNA concentrations were detected on the NanoDrop spectrometer alone using wavelength assessment as a quality guide, as demonstrated in Figure 4.3. DNA modification and bisulfite conversion QC were then performed and the samples were run on the Illumina 450K array.

Shortly after this array was performed, two additional steps were added to the DNA analysis. Firstly, a Qubit fluorometer was introduced in the laboratory and was used for subsequent

DNA concentration alongside NanoDrop spectrometry, and secondly the Illumina QC step to assess DNA quality prior to performing bisulfite conversion, was also introduced.
The concentrations for the first 10 samples were then retrospectively assessed using the Qubit fluorometer and subsequent samples were analysed using both the NanoDrop spectrometer and Qubit fluorometer. Wavelength values were also collected from NanoDrop to assist with quality assessment. The concentrations and wavelengths are recorded in the Appendix (Table 4A for these first 10 samples, Table 4B for the remaining FFPE samples and Table 4C for the FF samples). Overall, the concentrations on the Qubit fluorometer tended to be lower than that recorded on NanoDrop spectrometry. As such, the Qubit derived DNA concentrations were used for the calculations of the volumes required for further DNA modification.
The FFPE specimens were all extracted with needle dissection following microscopic review with a Gynae-oncology consultant histopathologist and the FF specimens were classified macroscopically by the same histopathologist at the time of specimen delivery and prior to freezing.

Figure 4.3: NanoDrop Spectrometry Assessment of Concentration and Wavelength



Note the 260/280 ratio of 1.97 and 260/230 ratio of 2.23, along with the wavelength shape peaking at 260-270nm were consistent with adequate DNA quality as per NanoDrop assessment

4.4. DNA Quality Analysis

The Illumina Infinium HD FFPE QC assay was designed to determine DNA quality and to predict if the DNA samples were candidates for FFPE restoration and likely to yield valid results on the subsequent Illumina 450K array analysis.

This was performed in retrospect for the first 10 samples to correlate with the array results and prospectively for the remaining samples. All specimens were run in triplicate to demonstrate intra-sample reproducibility on the 7300 real time-polymerase chain reaction (RT-PCR) system. The results of the QC analysis and Delta Cq (average quantification cycle for each sample minus the average quantification cycle for the control) for FFPE samples, based on the raw data from the RT-PCR analysis, are illustrated in Table 4.1. QC analysis was not required for FF specimens as the DNA was not degraded with formalin during the preservation process.

The protocol for the Illumina Infinium HD FFPE QC [255], as outlined in Chapter 2.2.4.1.3, advises that samples with a Delta Cq value <5 can be selected for FFPE restoration and array analysis, but also recommends the use of the Illumina restoration kit for further modification. A higher Delta Cq cut-off up to 11 was used here for a number of reasons. Firstly, the Delta Cq cut-off recommended by Illumina was based on their own restoration protocol while an alternate method [232] for DNA modification was utilised here with a tissue type that is less well studied and for which a different Delta Cq cut-off may apply. Secondly, there was a moderate amount of variation in Delta Cq values generated here, such that some samples with a higher Delta Cq were analysed further to better correlate the Delta Cq with this methodology. Thirdly, a less stringent approach was deemed worthwhile to balance the practicalities of limited sample number with a need to better define the optimum cut-off in this tissue type.

Table 4.1: Illumina Infinium HD FFPE QC results for all samples

Specimen	Delta Cq	**Specimen**	Delta Cq	**Specimen**	Delta Cq
1A[1]	5.27*	**18N**	10.07*	**29C**	8.49
2N[1]	-0.11*	**18A**	8.24*	**29C2**	8.94
2C[1]	2.45*	**18C**	8.39*	**30N**	4.74*
5N[1]	3.12*	**19N**	9.7*	**30A**	3.03*
5A[1]	2.90*	**19A**	9.4*	**30C**	2.86*
5C[1]	3.90*	**19C**	8.9*	**31N**	3.45*
6C[1]	4.59*	**20N**	4.42*	**31C1**	3.59*
7C[1]	5.66*	**20C**	4.71*	**31C3**	5.24*
9N[1]	8.45*	**21N**	2.83*	**32N**	2.04*
9C[1]	6.27*	**21A**	3.21*	**32C**	0.49*
3N	5.73*	**21C**	5.96*	**33N**	1.29*
4A	2.37*	**22N**	1.48*	**33C**	2.21*
6N	10.19*	**22C**	1.5*	**34N**	3.23*

6A	6.23*	**23N**	13.5	**34A**	2.43*
10N	6.15*	**23A**	14.3	**34C**	5.12*
10A	5.98*	**23C**	15	**35N**	5.38*
10C	3.47*	**24N**	0.73*	**35C**	6.36*
12N	6.16*	**24A**	1.26*	**36N**	4.95*
12C	9.78*	**24C**	0.53*	**36C**	4.31*
13N	2.6*	**25N**	10.5	**1N²**	9.18
13A	0.84*	**25C**	9.32	**1A²**	8.34
13C	3.14*	**26N**	10.19	**1C²**	9.66
14N	6.37*	**26C**	9.65	**3N²**	5.49*
14C	6.57*	**26C2**	10.45	**3A²**	6.02*
15N	6.48*	**27N**	5.87*	**3C²**	7.09*
15C	5.57*	**27A**	5.18*	**7N²**	9.42*
16N	7.18*	**27C**	5.20*	**7A²**	7.02*
16A	9.94*	**28N**	1.41*	**7C²**	8.73*
16C	9.08*	**28A**	0.18*	**8N²**	8.99*
17N	9.14*	**28C**	0*	**8A²**	8.7*
17A	7*	**29N**	9.06	**8C²**	11.3*
17C	8.58*	**29N2**	12.11	**11A²**	7.3*
				11C²	6.92*

[1]QC for first 10 samples was assessed in retrospect
[2]Samples that were re-extracted due to initial QC concerns or limitations in extraction
*Samples that underwent further modification for the arrays.

The consistency of the Delta Cq within a case as well as the tissue types available for analysis were also taken into consideration. If all specimens within a given patient case had a high Delta Cq, the case was not used; however, if the Delta Cq value was lower for one of the tissue types or if all 3 tissue types were available for potential comparison, than the entire case was used for analysis. For example, for cases 23, 25, 26 and 29, all the Delta Cq values were between 8 - 15 and 4 of these specimens only had normal and cancer tissue available for comparison, so these specimens were not analysed on the Illumina 450K array.
However, for both specimens 9 and 12, one of the Delta Cq values for the specimens was measured as 6, so deemed worthwhile to analyse the entire case. For specimens 7, 8, 16, 17, 18 and 19, there was a range in Delta Cq values, but as all 3 tissue types were available for comparison, it was again deemed worthwhile to analyse these cases.
Overall, 41 from 97 specimens (42%) had a Delta Cq <5, 49 specimens (51%) had a Delta Cq value ≤5, 38 specimens (39%) had a Delta Cq value between 6-9 and 10 specimens (10%) had a Delta Cq value ≥10.

Analysis was based on both raw numerical data, as in Table 4.1, as well as a visual assessment of the amplification plots produced on the RT-PCR System as illustrated in Figures 4.4 and 4.5. When interpreting the RT-PCR plots, it is important to be aware of the baseline, threshold and threshold cycle. The baseline of the reaction refers to the signal level during cycles 3-15,

where there is little change in the fluorescence signal and background 'noise' of the reaction is recorded. The threshold of the RT-PCR is the level of signal that reflects a statistically significant increase over the baseline signal (horizontal green line in the figures). The threshold cycle (Ct) is the cycle number at which the fluorescent signal of the reaction crosses the threshold, related to the amount of DNA present.

Figure 4.4: Amplification plot demonstrating the Difference between the QC Template and H_20 samples

Note that QC Template (left, red arrow) amplifies at an earlier cycle number as a DNA control. The H_20 sample (right, blue arrow) amplifies at a later cycle. Samples with good quality DNA should amplify close to the QC Template Reagent. All samples are run in triplicate and the triplicate readings for each sample should amplify at a similar cycle, as demonstrated for the QC template and H_20 samples below. The purple arrow indicates the low-level signal at the baseline of the PCR reaction.

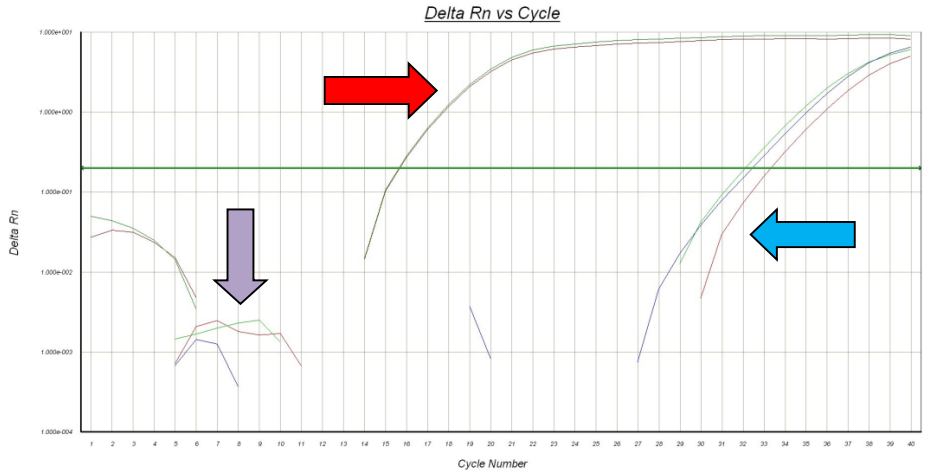
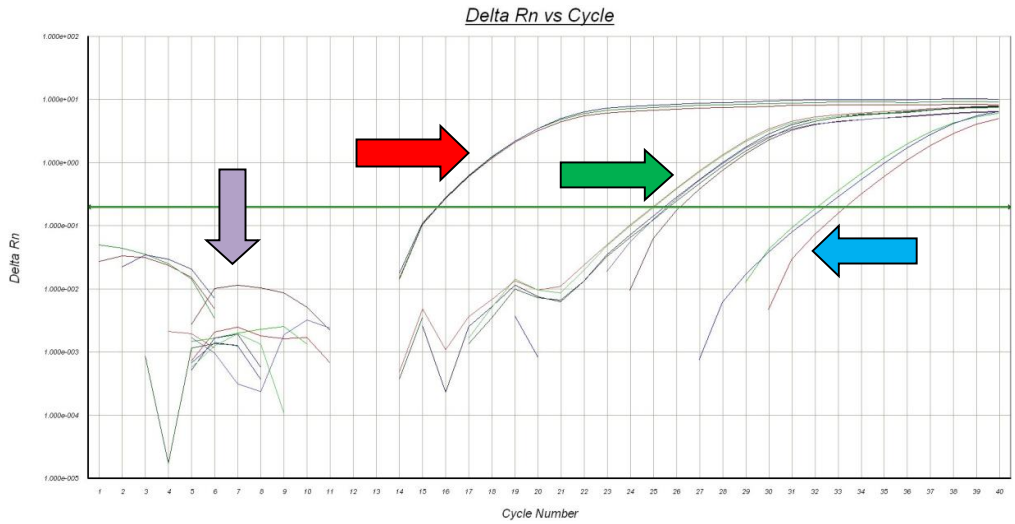


Figure 4.5: Amplification plot of QC Template (left, red arrow), H_20 (right, blue arrow) and 2 samples, 26N and 26C (centre, green arrow)

Note that the samples with good quality DNA will lie closer to the QC Template plot. The samples here are 26N and 26C, which did not demonstrate good quality DNA on this QC step and were not used for further analysis. All specimens were run in triplicate and demonstrate intra-specimen concordance.



4.5. DNA Bisulfite Conversion Analysis

This step determines the success or failure of the DNA modification and BIS conversion, through the use of RT-PCR and primer sets specific for BIS converted (actin positive) and unconverted DNA (actin negative).

Figures 4.6 and 4.7 demonstrate examples of the amplification plots generated on the RT-PCR system that are visually analysed to assess success of bisulfite conversion.

Table 4.2 lists the calculated bisulfite conversion percentage for each sample, based on the raw data from the RT-PCR System analysis.

Figure 4.6: Amplification plot of a Bisulfite Converted Lab Control with Actin positive (left, red arrow) and Actin negative (right, blue arrow) primers
Note the BIS converted DNA detected in the actin positive sample amplifies earlier than the Actin negative sample which detects unconverted DNA. The baseline low-level signal of the RT-PCR reaction is also shown (purple arrow).

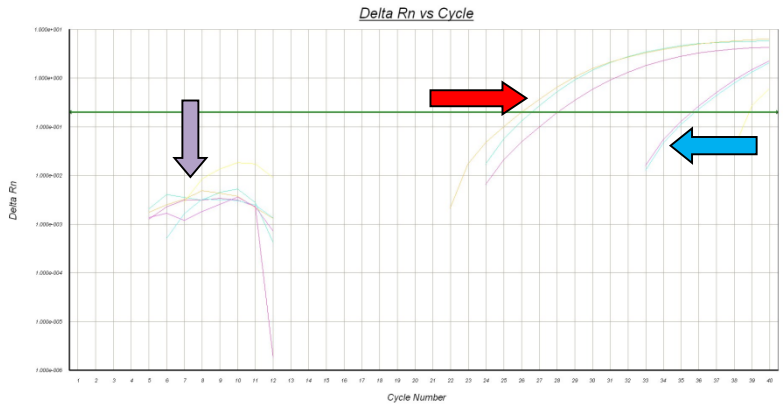


Figure 4.7: Amplification plot demonstrating successful Bisulfite Conversion of sample 24C
All samples are run in triplicate and these curves should amplify within 2-3 cycles of each other, as illustrated. The actin positive sample is on the right (red arrow) and amplifies earlier than the actin negative sample (blue arrow) when the DNA modification has been successful, as it detects the bisulfite converted DNA. The 24C plot lies close to the actin positive sample.

Table 4.2: Bisulfite Conversion Analysis for all FFPE and fresh frozen samples

Sample	Conversion %	**Sample**	Conversion %	**Sample**	Conversion %
1A	98.58	**18N**	52.38*	**34N**	95.79
2N	99.9	**18A**	87.13*	**34A**	96.30
2C	99.23	**18C**	69.81*	**34C**	91.05
5N	99.84	**19N**	98.52	**35N**	96.53
5A	99.58	**19A**	95.02	**35C**	95.25
5C	99.66	**19C**	98.88	**36N**	98.22
6C	99.3	**20N**	99.46	**36C**	99.62
7C	92.78	**20C**	99.79	**3N**	97.52
9N	56.97	**21N**	99.83	**3A**	96.17
9C	85.73	**21A**	99.88	**3C**	95.77
3N	94.42	**21C**	99.65	**7N**	62.10
4A	99.99	**22N**	99.89	**7A**	55.58*
6N	21.29	**22C**	99.98	**7C**	30.49
6A	99.93	**24N**	99.85	**8N**	75.41*
10N	92.15	**24A**	99.93	**8A**	63.77*
10A	98.81	**24C**	99.89	**8C**	38.84*
10C	-	**27N**	99.71	**11A**	82.64
12N	99.95	**27A**	99.39	**11C**	98.34
12C	98.40	**27C**	99.48	**37N**	89.98
13N	99.87	**28N**	99.77	**37C**	99.87
13A	99.81	**28A**	99.98	**38C**	97.36
13C	98.97	**28C**	99.92	**39C**	99.73
14N	98.95	**30N**	99.65	**40N**	98.5
14C	99.45	**30A**	99.62	**40C**	96.95
15N	99.84	**30C**	99.45	**41N**	96.9
15C	99.96	**31N**	98.28	**41C**	99.49
16N	86.44*	**31C1**	99.82	**41N2**	95.47
16A	98.43*	**31C3**	99.45	**41C2**	99.88
16C	97.37*	**32N**	99.58	**42N**	99.92
17N	71.39	**32C**	99.99	**42C**	98.95
17A	98.79	**33N**	99.80		
17C	83.87	**33C**	99.85		

* BIS QC run performed 1 week after conversion due to RT-PCR machine break-down. The BIS conversion for these samples was lower overall than for those analysed immediately.
- 10C was not assessable as the actin negative triplicate was not analysable by the QT-PCR.
Illumina Delta Cq>5

A cut-off of 98% is generally used to demonstrate successful BIS conversion. For the above specimens:

- 60 from 94 specimens (64%) showed BIS conversion ≥98%
- 12 specimens (13%) showed BIS conversion of 95-97% (16C, 19A, 34N, 34A, 35N, 35C, 3N, 3A, 3C, 38C, 40C, 41N)
- 4 specimens (4%) showed BIS conversion 90-94% (7C, 3N, 10N, 34C)

- 17 specimens (18%) showed BIS conversion less than 90% (9N, 9C, 6N, 16N, 17N, 17C, 18N, 18A, 18C, 7N, 7A, 7C, 8N, 8A, 8C, 11A, 37N). The correlation between BIS conversion percentage and Illumina Delta Cq is summarised in Table 4.6 below
- 1 specimen (1%) could not be analysed (10C)

Table 4.3: Correlation of BIS conversion and Illumina Cq values

BIS conversion percentage	Mean Illumina Cq value	Range of Illumina Cq values
>98%	4.0	-0.11 - 9.94
95-98%	6.4	2.43 - 9.4
90-94%	5.7	5.12 - 6.15
<90%	8.6	6.27 - 11.3

As demonstrated in Table 4.3, a higher Delta Cq value on the Illumina QC analysis correlated overall with a lower BIS conversion percentage, indicating potentially poorer DNA integrity that may not have been amenable to further modification. Note that the range on those specimens with a BIS conversion of greater than 95% however is quite broad and elevated Delta Cq does not always predict BIS conversion failure. These results are further correlated with the ChAMP analysis in section 4.6.

4.6. Sample Analysis from the ChAMP pipeline

The ChAMP pipeline [245] incorporates a number of steps for quality control and normalisation of the samples and their data. It can compare 2 variables at a time when generating data on MVPs and DMRs. As there were 3 tissue types for analysis here, the data was analysed through paired comparisons of normal endometrium, AEH and EEC tissue including all specimens initially and then based on grade and tissue type. This allowed selection of the best quality samples for data analysis and generated differential methylation data between the two comparator groups. A simultaneous comparison between all 3 tissue types could be performed to demonstrate clustering based on differential methylation on heatmaps and dendrograms, but did not generate specific information on DMRs and MVPs.

The first three comparator groups used all the modified DNA specimens and were between normal endometrium and endometrioid endometrial cancer (NvC), normal endometrium and atypical endometrial hyperplasia (NvA) and atypical endometrial hyperplasia and endometrioid endometrial cancer (AvC). Subsequent comparisons subdivided these main groups based on grade (1 or 3) and tissue type (normal, atypical or cancer). Comparison was also made just on those cases for which all 3 tissue types were available (referred to as matched samples and including cases 13, 21, 24, 27, 28, 30 and 34).

As detailed in the methods section 2.2.6.1, the ChAMP quality control and normalisation involves the following:

- filtering out probes with non-significant detection (p-value >0.01) in more than one sample and excluding samples with a fraction of failed probes >5-10%. Table 4.4 outlines the number of samples that were filtered in this step based on the three comparator groups outlined above.
- intra-array normalisation using BMIQ to adjust the data for the Illumina 450K array which uses both Infinium I and II probes. Infinium type II probes are generally less accurate and reproducible than the Infinium I probes [246], thus BMIQ normalisation is performed to overcome this bias. The pre- and post- normalised plots for NvC, NvA, AvC and NvAvC comparisons are presented in Figure 4.8.

  This step also generates a normalised multidimensional scaling (MDS) plot and sample cluster plot that gives a preliminary assessment of the 1000 most variable probes between the two groups, outlined in Figure 4.9.

  The MDS plot, also known as a Principal Coordinates Analysis, assesses the level of similarity of cases in a dataset. The axes values are not absolute but are a representation of the spread of the samples.
- Singular Value Decomposition (SVD) to control for batch effects associated with technical aspects of the array. The heatmaps generated at this step are presented in Figure 4.10. The ComBat normalization method then corrects for batch variation but requires at least 2 specimens per array per comparison.

After these steps, the data is then analysed to detect segmentation of MVPs and DMRs, as well as analysing CNV.

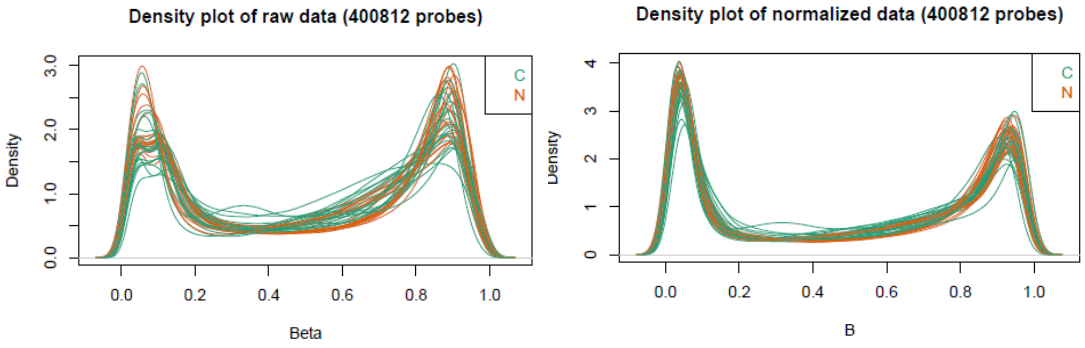
Table 4.4: Comparison of the total unfiltered evaluable samples and significant probes with the filtered samples and significant probes

Comparator groups	Total unfiltered samples & significant probes	Filtered samples and significant probes
N v C	69 & 12 750	39 & 400 812
N v A	53 & 24 737	26 & 424 168
A v C	57 & 14 422	29 & 405 496
N v A v C	-	47 & 400 515

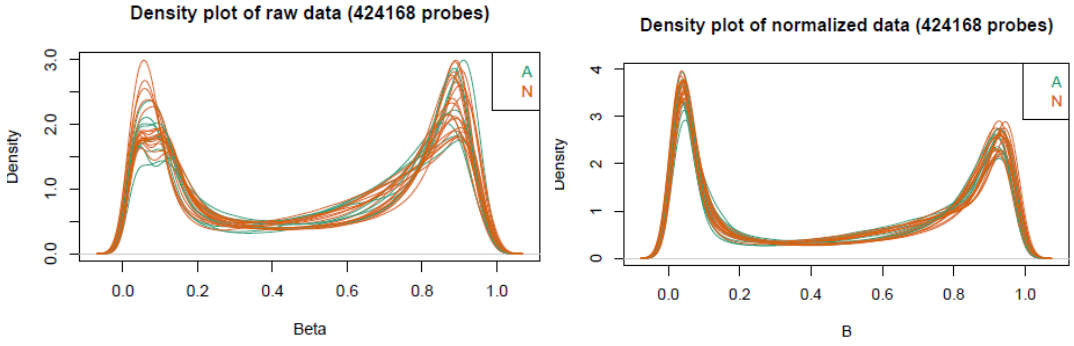
Figure 4.8: Pre and post normalisation Density plots for
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)
D: Normal Endometrium v Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (NvAvC)

Note that for all 4 plots showing the raw data on the left, there is a wider variation in density between all the samples related to the Infinium type I and II assays. This variation is reduced post normalisation within ChAMP, as demonstrated on the plots on the right side.

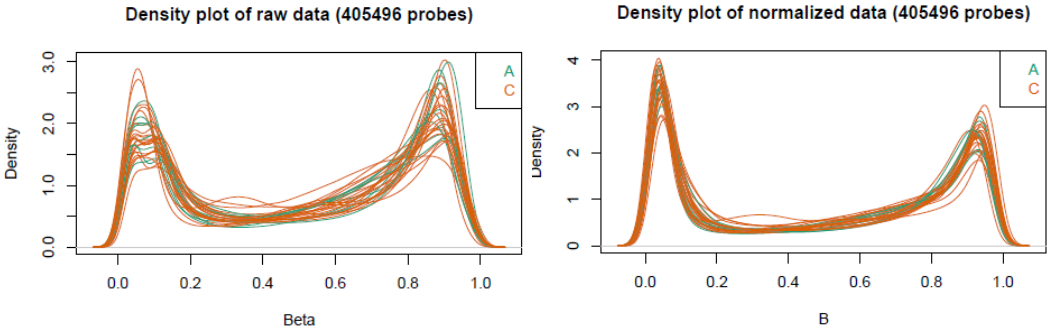
A:



B:



C:



D:

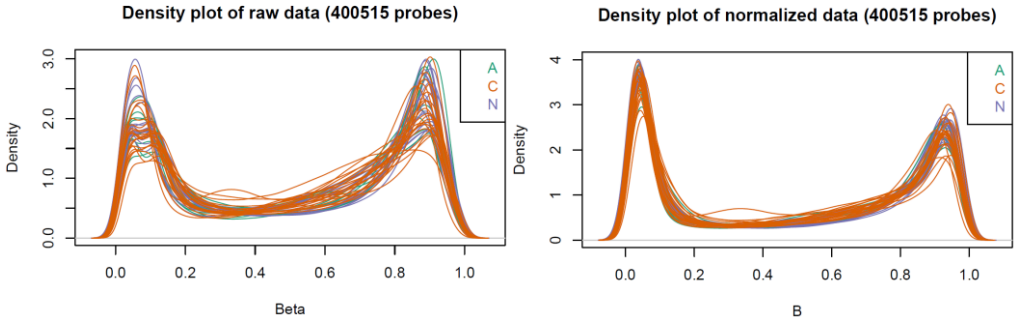
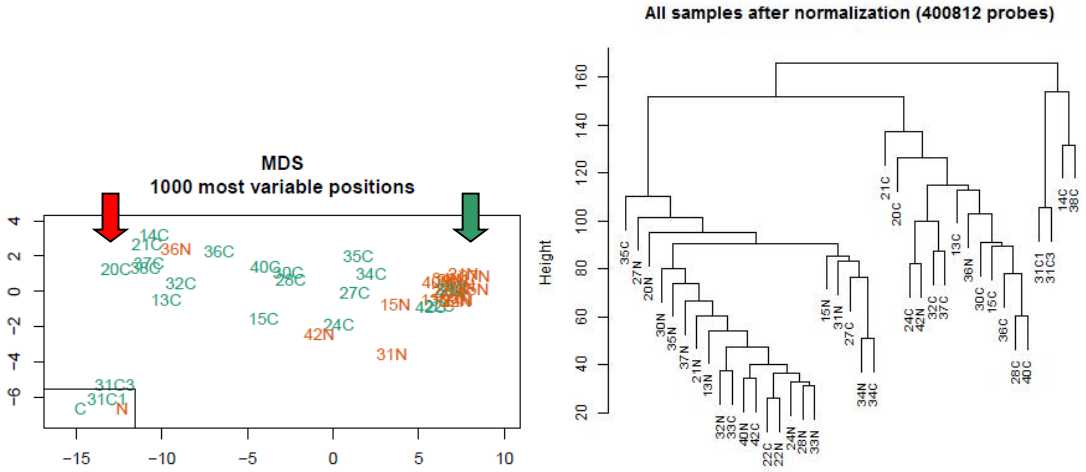
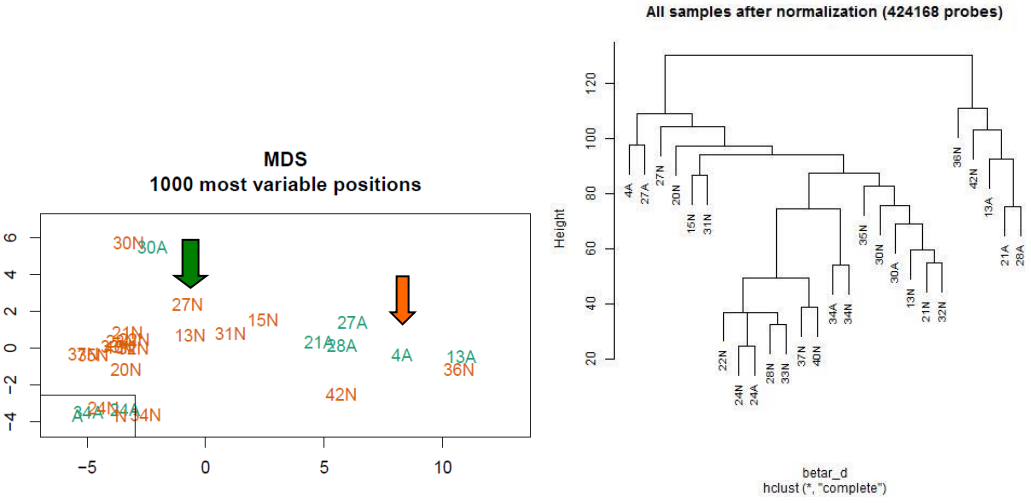


Figure 4.9: Normalised Multidimensional Scaling (MDS) and Cluster Plots for
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)
D: Normal Endometrium v Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (NvAvC)
The MDS plot (left) allows visualisation of the similarity of the samples based on the top 1000 most variable probes. The cluster plot (right) is a second way to visualise the similarity of samples based on all probes.

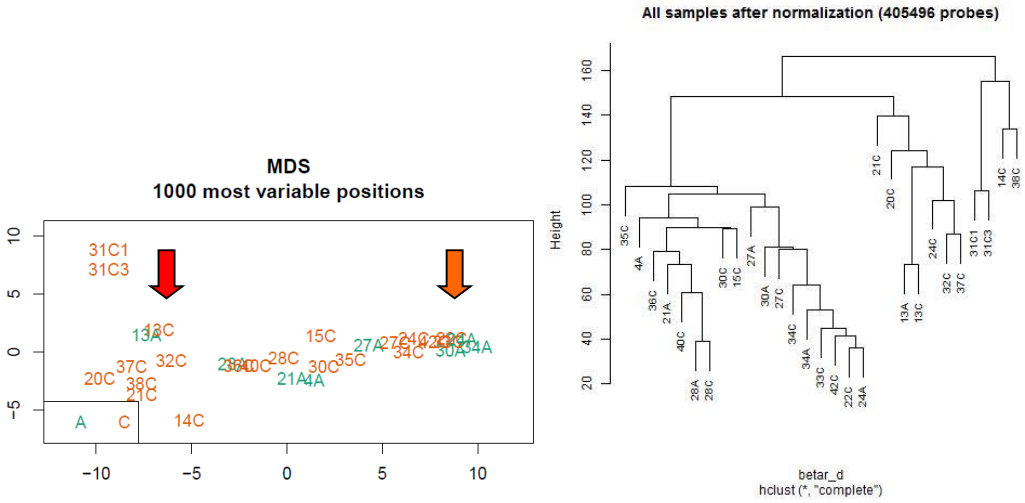
A: NvC MDS and cluster plots showing differential methylation between the 2 groups with normal samples on the right (green arrow) and cancer samples on the left (red arrow) on the MDS plot; and on the cluster plot, normal samples are clustered on the left and cancer samples on the right.



B: NvA MDS and cluster plots showing some differential methylation and some overlap between the 2 groups with normal samples in general on the left (green arrow) and atypical samples on the right (orange arrow) on both the MDS plot and on the cluster plot.



C: AvC MDS and cluster plots showing some differential methylation but greater overlap between the 2 groups, with atypical samples more on the right (orange arrow) and cancer samples on the left (red arrow) on the MDS plot; and on the cluster plot, atypical samples are on the left and cancer samples are on the right.



D: Normal v Atypical v Cancer MDS and Cluster plots showing a trend towards differential methylation between the three groups with normal samples in general on the left (green arrow), atypical samples in the centre (orange arrow) and cancer samples towards the right (red arrow) of the MDS plot and similarly on the cluster plot.

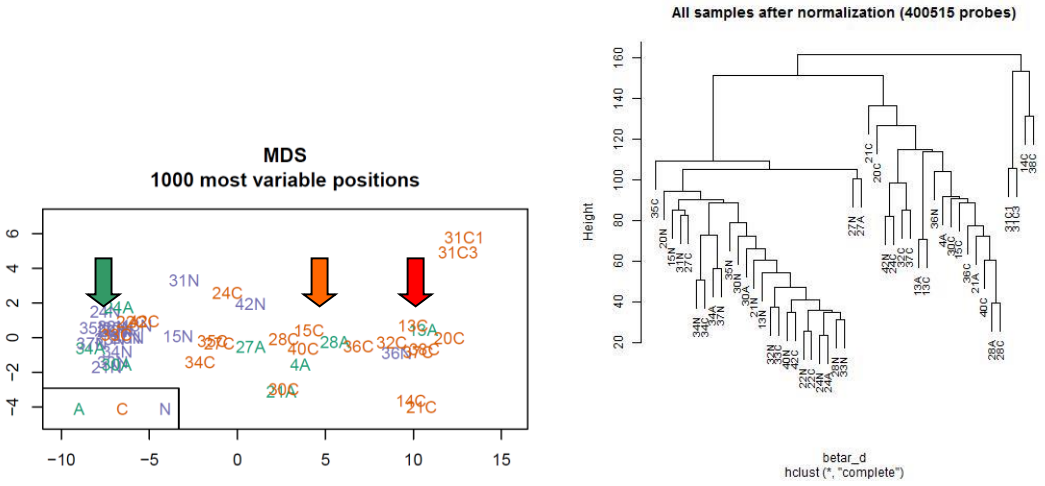
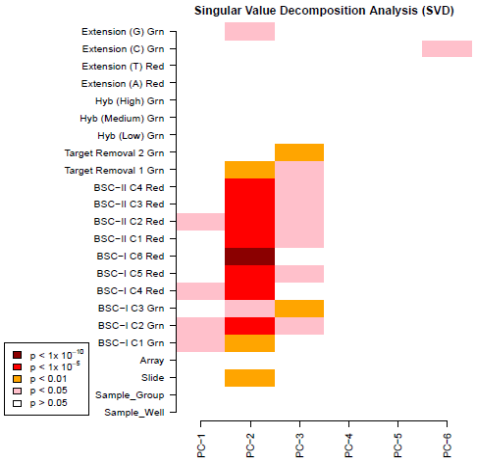
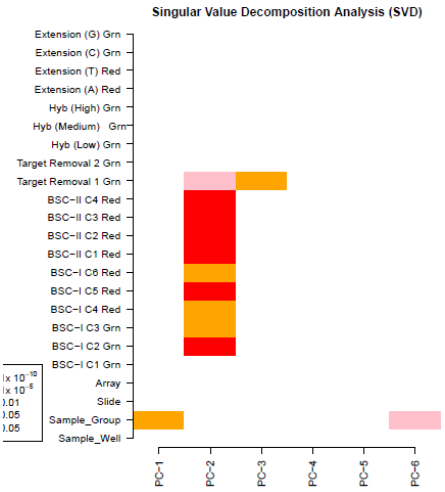


Figure 4.10: Singular Valve Decomposition (SVD) Heatmaps for
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)
D: Normal Endometrium v Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (NvAvC)

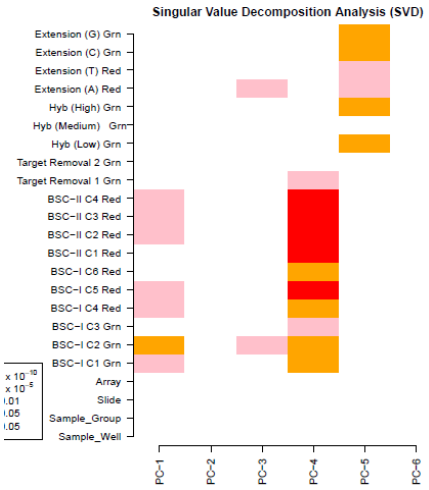
A:



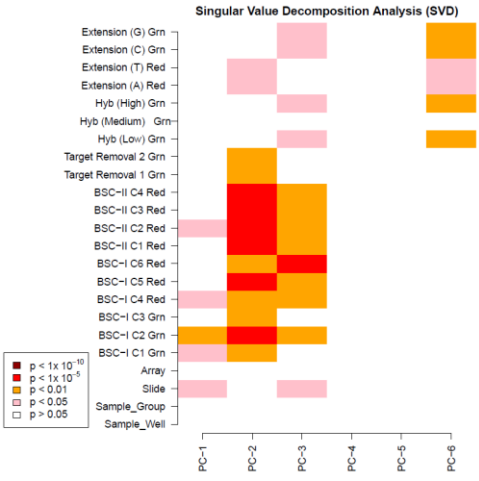
B:



C:



D:



Note that all of these comparisons have technical batch effects demonstrated, particularly in the bisulfite conversion parameters. The ComBat normalisation step corrects for batch variation.

From the MDS and cluster plots, there is evidence of differential methylation between normal endometrium, AEH and EEC samples, based on the 1000 most variable probes.

An important control in this study was that the FF samples behaved similarly to the FFPE samples in terms of methylation status. Other than specimen 42 (discussed in section 4.8), this was evident on the MDS and cluster plots.

There were some samples however that did not demonstrate differential methylation between tissue types. This included 4 specimens (33C, 36N, 42N, 42C) in the NvC comparison, as well as 3 specimens (24A, 30A, 34A) in the NvA comparison. Note that specimen 2N had to be removed for the ComBat normalisation and subsequent ChAMP analysis, as it was the only specimen from the 1st array.

The analyses for NvC and NvA were then rerun on ChAMP with these further specimens excluded to re-assess clustering and the spectrum of differential methylation. The AvC comparison was also rerun with these samples excluded, but is not illustrated here as it showed a greater overlap between specimens in terms of methylation patterns and no significant DMRs nor MVPs.

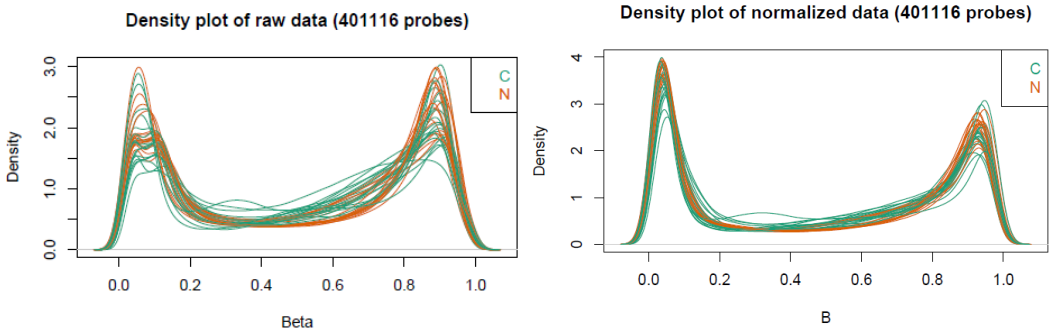
The pre and post normalisation density plots, MDS and cluster plots and SVD heatmaps for the 2 comparisons NvC and NvA are presented in Figure 4.11. These demonstrate more consistent sample clustering in the MDS and cluster plots, with a reduction in batch effects on the SVD plots. They also generate a higher number of MVPs and DMRs for the NvC and NvA comparisons though not for the AvC comparison, as illustrated in Table 4.5.

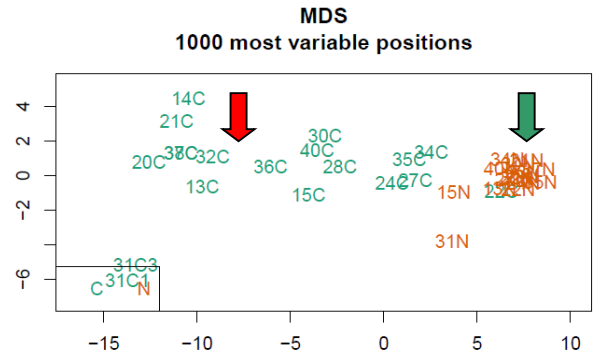
However, due to the already reduced sample size, further analysis of the MVPs and DMRs was based on the samples and probes after the initial ChAMP sample filtering, rather than with further specimens excluded, as outlined in the bolded column in Table 4.5. This is detailed further in section 4.8 of this chapter.

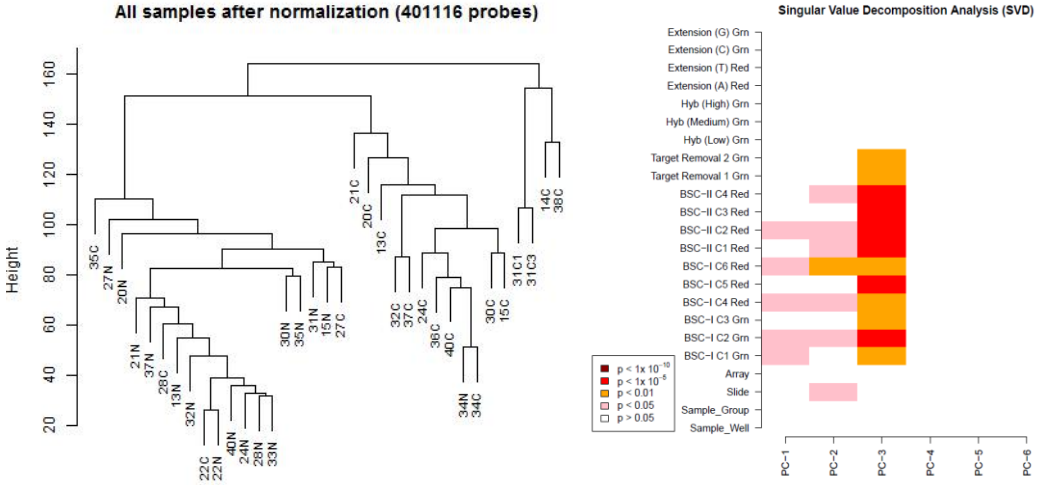
Figure 4.11: Pre and post normalisation density plots, Multidimensional scaling (MDS) and cluster plots and Singular Valve Decomposition (SVD) heatmaps for
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC) with samples excluded
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA) with samples excluded

A: NvC Comparison with 4 specimens excluded

On the density plots, there is wider variation on the raw data density plot on the left, while the normalised plot on the right has narrower density readings. The MDS plot shows the cancer specimens clustered on the left (red arrow) and normal specimens on the right (green arrow) based on the 1000 most variable probes, except for specimen 22C that lies with the normal specimens. The cluster plot shows the normal specimens towards the left and cancer specimens towards the right based on the 401 116 probes analysed. The SVD plot shows less batch effect than that when all NvC specimens were analysed.

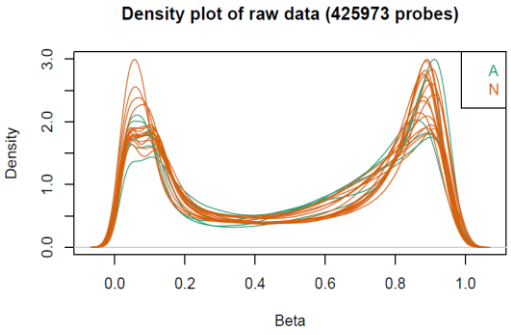


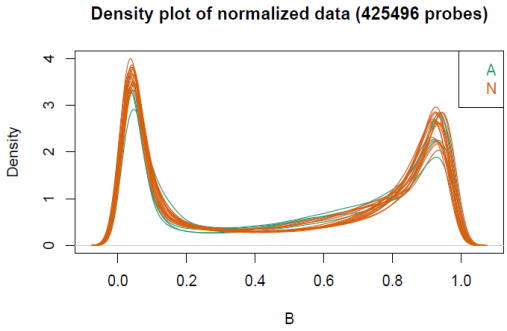


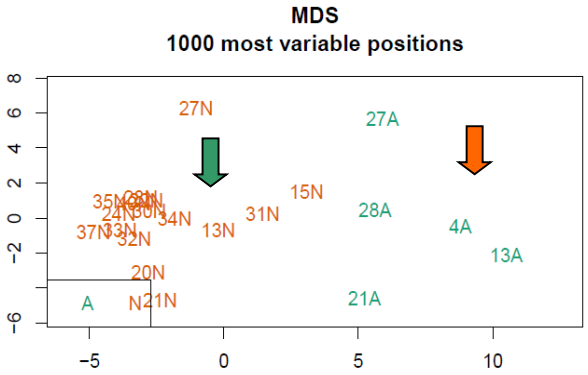


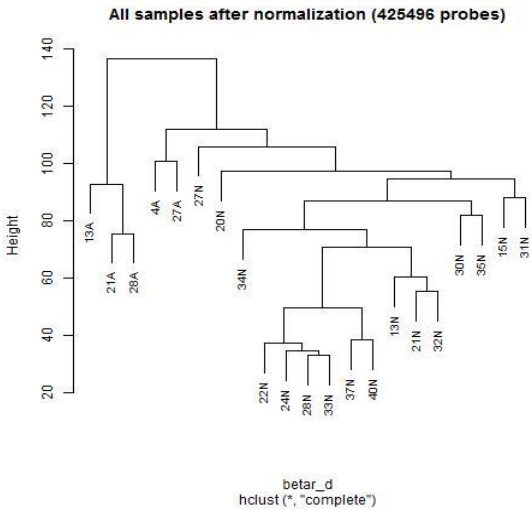
B: NvA Comparison with 3 specimens excluded

On the density plots, there is wider variation on the raw data density plot on the left, while the normalised plot on the right has narrower density readings. The MDS plot shows the normal specimens clustered on the left (green arrow) and atypical specimens on the right (orange arrow) based on the 1000 most variable probes. The cluster plot shows the normal specimens towards the right and atypical specimens towards the left based on the 425 496 probes analysed. The SVD plot shows less batch effect than that when all NvA specimens were analysed.









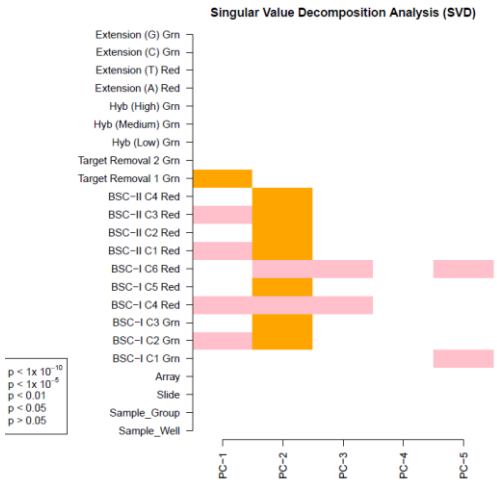


Table 4.5: Comparison of the total unfiltered evaluable samples and significant probes with the filtered samples and significant probes, including assessment of Methylation Variable Positions (MVPs) and Differentially Methylated Regions (DMRs)

Comparator groups	Total unfiltered samples & significant probes	**Filtered samples and significant probes**	**Significant MVPs and DMRs**	Revised sample number*	MVPs and DMRs
N v C N v A A v C	69 – 12 750 53 – 24 737 57 – 14 422	**39 – 400 812** **26 – 424 168** **29 – 405 496**	**23572, 706** **4, 0** **0, 0**	35 – 401 116 21 – 425 496 25 – 405 400	54841,1832 27908, 727 0,0

*removed 4 specimens (36N, 42N, 42C, 33C) from NvC; 5 specimens (30A, 24A, 34A, 36N 42N) from NvA; 5 specimens (30A, 24A, 34A, 33C, 42C) from AvC
The columns marked in bold indicate the sample number, probes, MVPs and DMRs that were used for further analysis.

4.7. Comparison of the Illumina QC and ChAMP QC Steps

To assess the utility of the Illumina Infinium QC step and bisulfite conversion percentage in relationship to the ChAMP analysis, the values for samples that passed and failed on ChAMP were compared, as outlined in Table 4.6. As expected, the FF specimens 37-42 all passed the ChAMP QC steps and are not included in the summary table.

In the majority of cases (68 from 82 specimens; 82%), there was a correlation between the Illumina QC value being ≤5 and the successful specimens on ChAMP. In addition, 11 specimens (13%) had an Illumina QC≤5 but failed on ChAMP, while a further 3 specimens (5%) had an Illumina QC>5 but were successful on ChAMP, though all of these were ≤6.

From the 11 specimens that had an Illumina QC≤5 but failed on ChAMP, 7 were analysed on the first array, which seemed to behave poorly compared to subsequent arrays. This may have been due to the use of NanoDrop derived concentrations for subsequent modification and array calculations, such that less DNA was actually available for analysis. As such, a Delta Cq cut-off ≤5, rather than the Illumina cut-off <5 was reviewed further here, to define a broader number of evaluable samples.

Samples that had poor percentage bisulfite conversion also tended to have a higher Illumina Delta Cq as illustrated in Table 4.6, as DNA with poor integrity may have been less amenable to subsequent modification. This was confirmed on correlation with the ChAMP QC.

For the 52 samples with BIS conversion ≥98%:

- 41 had Cq ≤5, of which 7 failed on ChAMP; 11 had Cq ≥6, of which 9 failed on ChAMP

For the 9 specimens with BIS conversion 95-97%:

- 4 had Cq≤5, of which 1 failed ChAMP; 5 had Cq ≥6, of which 4 failed ChAMP

For the 4 specimens with BIS conversion 90-94%:

- 3 had Cq≤5, of which 2 failed ChAMP; 1 had Cq ≥6 and failed on ChAMP

For the 16 FFPE specimens with BIS conversion <90%, all had a QC>5 and failed ChAMP QC.

Table 4.6: Correlation of the Illumina Infinium QC and Bisulfite Conversion values with Filtered samples through ChAMP

Sample	Illumina Cq/ BIS QC	ChAMP result	Sample	Illumina Cq/ BIS QC	ChAMP result	Sample	Illumina Cq/ BIS QC	ChAMP result
1A	5.27/98.58	F*	**16C**	9.08/97.37	F	**30A**	3.03/99.62	P
2N	-0.11/99.9	P	**17N**	9.14/71.39	F	**30C**	2.86/99.45	P
2C	2.45/99.23	F*	**17A**	7/98.79	F	**31N**	3.45/98.28	P
5N	3.12/99.84	F*	**17C**	8.58/83.87	F	**31C1**	3.59/99.82	P
5A	2.90/99.58	F*	**18N**	10.07/52.38	F	**31C3**	5.24/99.45	P
5C	3.90/99.66	F*	**18A**	8.24/87.31	F	**32N**	2.04/99.58	P
6C	4.59/99.3	F*	**18C**	8.39/69.81	F	**32C**	0.49/99.99	P
7C	5.66/92.78	F*	**19N**	9.7/98.52	F	**33N**	1.29/99.8	P
9N	8.45/56.97	F	**19A**	9.4/95.02	F	**33C**	2.21/99.85	P
9C	6.27/85.73	F	**19C**	8.9/98.88	F	**34N**	3.23/95.79	P
3N	5.73/94.42	F*	**20N**	4.42/99.46	P	**34A**	2.43/96.30	P
4A	2.37/99.99	P	**20C**	4.71/99.79	P	**34C**	5.12/91.05	P
6N	10.19/21.29	F	**21N**	2.83/99.83	P	**35N**	5.38/96.53	P
6A	6.23/99.93	F	**21A**	3.21/99.88	P	**35C**	6.36/95.25	P*
10N	6.15/92.15	F	**21C**	5.96/99.65	P	**36N**	4.95/98.22	P
10A	5.98/98.81	F*	**22N**	1.48/99.89	P	**36C**	4.31/99.62	P
10C	3.47/-	F*	**22C**	1.5/99.98	P	**3N**	5.49/97.52	F*
12N	6.16/99.95	F	**24N**	0.73/99.85	P	**3A**	6.02/96.17	F
12C	9.78/98.40	F	**24A**	1.26/99.93	P	**3C**	7.09/95.77	F
13N	2.6/99.87	P	**24C**	0.53/99.89	P	**7N**	9.42/62.1	F
13A	0.84/99.81	P	**27N**	5.87/99.71	P	**7A**	7.02/55.58	F
13C	3.14/98.97	P	**27A**	5.18/99.39	P	**7C**	8.73/30.49	F
14N	6.37/98.95	F	**27C**	5.20/99.48	P	**8N**	8.99/75.41	F
14C	6.57/99.45	P*	**28N**	1.41/99.77	P	**8A**	8.7/63.77	F
15N	6.48/99.84	P*	**28A**	0.18/99.98	P	**8C**	11.3/38.84	F
15C	5.57/99.84	P	**28C**	0/99.92	P	**11A**	7.3/82.64	F
16N	7.18/86.44	F	**30N**	4.74/99.65	P	**11C**	6.92/98.34	F
16A	9.94/98.43	F						

P: passed, F: failed

*samples where the Illumina recommended Delta Cq≤5 did not predict the ChAMP filtering

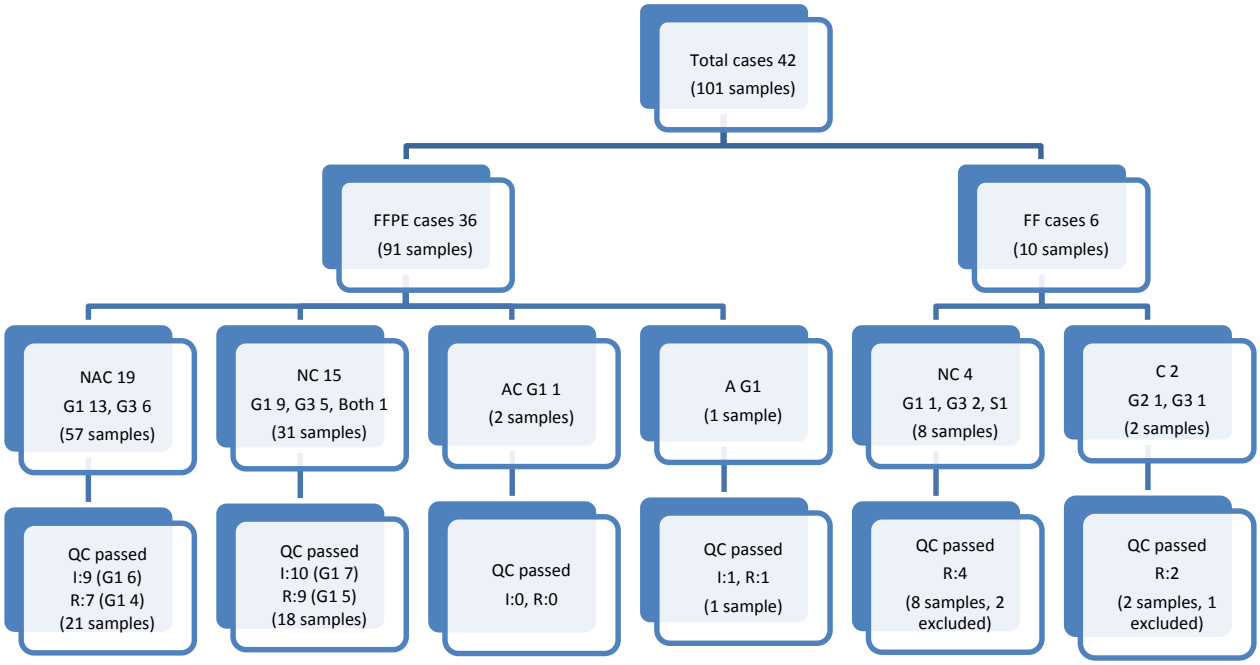
Using an Illumina Delta Cq≤5, 49 from 82 specimens (60%) met the criteria for DNA modification. From these, 38 specimens (46%) were evaluable for analysis on ChAMP. Using an Illumina Delta Cq≤6, 59 from 82 specimens (72%) met the criteria for DNA modification. From these, 41 specimens (50%) were evaluable for analysis on ChAMP.

Using a Cq cut-off ≤6 compared to ≤5 captures a greater number of cases and specimens that are then analysable on ChAMP, but also results in a greater proportion of sample failures. In terms of the number of specimens, using a Cq≤5 yielded 38 analysable specimens on ChAMP from 49 identified by the Illumina QC method (78% correlation). Using a Cq≤6 yielded 41 analysable specimens on ChAMP from 59 identified by the Illumina QC method (69%

correlation), with an extra 3 specimens that were analysable. However the subsequent failure rate also increased from 11 from 49 specimens (22%) to 18 from 59 specimens (31%). Thus, a Delta Cq≤5 and BIS conversion ≥95% seemed to correlate best with these samples for further analysis on ChAMP.

A summary of the patient cases and samples selected that then went on to pass each QC analysis is presented in Figure 4.12. In total from the ChAMP analysis, there were 47 samples of normal endometrium, AEH and EEC that passed the QC process. This included 40 FFPE samples (2N was excluded from the total of 41 analysable samples as it was the only sample from the 1st array and as such could not be analysed via the ComBat normalisation method) and 7 FF samples (sample 39 was excluded as it was grade 2 and 41 was excluded as it was sarcomatous).

Figure 4.12: Samples available for DNA extraction and that passed the Illumina and ChAMP QC Analyses



QC: quality control, I: Illumina QC, R: R console ChAMP QC
N: normal endometrium, A: atypical endometrial hyperplasia, C: endometrial cancer.
G1: grade 1, G3: grade 3, S: sarcoma

4.8. Review of Sample Selection

Based on the data in section 4.6, dissected specimens were re-reviewed to see if the specimens that clustered with another tissue sub-group were at higher risk of specimen contamination at the time of DNA extraction due to location and/or size. This is summarised in Table 4.7.

Table 4.7: Evaluation of Sample Slides and Risk of Contamination

Sample	Slide review	Risk (high/low)	Cluster pattern
2*	Small area of A next to N	High (for NvA)	Not run
4	N and A on separate slides	Low	Only A dissected*
13	N and C adjacent on same slide, A on separate slide	High (for NvC)	N&C sep A&C same
14	N and C on separate slides	Low	N failed on ChAMP
15	N and C on separate slides	Low	N&C sep
20	N and C on separate slides	Low	N&C sep
21	N, A, C on separate slides	Low	N&C sep
22	N and C on separate slides	Low	N&C same
24	N, A, C on separate slides, A very small	High (for A)	N&A same C sep
27	N and C adjacent on one slide, A on separate slide	High (for NvC)	N&C sep N&A same
28	A and C adjacent on one slide	High (for AvC)	N sep A&C same
30	N, A adjacent on same slide; C separate on same slide	High (for NvA)	N&A same C sep
31	N and C on separate slides	Low	N&C sep
32	N and C on separate slides, N small	High (for N)	N&C sep
33	N and C on separate slides but both small areas	High (for NvC)	N&C similar
34	N on separate slide, A and C on adjacent slide and small.	High (for AvC)	N,A,C same
35	N, C on separate slides; both centrally located.	High (for NvC)	N&C similar
36	N and C on separate slides, N small	High (for N)	N&C similar
42	Query at time of dissection if could differentiate N and C	High (for NvC)	N&C same

N: normal endometrium; A: atypical endometrial hyperplasia; C: endometrioid endometrial cancer

Sep: specimens in separate clusters on cluster plot; Same: specimens in the same cluster on cluster plot; Similar: specimens lie in the same main cluster but not directly adjacent on cluster plot

*sample 2 was not analysable on ChAMP as it could not be reviewed in the ComBAT normalisation step. Sample 4 normal tissue was centrally located and not amenable to needle dissection.

Although the specimens 24A, 30A, 33C, 34A, 36N, 42 would be at risk of being contaminated with other tissue types, there are other evaluable samples that were also at risk of contamination that clustered separately as well as low risk samples that appeared to cluster together. In addition, the clustering that was demonstrated for these samples was not always consistent with what may occur from dissection technique. For example, for specimen 24, the 3 histological subtypes were on different slides. Thus, the clustering of the normal and AEH samples together may not be explained by technical dissection factors alone.
As such, rather than exclude any of these specimens, all were included in evaluation of MVPs and DMRs through ChAMP for this study, as the differences seen in clustering could not be attributed to contamination alone but may equally have reflected the differential methylation being investigated. Particularly with the small number of atypical specimens available for comparison, sample exclusion would also restrict subsequent analysis and interpretation. Only the cluster dendrograms and heatmaps in Chapter 5 (Figure 5.2) were analysed with these specimens excluded to visualise the spectrum of differential methylation compared to that with all specimens included.

The slides of three cases, two deemed at high risk and one at low risk of contamination, are shown in Figure 4.13. The remainder are included in the Appendix section 4A.

Figure 4.13: Examples of the Slide Review of ChAMP Evaluable Samples

Case 21 is deemed low risk as the normal (N), atypical and cancer (T) are all on separate slides. All cluster with the corresponding tissue group on ChAMP analysis.
Case 27 is deemed high risk as it has normal (N) and cancer (Ca) tissue adjacent on the same slide (centre slide). Both cluster with the corresponding tissue group on ChAMP analysis.
Case 30 is deemed high risk as the normal (N), atypical and cancer (Ca) tissue is adjacent on the same slide (on the right). 30A clusters with the normal specimens on ChAMP analysis.

21 27 30

4.9. Discussion

Given the complexity of the sample processing for analysis, detailed evaluation of each step for QC was performed. This included selection of specimens that were accurate for tissue type, extraction of quality DNA for analysis, accurate concentration assessment and quality bisulfite conversion.

4.9.1. Review of Dissection Technique

A number of techniques are available for FFPE DNA extraction including needle dissection, laser microdissection and core dissection using a tissue microarrayer. Needle dissection was the method chosen for DNA extraction, and relies on the replication of the same area of tissue, on each sectioned slide such that the tissue type of interest can be dissected based on an H&E slide template. In cases where the tissue area of interest is of a reasonable size and in a location amenable to dissection, this method is preferred. For smaller areas of tissue where there may be uncertainty regarding reproducibility in sectioned slides, each slide was reviewed under light microscopy at the time of dissection. Needle dissection however does exclude some samples where the tissue of interest is of small volume in the centre of the slide for example, and the reliability of accurate dissection is reduced. For large areas of tissue, this method is also relatively time efficient in that only 10-20 slides may need to be dissected. However, for smaller area of tissue, up to 40 slides will need dissection to ensure adequate DNA extraction.

Laser capture microdissection (LCM) is an alternative method ideal for small areas of tissue or in difficult locations, such as in the centre of a slide. This involves direct visualisation of the cells of interest via microscopy and transfer of the cells via laser energy into specific collection tubes [256]. This method requires specific slides, at increased cost, for tissue sectioning, and it is also more time consuming for small amounts of tissue, again with up to 40 slides required for dissection and each slide taking 15-20 minutes to process. Core dissection via a tissue microarrayer is the most time efficient method of collection and uses the microarrayer to extract larger cores of tissue measuring up to 2mm in diameter and 6-8mm in depth. This method again relies on the reproducibility of the H&E slide with the tissue block but there is a higher risk of tissue contamination, as there is no option to microscopically review the tissue at the time of dissection. Thus, although there were some limitations to needle dissection, it was deemed the most suitable for DNA extraction in this study as it best balanced time and cost efficiency with tissue accuracy, as would be required to expand the clinical utility of this analysis.

In terms of FF tissue dissection, this was based on macroscopic assessment by the histopathologist at the time of initial specimen review. In this study, there was no microscopic review of these samples as the tissue type was clear to the histopathologist at macrodissection, except for one specimen (42). To improve on the process in cases such as these, the FF tissue could also be sectioned while frozen into 5-10 micron thickness slides and a representative H&E slide evaluated every 25-50 microns to check changes in architecture and tissue type.

4.9.2. Review of Concentration Analysis

As the DNA concentration forms the basis of the subsequent modification steps, accuracy at this point was critical and was improved with the introduction of Qubit fluorometer to the protocol. As outlined in the methods section 2.2.4.1.2, NanoDrop is a spectrophotometric instrument that measures the DNA concentration based on the peak light absorption of a sample. It is time-efficient in that no dilutions are required to measure the concentration but it does not reliably distinguish between double and single stranded DNA, RNA, free nucleotides and other contaminants which scatter light or are UV-absorbing materials. Thus although it is quick and easy to run, it tends to overestimate the concentration of DNA that is present in the sample.

As the Qubit fluorometer utilises fluorescent dyes that only emit signals when bound to the specific target molecule, it provides a more accurate assessment of DNA concentrations compared to NanoDrop. It is more time consuming to use, in that each of the specimens has to be diluted and then processed on the fluorometer, but results achieved on the later 450K arrays compared to the initial array were improved, which in part may have been due to more accurate assessment of concentration with the Qubit fluorometer. For the initial array, even with Illumina Infinium QC≤5, the majority of specimens were not assessable on ChAMP, potentially due to the lack of DNA available for bisulfite conversion and analysis if the NanoDrop derived concentration was incorrect.

4.9.3. Review of the Illumina quality control assay for DNA modification

The Illumina QC assay is designed to select the specimens with the best quality DNA to take forward to further modification and analysis. It is based on the assessment of FFPE DNA prior to performing the Illumina Infinium HD restoration protocol, rather than the alternative method of ligation and bisulfite conversion that was used here. As limited EC specimens have been modified using this method of ligation and bisulfite conversion, a range of specimens with different QC values was analysed to determine if the pre-defined Illumina cut-off was consistent with what was generated with endometrial tissue and this methodology. Overall,

the majority of specimens with an Illumina Cq value ≤5 were successful in subsequent modification and as analysed on ChAMP, with a correlation of 78%. This allows a broader definition of specimens than the Illumina Delta Cq cut-off <5 used for the Illumina restoration protocol. Accurate definition of evaluable samples is of particular importance in the clinical setting when trying to maximise sample utility but eliminate samples that will fail on the Illumina 450K array that is expensive to run.

4.9.4. Review of Tissue Sample Selection

The ChAMP pipeline for processing the Illumina 450K methylation data includes stringent QC parameters to optimise the results generated. This study showed early evidence of differential methylation between normal endometrium, AEH and EEC, reviewed in greater detail in Chapter 5. Whether due to tissue contamination or other factors, there were some samples that demonstrated a similar methylation pattern to the comparator tissue type. Rather than excluding these samples, all samples were used for further analysis for a number of reasons. Firstly, although some of these samples demonstrated a risk of contamination based on review of the dissection techniques, this was not consistent across all the samples. Secondly, in an already small number of samples, further reduction in the sample size would not be warranted and could bias results.

Future research would need to validate these early findings in a larger group of patients, ideally collected across institutions and for FFPE and FF samples, where infrastructure is available. Tissue both from early stage and advanced disease should also be included, and aim to compare methylation patterns between these stages of disease, as well as between early stage disease that recurs and early stage disease that does not recur to see if a methylation pattern exists that may be used to predict outcomes and guide treatment intensity. Similarly in the advanced disease setting, there may be a methylation pattern that predicts response to chemotherapy and indeed, targeted agents. The methodology described here in terms of needle dissection and the methods of DNA modification are reproducible and feasible, though were made more difficult by limitations in the amount of normal and atypical tissue in particular. Once cancer is diagnosed in the sample, it is seen as less imperative by the histopathologists to block out adequate amounts of normal and atypical tissue for study, thus making this type of analysis challenging. Communication with the surgical and histopathology teams is essential to optimise the amount and quality of available tissue. Attention to processing and preservation techniques may improve the results and conclusions that can be drawn from further research, particularly when using FFPE tissue.

Overall, the methodology and QC checks utilised in this study were feasible, as outlined in Figure 4.14, and have the potential to be reproducible across centres with adequate education and communication.

Figure 4.14: Key Findings in the Review of Sample Management for Analysis

- Needle dissection is feasible to accurately dissect endometrial tissue types
- DNA concentration assessment via fluorometry is preferred to spectrophotometry
- A Delta Cq ≤5 on the Illumina Infinium HD FFPE QC assay along with bisulfite conversion ≥95% can predict successful analysis on the Illumina 450K array and ChAMP pipeline for analysis of differential methylation

CHAPTER 5: Methylation and CNV Analysis of Endometrial Specimens

5.1. Introduction

DNA methylation is one of the most well documented mechanisms of epigenetic aberration. It comprises reversible covalent modification of cytosines through the addition of a methyl group to the carbon-5 position of cytosine bases by one or more DNA methyltransferase (DNMT) enzymes, occurring predominantly in the context of cytosine-guanine dinucleotides (CpGs). Broadly, this affects chromatin packaging and gene expression [186]. Although reversible, it is much less dynamic than acetylation. CpGs that occur randomly throughout the genome tend to be methylated. CpGs also occur in clusters, known as CpG islands (CGIs), ranging from 500 up to 2000 base pairs and these tend to be unmethylated, unequally distributed and localized within and around the promoter regions of mammalian genes [188]. CGIs remain unmethylated in normal genes, except where DNA methylation is required for normal embryonic development, X-chromosome inactivation and genomic imprinting [257, 258]. DNA methylation is one of the primary heritable epigenetic gene expression processes, but epigenetic changes can also be affected by age and environment [259-261].
In the setting of carcinogenesis, promoter hypermethylation involving a tumour suppressor gene may lead to decreased gene expression, and thus promote carcinogenesis. Conversely, promoter hypomethylation of an oncogene could lead to its increased expression [262].

DNA methylation has been shown to differ between normal and cancer tissue, with one of the first reports being of hypomethylation of the genome of cancer cells in comparison to normal tissue [192]. *De novo* methylation of CGIs around the promoter region of genes causing repression of affected genes is well documented [188] and recent studies suggest that within each tumour, hundreds of genes may be silenced by DNA methylation, compared to 10-15 genes that may be silenced by genetic mutations [193].
Moreover, DNA methylation in cancer appears to be tissue and tumour-type specific [190], with the potential that DNA methylation profiles may act as biomarkers for diagnosis, prognosis and response to treatment, as well as identifying biological pathways which are disrupted in tumour development and may be targeted.
Epigenetic alterations and aberrant DNA methylation may be more frequent than genetic alterations in EC with documented epigenetic abnormalities in genes encoding tumour suppressors, cell cycle regulators, steroid receptors, transcription factors, angiogenesis modulators and oncoproteins [196]. As detailed in section 1.4.2.2, such abnormalities may occur in the RAS-MAPK, FGF, PI3K, Wnt and MSI pathways [200, 203-223], though clinical correlation in these earlier studies is often lacking. More recently, TCGA has analysed genetic

and epigenetic data in FF EC [30] as outlined in section 1.4.1.1, to which FFPE data in this study is compared.

What has not yet been studied is the spectrum of epigenetic changes that occurs between normal endometrial tissue, AEH and EC. In this study, this spectrum of epigenetic changes across normal, AEH and EC tissue has been examined in order to identify events that occur early in endometrial carcinogenesis which may be relevant to driving the oncogenic process. EEC alone was analysed here to provide a homogenous group of samples for comparison.

As outlined in Figure 5.1, this chapter presents the DNA methylation results from analysis of FFPE normal endometrial, AEH and EEC tissue on the Illumina HumanMethylation 450K array platform. Assessment is made as to whether differential methylation exists between these tissue types that may be of prognostic or predictive value. Data was compared to available TCGA data. In addition, an in-depth review of the genes and molecular pathways that were differentially methylated between normal endometrium and EEC was performed, using Web Gestalt and the R console. As there was not a significant number of MVPs or DMRs between the normal endometrium v AEH and AEH v EEC comparisons (Table 5.1), detailed molecular analysis could not be performed between these histological types.

CNV was also analysed across the samples to identify genomic changes that occur during the normal endometrium, AEH to EEC transition in endometrial carcinogenesis. Again, as discussed in Chapter 1 and particularly relevant with emerging TCGA data [263], there is limited published information that compares these changes across the histological types. By investigating the genomic changes that occur early in carcinogenesis, management strategies and novel therapies to improve patient outcomes may be better identified.

Figure 5.1. Aims of the Epigenetic and CNV Analysis in EEC

- Assess feasibility and reproducibility of epigenetic and CNV analysis from FFPE and FF EEC tissue on the Illumina 450K array
- Analyse methylation patterns in normal endometrium, AEH and EEC and correlate with clinical data
- Correlate epigenetic and CNV data with TCGA and other published literature
- Investigate molecular pathways and candidate genes with differential changes between normal endometrium, AEH and EEC.

5.2. Results of the Epigenetic Analysis

5.2.1. Demonstration of Differential Methylation in Normal, Atypical and Endometrioid Endometrial Cancer tissue

The evaluation of differential methylation between normal endometrial, AEH and EEC tissue utilised the ChAMP pipeline, as detailed in Sections 2.2.6.1 and 4.6. To analyse significant MVPs and DMRs, only 2 tissue types could be compared per analysis, as detailed in Section 4.6 and Table 5.1 below. To review the methylation data diagrammatically on cluster dendrograms and heatmaps, all 3 histological types could be analysed at one time, detailed in Figures 5.2-5.6. Therefore a number of different analyses were performed, either with 2 or 3 histological groups, as per Table 5.1. The following comparisons were performed: (1) all normal endometrial, AEH and EC samples, (2) matched samples where all 3 histological subtypes were available for an individual case (NAC matched in Table 5.1) and (3) grade 1 and 3 samples between and within histological subtypes.

As discussed in sections 2.2.6.1 and 4.6, the ChAMP pipeline filters out probes on the array that fall outside levels of significance for detection ($p > 0.01$) in more than one sample. For example, from the 69 unfiltered samples in the NvC comparison, only 12 750 from 450 000 probes were suitable for further analysis. Once samples were excluded for which the fraction of failed probes exceeded 5-10%, the NvC comparison included 39 high quality samples where 400 812 from 450 000 probes could be analysed.

For each comparison, Table 5.1 details: (1) the total unfiltered sample number and statistically significant probes per comparison, (2) the filtered sample number and statistically significant probes per comparison and (3) the number of significant MVPs and DMRs per comparison.

Table 5.1: Differential Methylation Analysis by Tissue Type and Grade

Comparator groups	Total unfiltered samples & significant probes	Filtered samples & significant probes	Significant MVPs & DMRs
N v C	69 & 12 750	39 & 400 812	23572 & 706
N v A	53 & 24 737	26 & 424 168	4 & 0
A v C	57 & 14 422	29 & 405 496	0 & 0
NAC	-	47 & 400 515	ne
NAC matched			
N v C	-	14 & 428839	1032 & 17
N v A	-	14 & 438 267	0 & 0
A v C	-	14 & 429047	0 & 0
NAC	-	21 & 425 123	ne
Grade 1			
N v C	47	21 & 413 732	1& 0
N v A	38	14 & 428 972	0 & 0
A v C	39	16 & 413 589	0 & 0

Grade 3 N v C N v A A v C	 23 16 18	 19 & 413 752 12 & 426 675 14 & 417 283	 0 & 0 0 & 0 0 & 0
Normal G1 v G3	33	19 & 422 800	0 & 0
Atypical G1 v G3	20	6 & 444 235	0 & 0
Cancer G1 v G3	39	22 & 407 535	0 & 0

MVPs: methylation variable positions, DMRs: differentially methylated regions, ne; not evaluable, N: normal endometrium, A: atypical endometrial hyperplasia, C: endometrioid endometrial cancer, G: grade

Figures 5.2 – 5.6 demonstrate diagrammatically the differential methylation of the 1000 most variable probes from the above comparisons. For grade 1, grade 3, normal, atypical, cancer and matched subgroup comparisons, the MDS, density plots and sample cluster plots are included in the Appendix section 5A. Only the cluster dendrograms (DGs) and heatmaps (HMs) are depicted here.

In Figures 5.2A and 5.2B, two HMs were generated for each of the NvC and NvA comparisons with certain samples excluded in the second HM (detailed below). As described in Section 4.6 on the MDS and cluster plots for these comparisons, 4 specimens in the NvC and 5 specimens in the NvA comparisons were outliers and clustered with the alternate histological tissue type. The HMs below show clustering based on the 1000 most variable methylation array probes with and without the outlier samples to assess if the differences seen in the MDS and cluster plots also occurred in the HMs. As also noted in section 4.6, subsequent analysis of the MVPs and DMRs was based on all the ChAMP filtered samples to maximise sample size of significant probes.

Figure 5.2: Cluster Dendrograms (DG) and Heatmaps (HM) based on the 1000 most variable probes for comparisons of
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)
D: Normal Endometrium v Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (NvAvC) (all grades included)

For all heatmaps (HM):

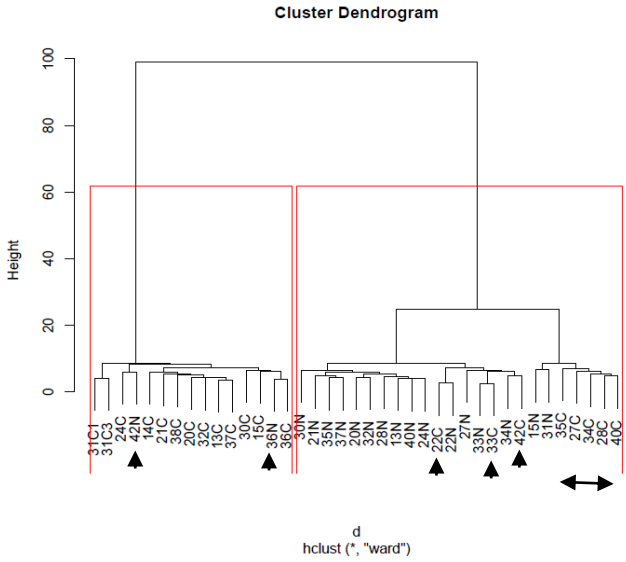
Normal endometrium: dark green, AEH: orange, EEC: red
Grade 1: light green, Grade 3: purple

The colour key depicts the colour corresponding to the beta value between 0 and 1. The beta value is the fluorescence intensity ratio between methylated and unmethylated probes. A ratio value of 0 (yellow) equals non-methylation of the locus, 1 equals total methylation (dark blue) and 0.5 means that one copy is methylated and the other is not (light blue).

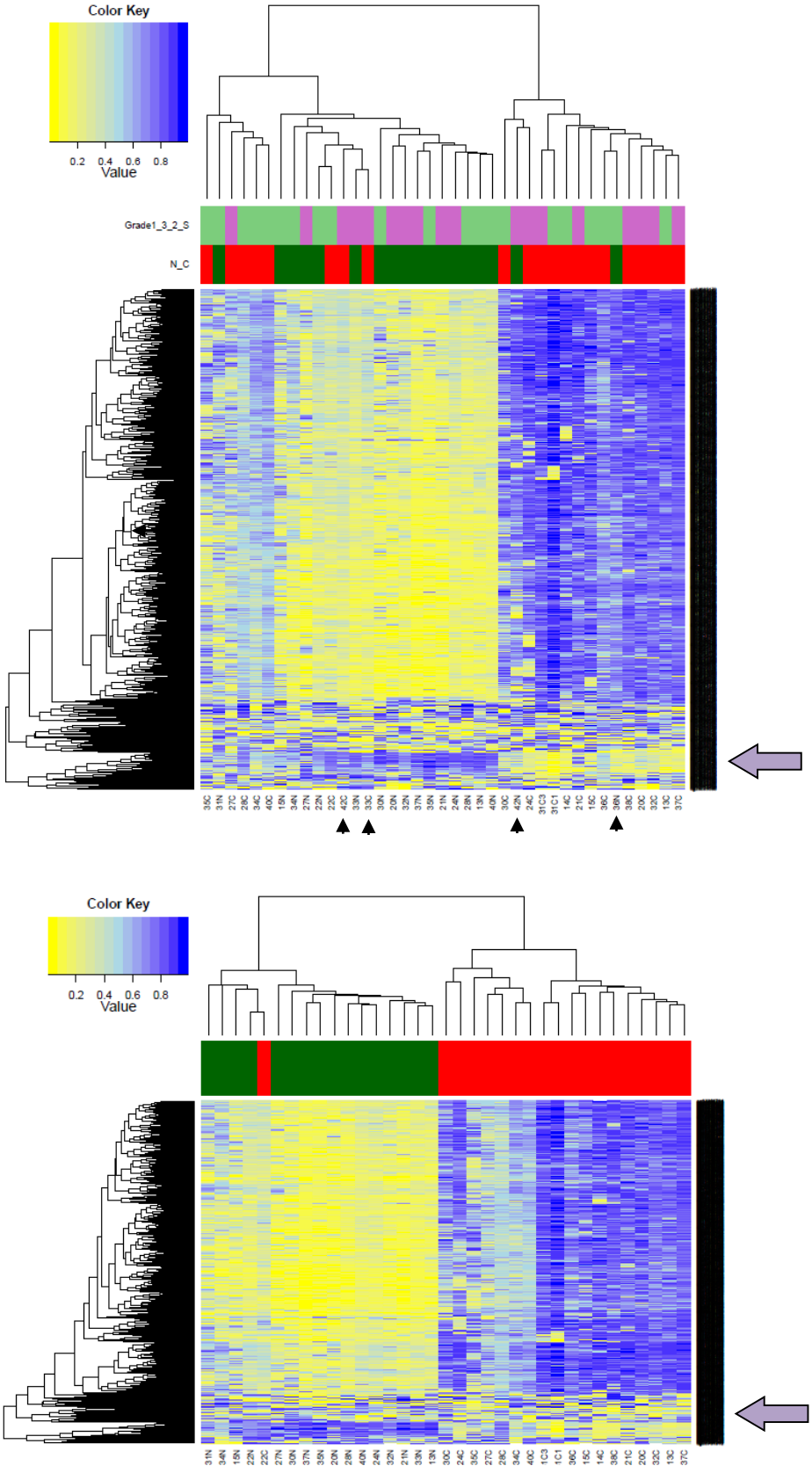
The dendrogram at the top of the heatmap shows sample clustering based on the 1000 most variable probes, with the samples detailed along the y-axis.
The dendrogram down the left side shows the clustering of the 1000 most variable probes from the Illumina 450K array.

A: NvC

The cluster DG below shows discrete clustering of the normal and cancer specimens except for 5 specimens: 22C, 33C, 36N, 42N, 42C. The 5 furthermost right cancer specimens on the DG (black arrow) are a separate cluster showing a level of hypermethylation intermediate between the main normal and cancer clusters, as demonstrated in the HM cancer specimens on the left.

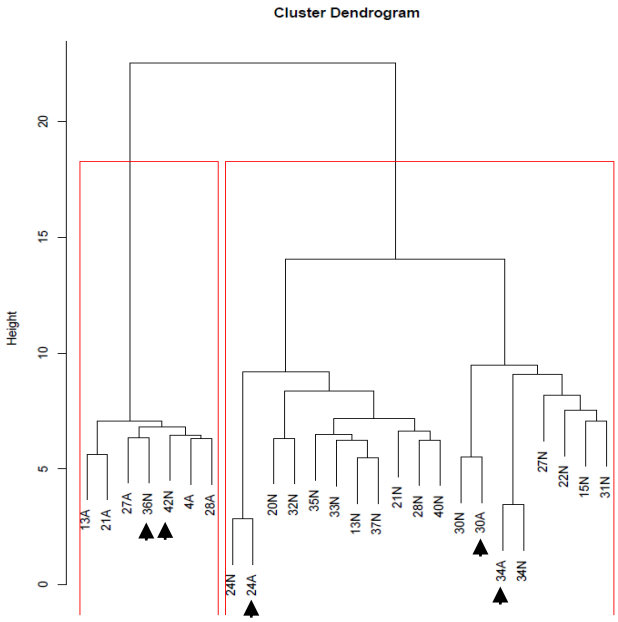


The top HM includes all specimens. The bottom HM excludes 4 outlier specimens: 33C, 36N, 42N, 42C (marked with black arrowheads). Both HMs show an increase in hypermethylation between the normal endometrium and EEC specimens with a beta value close to 1 (dark blue) in the majority of probes for the EEC specimens. Note that a subset of probes (bottom of HM, indicated by purple arrow) show hypomethylation in the EEC specimens, with the beta value close to 0.

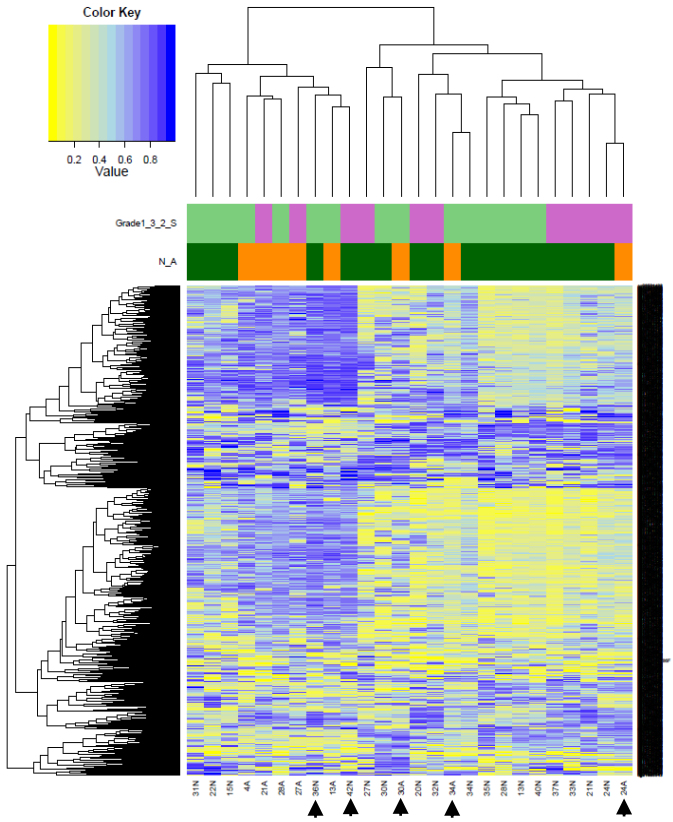


B: NvA

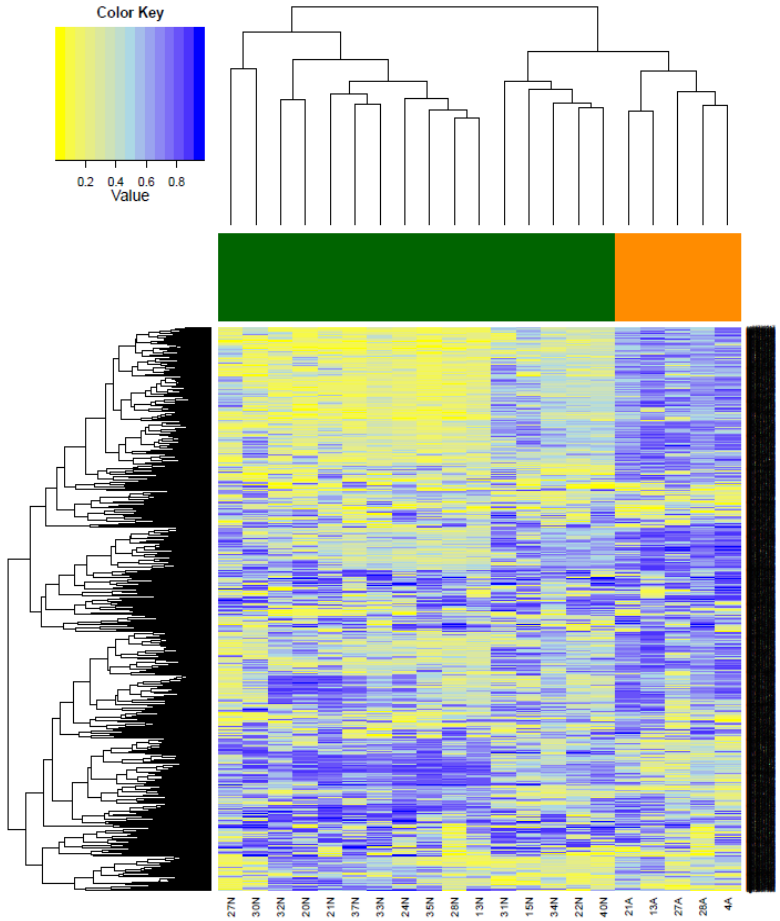
The cluster DG shows two main groups with the majority of specimens dividing according to normal and atypical histology, except 24A, 30A, 34A, 36N, 42N (black arrows).



The HM below shows all specimens in the analysis, with hypermethylation of the atypical specimens in many probes, but hypomethylation in some.

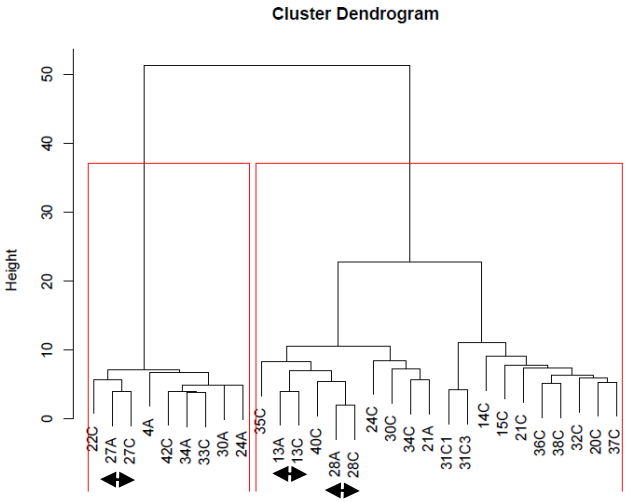


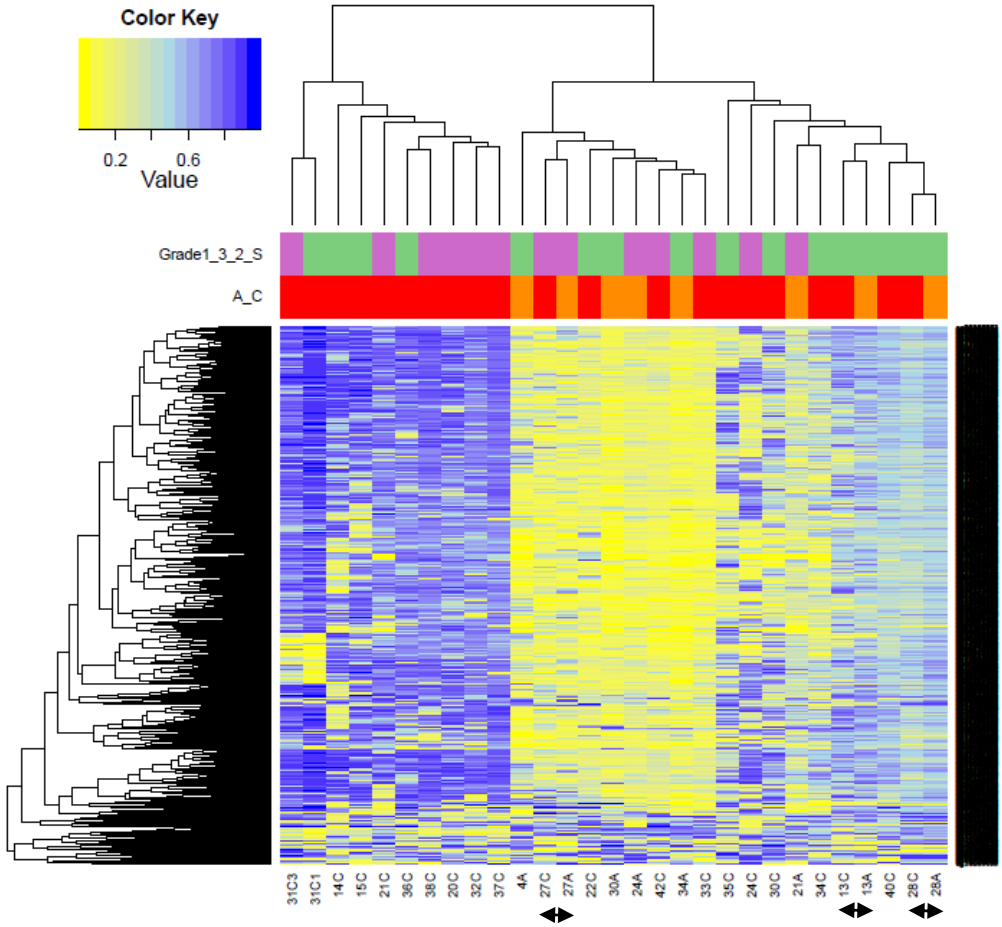
The following HM excludes the 5 outlier samples. There is a clear difference in clustering between the normal and atypical samples based on the DG at the top of the HM. On review of individual probe methylation status, there is a group of probes in the top quarter of the HM that are hypermethylated in the atypical samples compared to the normal samples. The probes in the rest of the HM show hypermethylation of some probes in the atypical samples compared to normal, but hypomethylation in others.



C: AvC

The DG does not show as clear a difference between AvC specimens, as was seen in the NvC and NvA comparisons. The HM demonstrates hypermethylation between some atypical and cancer specimens (towards the centre and left of the HM), while others show similar methylation pattern often in paired samples (for example 27 A/C, 13 A/C and 28 A/C, on the right of the HM and marked with black arrowheads).

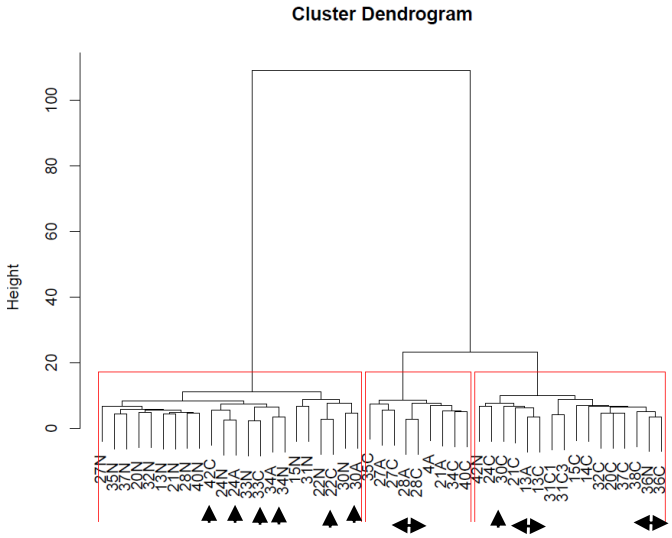


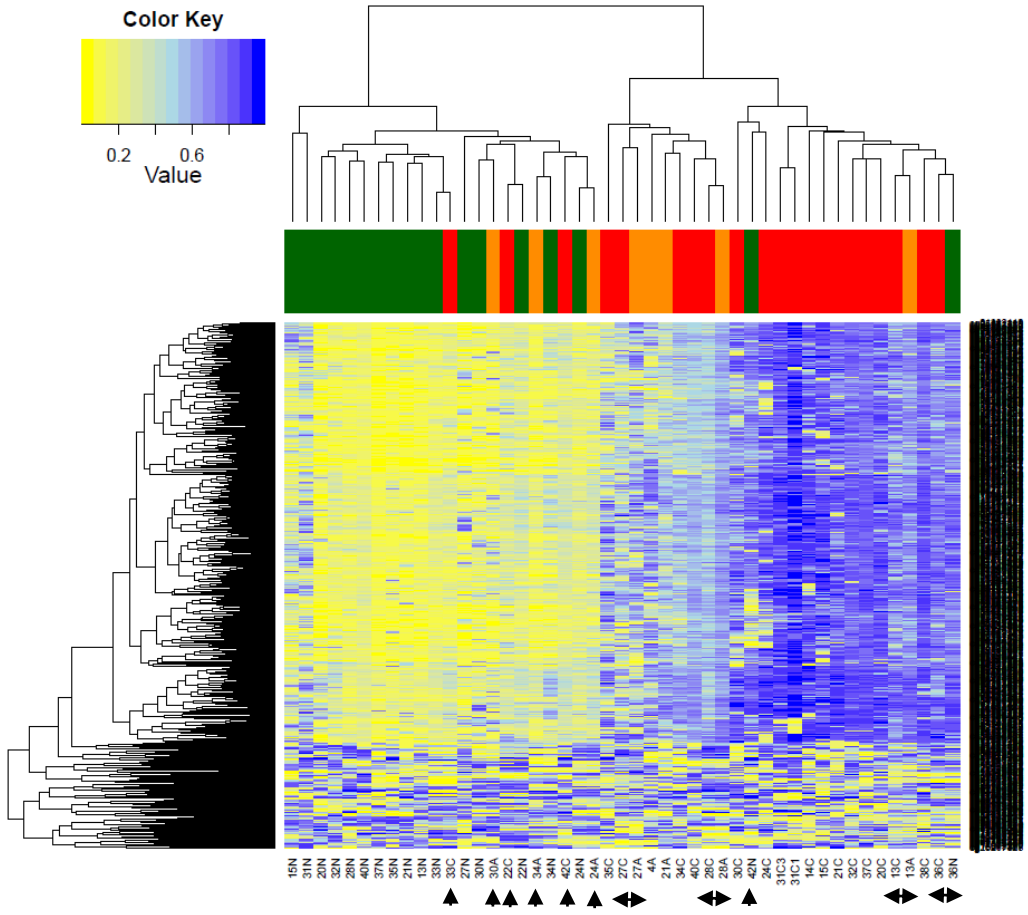


D: NvAvC

On review of all ChAMP filtered samples, there is evidence for differential methylation between the 3 histological subtypes. The DG shows 3 clusters, broadly corresponding to normal endometrium, AEH and EEC. On the HM, there is increasing hypermethylation between normal endometrium, AEH and EEC specimens, evidenced by the increasing beta values across the probes.

Note as previously identified on the paired comparisons, 13A, 22C, 24A, 27C, 28C, 30A, 33C, 34A, 36N, 42N, 42C appear as outliers, lying separate to their histological type. 13A/C, 22N/C, 24N/A, 27A/C, 28A/C, 30N/A, 33N/C, 36N/C lie adjacent (marked on DG and HM with black arrowheads). All are included in the HM and DG below, and in subsequent analysis.





The 4 comparisons above and the finding of MVPs and DMRS in the NvC and NvA comparisons, outlined in Tables 4.8 and 5.1, demonstrate differential methylation between the normal endometrium, AEH and EEC FFPE and FF specimens. Overall, the DGs above show clustering of the 3 histological subtypes based on the methylation pattern of the 1000 most variable probes, though there are some outlier specimens which do not cluster in a consistent manner. In addition, the HMs show a progressive increase in hypermethylation between the 3 histological subtypes, as evidenced by the change in the beta values across the probes between the normal endometrium, AEH and EEC specimens. As described, the majority of probes show increasing hypermethylation, though there is also a subset of probes that show hypomethylation between normal endometrium, AEH and EEC.

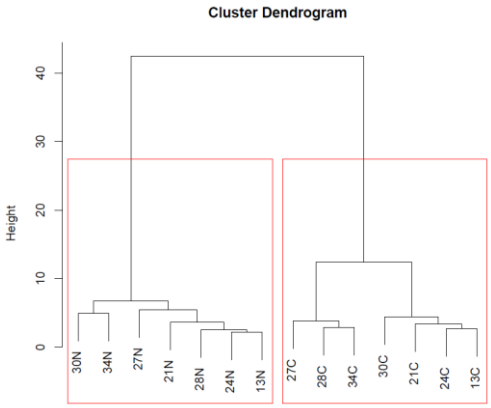
In particular, the heatmap of the 3 histological subtypes (Figure 5.2D) shows the spectrum of hypermethylation that occurs between normal, atypical and cancer tissue. Interestingly, some of the atypical samples show intermediate methylation between the normal and cancer samples, while a smaller number lie closer to their corresponding normal or cancer sample. For example, the AEH and EEC components of specimens 13, 27 and 28 have similar methylation patterns. As reviewed in section 4.8, this methylation pattern is unlikely to be attributable to dissection technique and suggests there may be a spectrum of methylation changes that occur either early or late in endometrial tumorigenesis. Specimens 13, 27 and 28 may represent those samples where the methylation changes occur between the normal to AEH transition, such that the AEH and EEC methylation patterns are similar. Conversely, specimens 24, 30 and 34 may represent those samples where the methylation changes occur between the AEH to EEC transition, such that the normal and AEH methylation patterns are similar. This is discussed further in section 5.4, but such patterns of changes may represent distinct methylation signatures of endometrial tumorigenesis. It certainly appears however, that the methylation patterns in AEH and EEC are different from normal tissue.

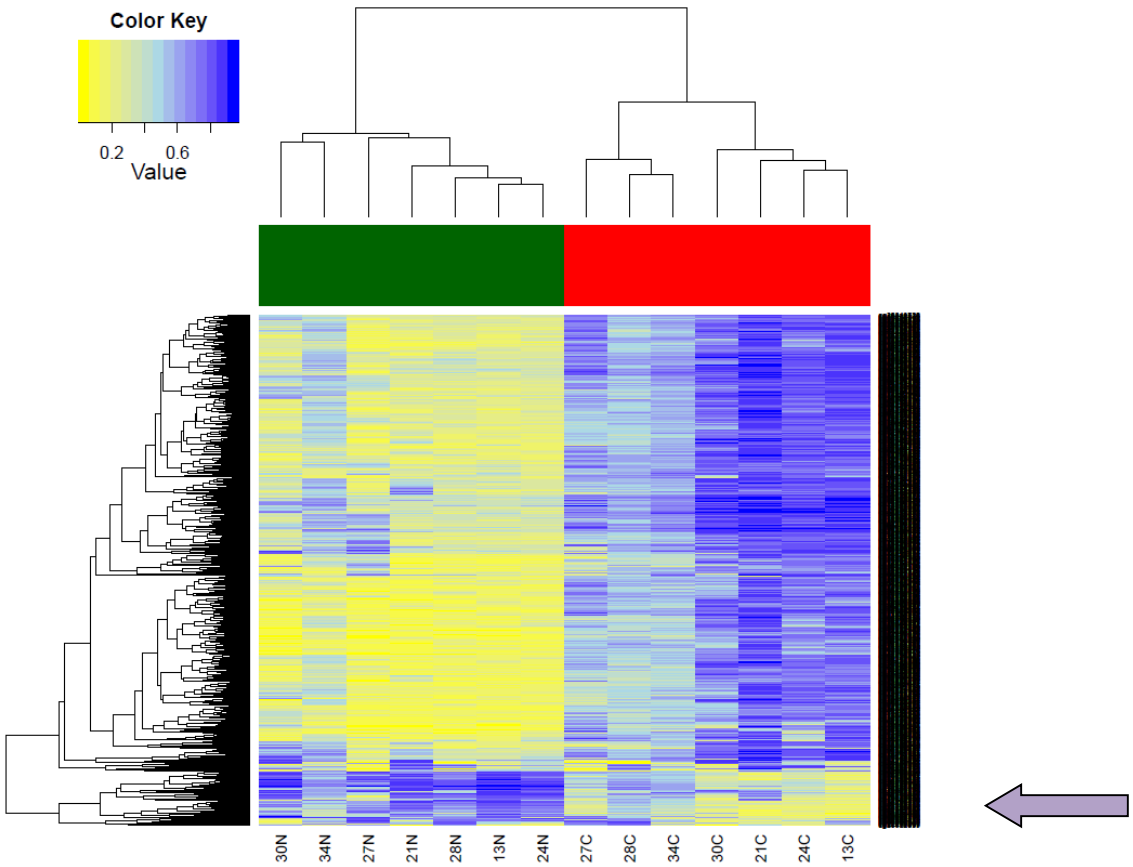
The differential methylation changes evident in the above all grade comparisons are further supported by the matched sample analysis in the following Figure 5.3.

Figure 5.3: Cluster Dendrograms and Heatmaps based on the 1000 most variable probes for matched samples
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)
D: Normal Endometrium v Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (NvAvC)

A: NvC

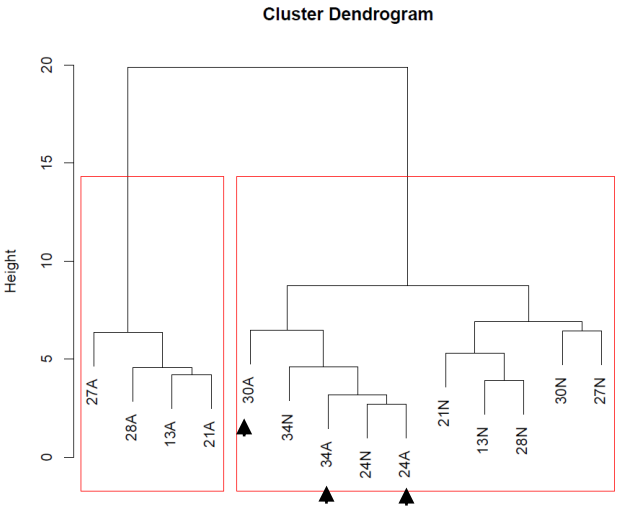
The DG shows 2 clusters between normal and cancer specimens. The HM also shows these 2 clusters, with hypermethylation in the majority of probes in the cancer specimens and hypomethylation in a subset of probes (bottom section of HM, purple arrow).

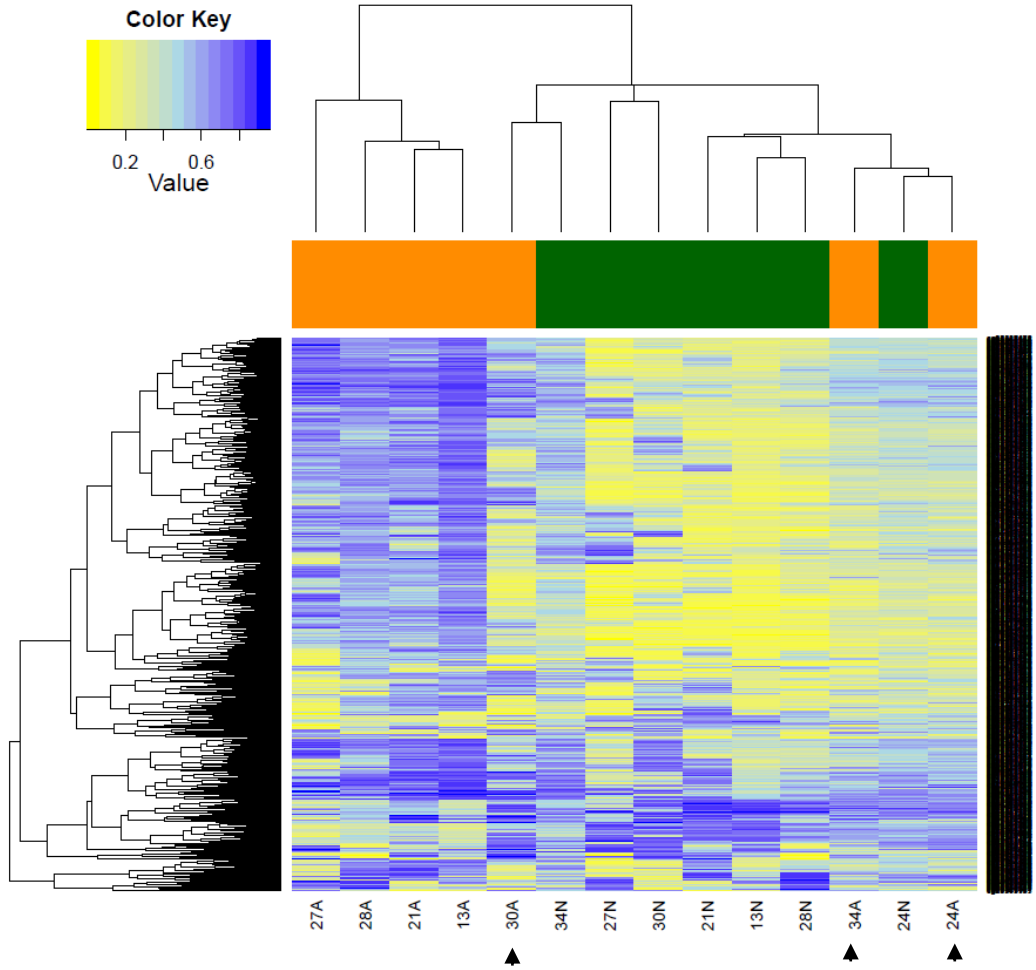




B: NvA

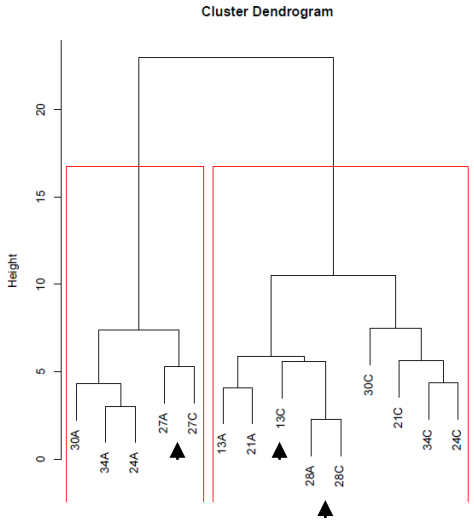
The DG shows two clusters of normal and atypical specimens, except for 24A and 34A (black arrows). On the HM, there is hypermethylation of the atypical specimens in over half of the probes. Note that fewer probes appear to be hypermethylated in this comparison compared to the NvC comparison. The HM also shows that 24A, 30A and 34A have intermediate hypermethylation changes compared to the main cluster of normal specimens.

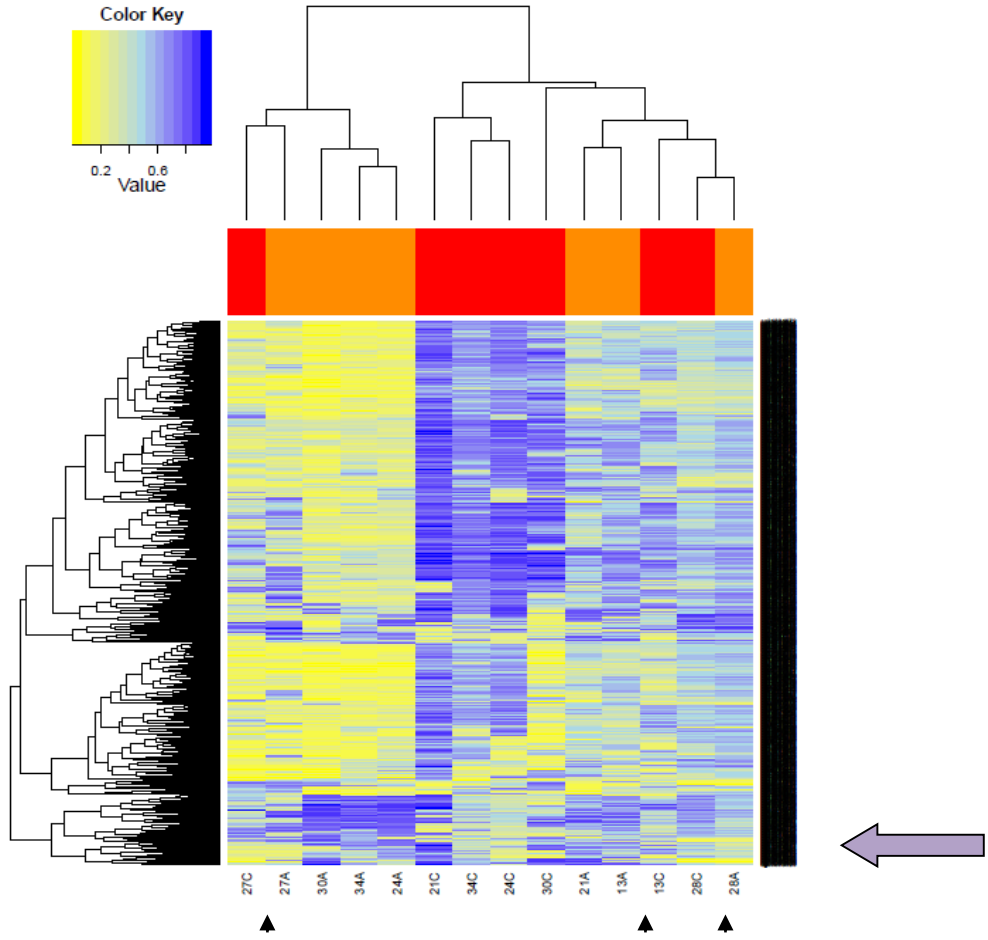




C: AvC

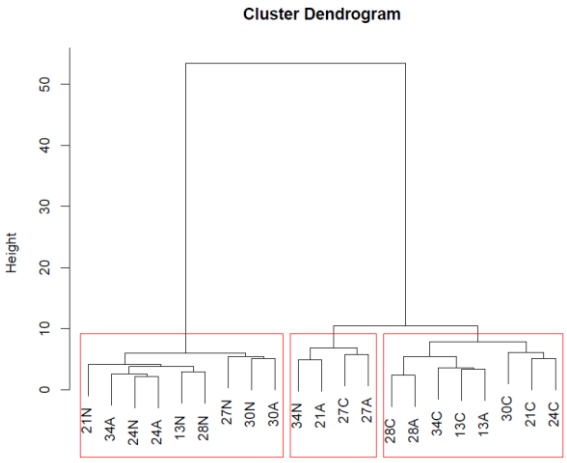
The DG and HM show differential methylation between a number of the atypical and cancer specimens, though the paired samples 13, 27 and 28 show a similar methylation pattern. Where these differential changes occur, the majority of probes show hypermethylation, though there is a smaller subset with hypomethylation (indicated by the purple arrow).

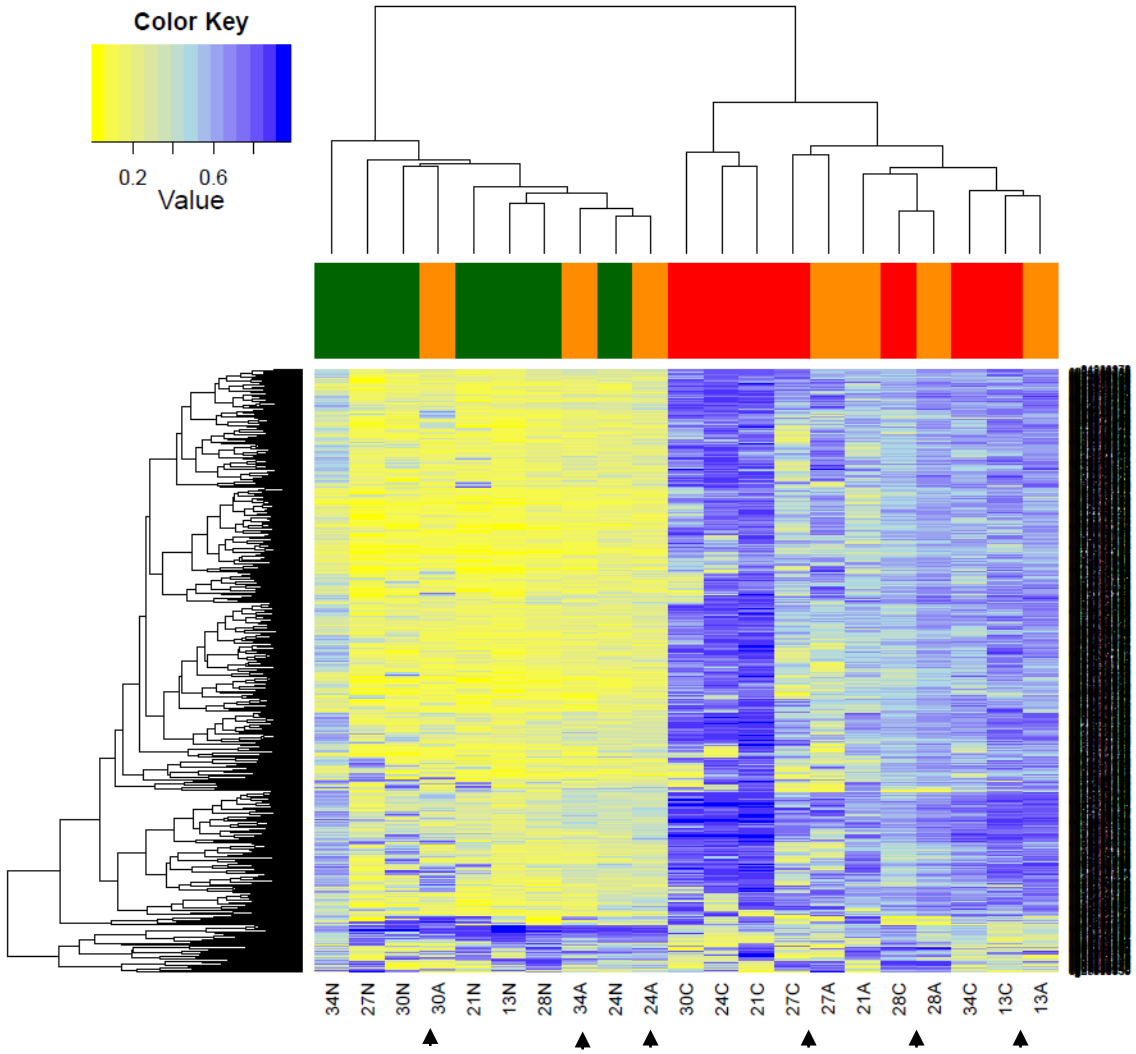




D: NvAvC

There is a spectrum of differential methylation changes between the 3 histological types on the HM, with corresponding clustering on the DG. The atypical specimens on the right half of the HM show an intermediate level of hypermethylation level compared to the normal and cancer clusters. Outliers 24A, 30A and 34A, as well as clustering of paired samples A/C 13, 27 and 28 are again noted.





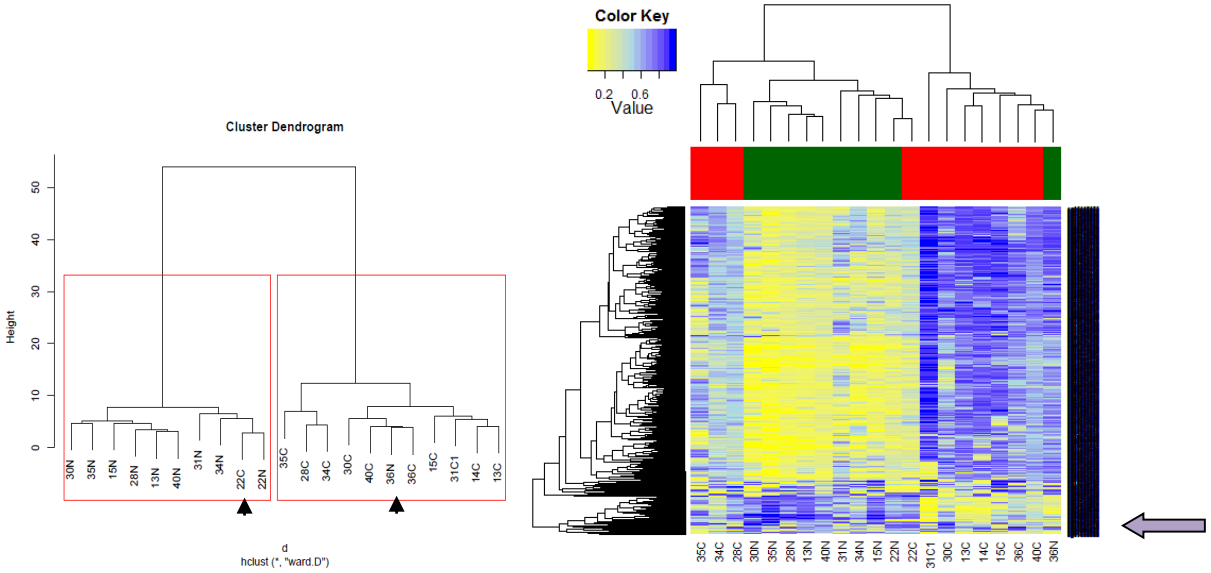
This comparison of cases where all 3 tissue types were available for analysis is consistent with the data generated by analysis of all specimens, in that there is a spectrum of methylation changes that occurs between normal endometrium, AEH and EEC. This is predominantly hypermethylation though there is a smaller subset of probes that show hypomethylation. The similarity between some normal and atypical samples, and between other atypical and cancer samples, is as described with Figure 5.2.

Figures 5.4 and 5.5 show comparison of NvC, NvA and AvC in grade 1 and grade 3 samples, to see if the grade of the tumour affected the degree of differential methylation that occurred. Figure 5.6 shows the comparison of grade 1 and grade 3 samples for normal, atypical and cancer tissue types, to see if within these histological types, there was a difference in methylation between grade 1 and 3.

Figure 5.4: Cluster Dendrograms and Heatmaps based on the 1000 most variable probes for comparisons of Grade 1 Specimens
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)

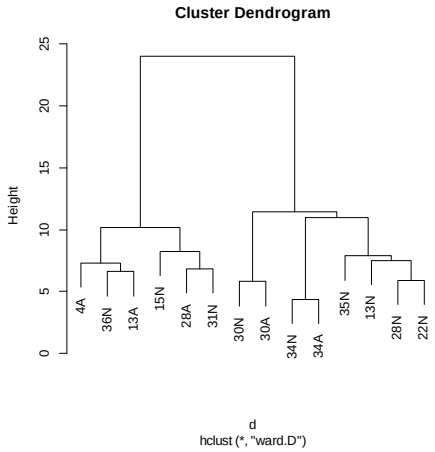
A: NvC

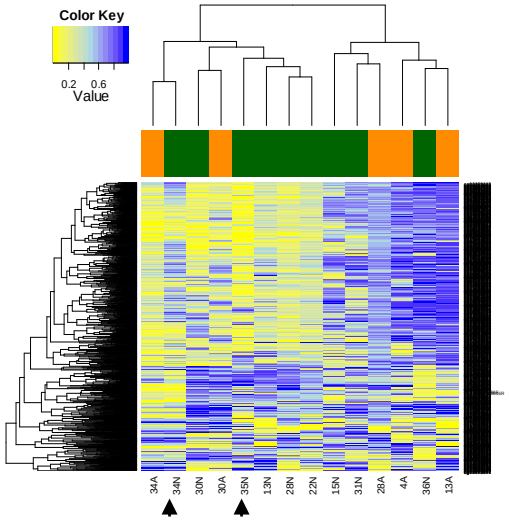
The DG shows 2 clusters correlating with normal and cancer specimens with only 2 exceptions, 22C and 36N. The HM shows hypermethyation of the cancer specimens with 28C, 34C and 35C on the left showing intermediate methylation between the normal and cancer specimens and 36N at the right side of the cancer specimens. There is a smaller subset of probes that are hypomethylated between normal and cancer (purple arrow).



B: NvA

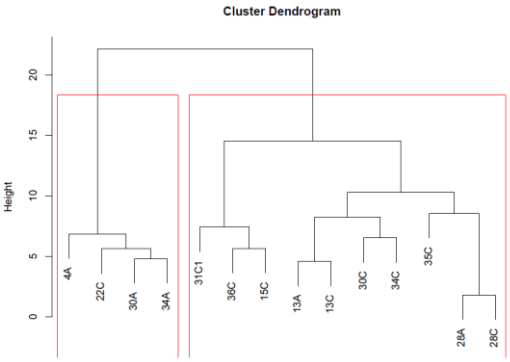
Clustering of the 2 histological types is not discrete on the DG. The HM however suggests hypermethylation in the majority of atypical specimens compared to normal, except for 30A and 34A, which are similar to normal.

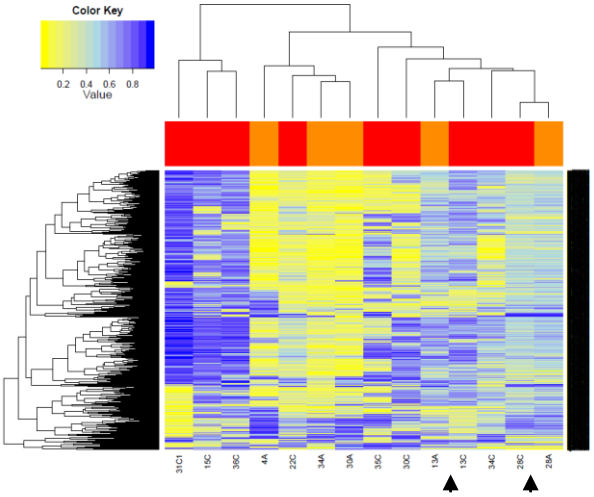




C: AvC

The DG shows 2 groups which broadly correspond to AEH and EEC histology. The HM shows some specimens with hypermethylation between atypical and cancer but the paired samples 13 and 28 have a similar pattern on the right.



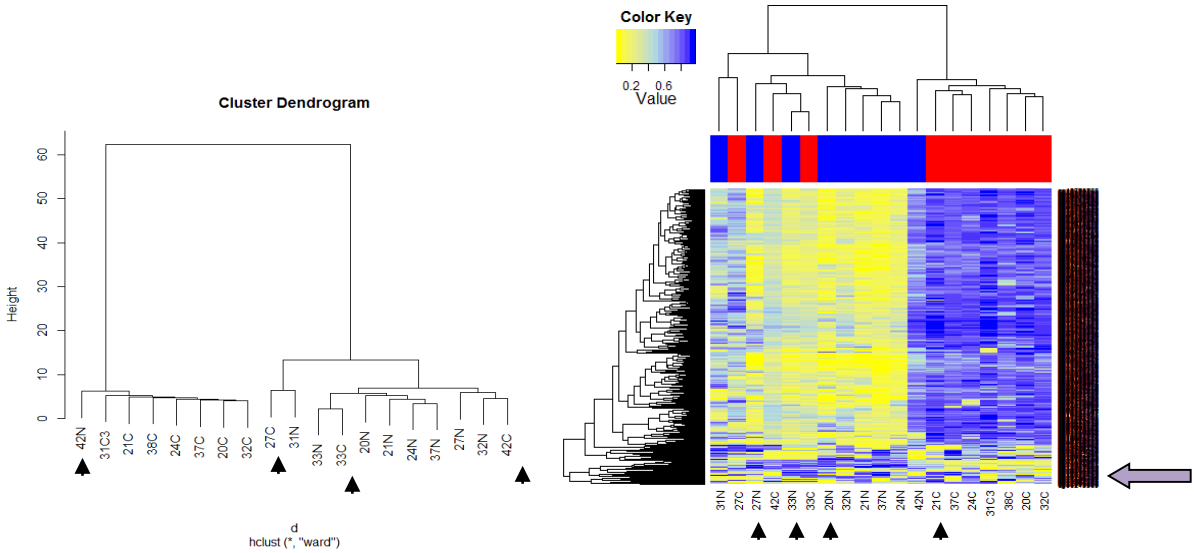


For the Grade 1 comparisons, there is evidence for hypermethylation between the normal and cancer samples with a small number of hypomethylated probes. In the normal vs atypical group, there appears to be a trend towards hypermethylation in the AEH specimens, though 30A and 34A show a similar pattern to the normal samples. Similarly, in the atypical vs cancer specimens, there are some specimens where the cancer tissue is hypermethylated compared to the atypical specimens but samples 13 and 28 have a similar methylation pattern.

Figure 5.5: Cluster Dendrograms and Heatmaps based on the 1000 most variable probes for comparisons of Grade 3 Specimens
A: Normal Endometrium v Endometrioid Endometrial Cancer (NvC),
B: Normal Endometrium v Atypical Endometrial Hyperplasia (NvA),
C: Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC)

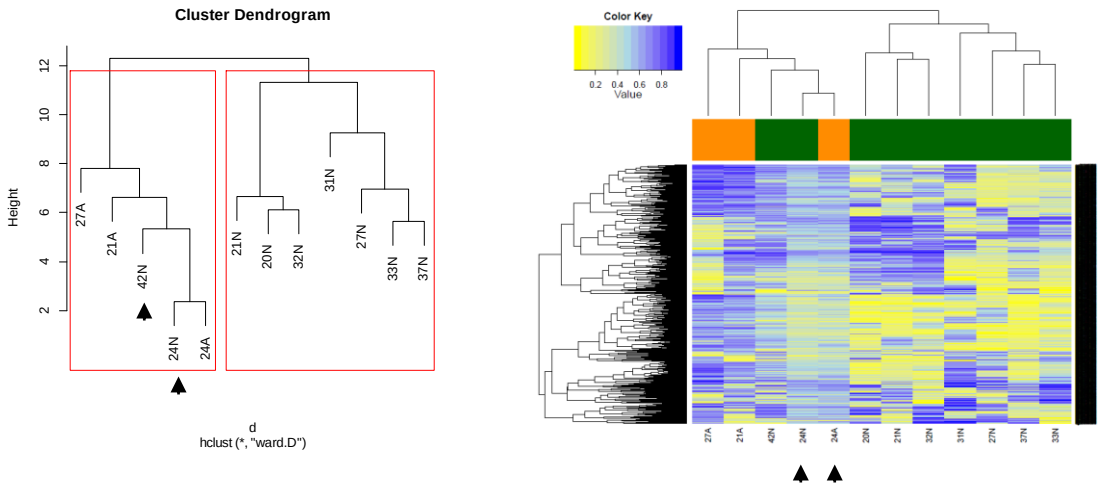
A: NvC

Other than 42N/C, 27C and 33C (black arrows), both the DG and HM show differential hypermethylation between the normal and cancer specimens, with a smaller subset of probes demonstrating hypomethylation (purple arrow).



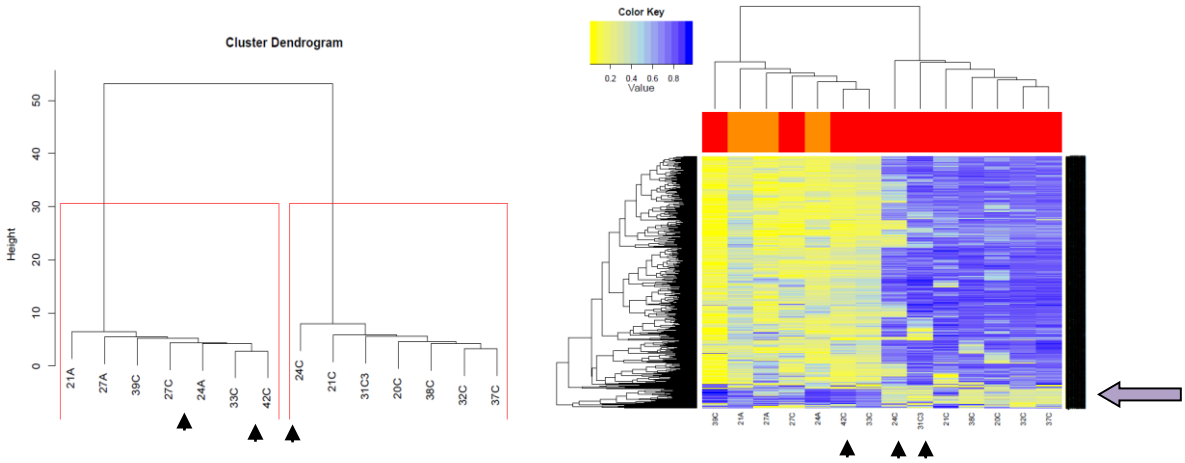
B: NvA

Despite the small number of grade 3 atypical samples, the DG shows the separate clusters of normal and atypical specimens with evidence for hypermethylation in the atypical specimens on the HM. The normal samples 24N and 42N (black arrows) cluster with the atypical samples and show a similar methylation pattern.



C: AvC

The DG shows separate clusters for the atypical and cancer specimens, though there are only 3 atypical samples and 27C, 33C and 42C (black arrows) are in that cluster. The main group of cancer specimens are hypermethylated compared to the atypical cluster, though there is also a smaller number of probes with hypomethylation (purple arrow).



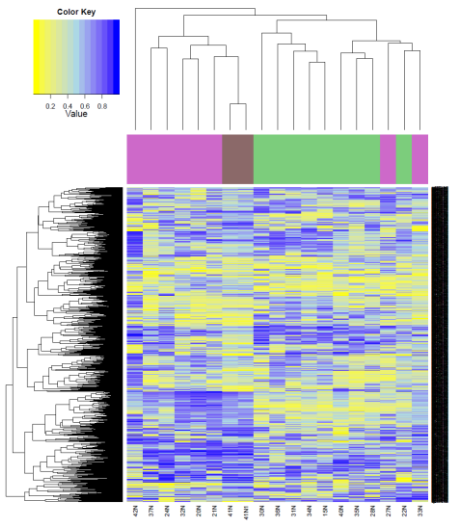
In the grade 3 comparisons, cancer specimens are hypermethylated relative to normal specimens, though again there is a small number of hypomethylated probes. There is differential methylation between some normal and atypical samples and similarly between some atypical and cancer samples, though others show a similar methylation pattern, as previously noted.

Overall, the changes in methylation described when all specimens were analysed together apply to both the grade 1 and grade 3 analyses.

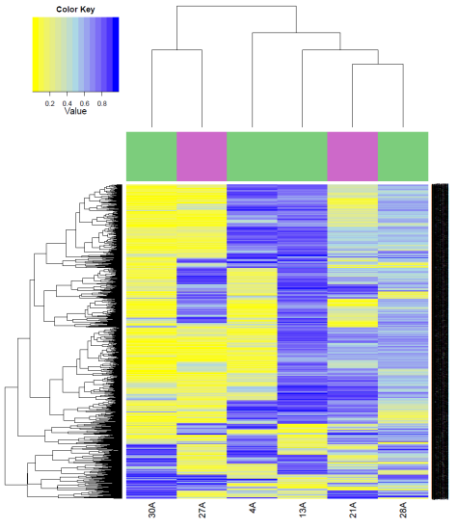
Figure 5.6: Heatmaps based on the 1000 most variable probes for comparisons of Grade 1 and Grade 3 specimens for A: Normal, B: Atypical, and C: Cancer (G1: green, G3: purple)

The normal samples, associated with grade 1 and 3 cancers, show differential clustering but no distinct methylation pattern on the HM. There is no consistent difference in methylation patterns between the associated atypical specimens. Some of the grade 3 cancer specimens are hypermethylated compared to the grade 1 cancer specimens.

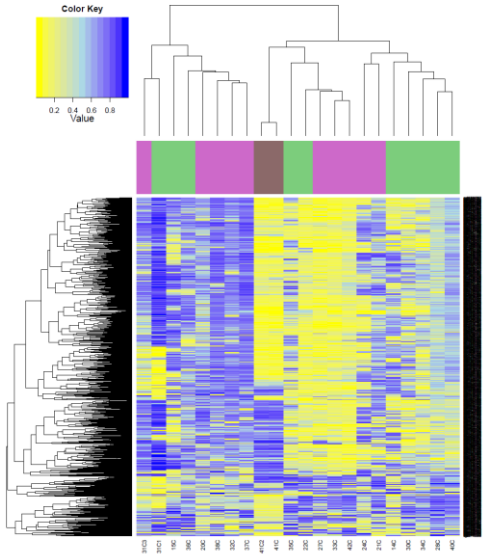
Normal



Atypical



Cancer



In the normal specimens, the cluster dendrogram at the top of the heatmap shows separate clusters of the normal tissue associated with grade 1 and grade 3 cancer samples, though there is a spectrum of methylation changes demonstrated. There is no clear difference in methylation between atypical specimens associated with grade 1 and grade 3 cancer, though this group has the smallest sample size for comparison with only 6 samples. Within the cancer specimens, there is some evidence for hypermethylation of grade 3 specimens compared to grade 1, as well as specimens of both grades that have similar features.

5.2.2. Correlation of Methylation Analysis with Clinical Data

Clinical data were collected for those samples that were evaluable through the ChAMP pipeline and are outlined in Tables 5.2 and 5.3. Note that all samples were EEC, though a few samples had smaller co-existing areas of other pathology. For specimens with grade 2 disease, it was either the grade 1 or 3 changes that were dissected for analysis.

The clinical data was correlated with the methylation data to ascertain if there was any association between clinicopathologic features and differential methylation. This was analysed on the NvC total group, as well as the NvC matched sample group. Figures 5.7 and 5.8 show the heatmaps with clinical factors and methylation status.

Table 5.2: Pathologic Features of ChAMP Evaluable Specimens

Sample number	Age	Stage	Grade	Tumour size (mm)	MMI> 50%	LVSI	Cerv invasion	LN +
4	53	IA	1	10	N	N	N	N
13	73	IIIB	1 (F G2)	32	Y	Y	N	N
14	55	IA	1	22	N	N	N	N
15	55	II	1	6	Y	N	N	N
20	77	IA	2 (F G3)	50	N	Y	N	N
21	50	IB	2 & 3	50	N	Y	N	N
22	66	IB	1	16	N	Y	N	N
24	72	IB	2 (F G3)	30	N	Y	N	N
27	65	IA	2 (F G3)	11	N	N	N	N
28	64	IA	1	10	N	N	N	N
30	70	IA	1	8	N	N	N	N
31	61	II	2 (F G3)	32	N	N	N	N
32	71	IIIC	3	35	Y	Y	N	Y
33	71	IB	3	46	Y	Y	N	N
34	36	IA	1	<10	na	na	na	N
35	61	IA	2	12	N	N	N	N
36	72	IA	1 (F G2)	15	N	N	N	N
37	66	IB	3	42	Y	Y	Y	N
38	79	IA	3	30	N	Y	N	N
40	76	II	1 &2	25	N	N	Y	N
42	58	IB	3 (NE)	65	Y	Y	N	N

MMI: myometrial invasion; LVSI: lymphovascular space invasion, Cerv: cervical, Pos: positive, F: focal, G: grade, NE: neuroendocrine, S: sarcomatous, na: not assessable

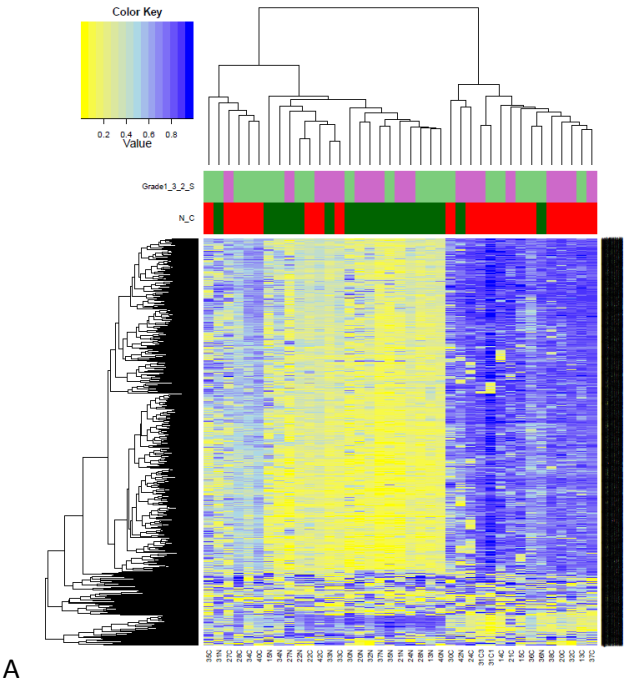
Table 5.3: Treatment Received and Patient Outcomes of ChAMP Evaluable Specimens

Sample	Adj chemo (Yes/No)	Adj RT (Yes/No)	Adj hormones (Yes/No)	Mets (Yes/No)	Time to Mets (m)	Survival
4	N	N	N	N	-	Y
13	Y	Y	N	N	-	Y
14	N	N	N	N	-	Y
15	Y	N	N	N	-	Y
20	N	Y	N	N	-	Y
21	Y	Y	N	N	-	Y
22	N	N	N	N	-	Y
24	Y	Y	N	Y	19	N (31m)
27	N	N	N	N	-	Y
28	N	N	N	N	-	Y
30	N	N	N	N	-	Y
31	N	Y	N	N	-	Y
32	Y	Y	N	N	-	Y
33	Y	Y	N	N	-	Y
34	N	N	Y	N	-	Y
35	N	N	N	N	-	Y
36	N	N	N	N	-	Y
37	N	Y	N	N	-	Y
38	N	Y	N	N	-	Y
40	N	N	N	N	-	Y
42	Y	Y	N	N	-	Y

Adj: adjuvant, Mets: metastases, m: months
Chemo: chemotherapy, RT: radiotherapy, Y: yes, N: no

Figure 5.7: Clinicopathological data and Normal Endometrium v Endometrioid Endometrial Cancer Methylation Status for all samples, A: grade, B: stage and age, C: tumour features, D: primary treatment

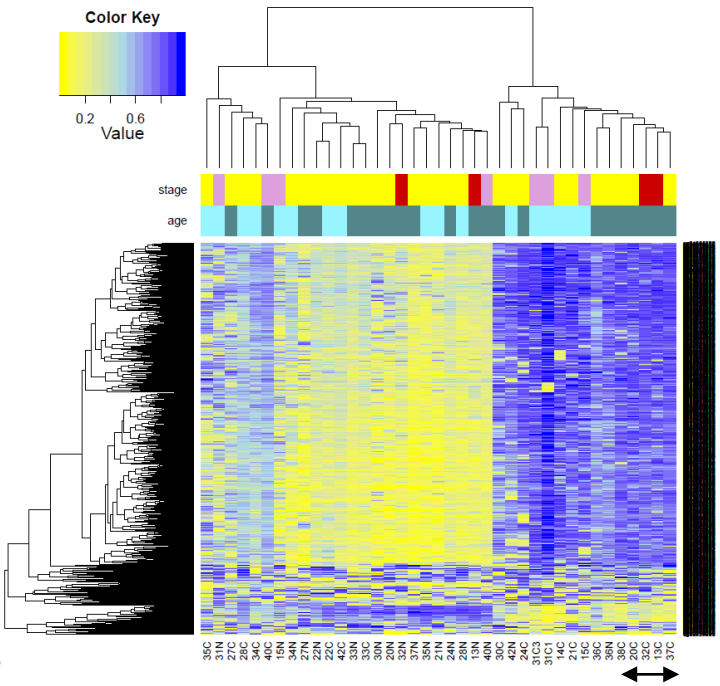
There appears to be a trend between patients' age ≥65 years and hypermethylation of EC samples in HM B, as well as tumour size ≥2cm and hypermethylation in HM C (both marked by black arrowheads). Otherwise, there is no definitive correlation between clinicopathological data and methylation status



Grade 1: green
Grade 3: purple

Normal: dark green
Cancer: red

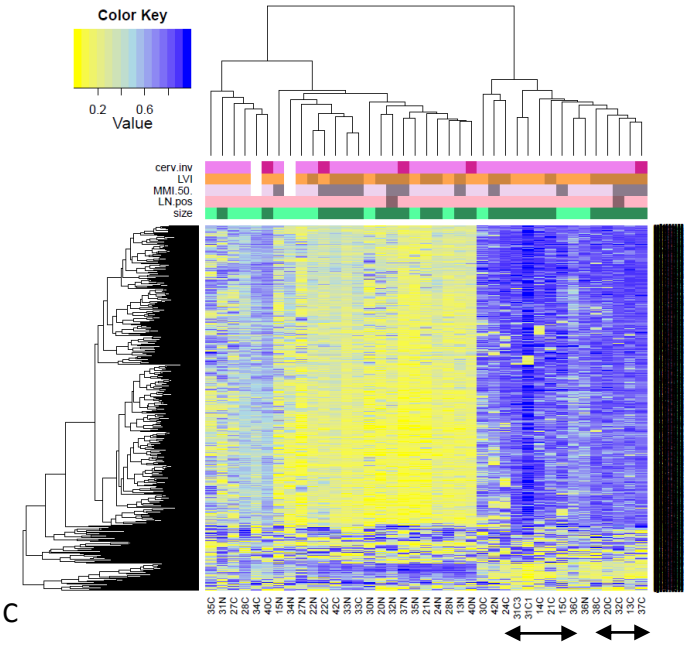
A



Stage I: yellow
Stage II: plum
Stage III: red

Age
<65yrs: light blue
≥65yrs: olive

B



C

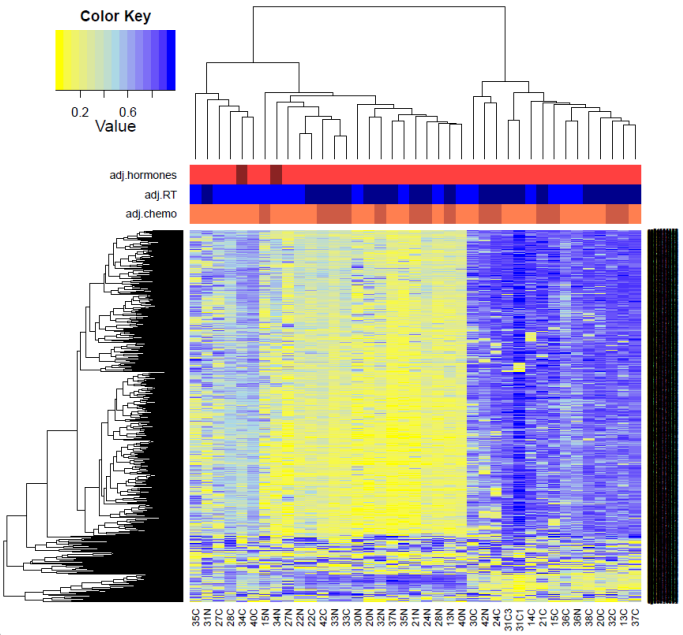
Cervical invasion
No: violet; Yes: violetred

Lymphovascular invasion
No: tan; Yes: dark tan

Myometrial invasion>50%
No: grey; Yes: dark grey

Lymph node positive
No: pink; Yes: dark pink

Size
<2cm: light green; ≥2cm: dark green



D

Adjuvant hormones
No: brown; Yes: dark brown

Adjuvant RT
No: blue; Yes: dark blue

Adjuvant chemotherapy
No: coral; Yes: dark coral

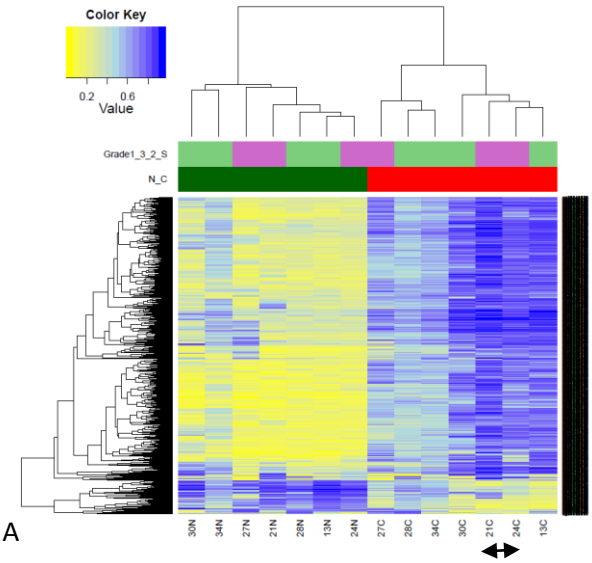
In the NvC analysis with all samples in Figure 5.7 above, there was limited correlation between clinicopathologic factors and differential methylation status. A group of patients aged ≥65 years and a group with tumour size ≥2cm appeared to cluster together associated with hypermethylation in EC samples, but otherwise there was no apparent clustering based on stage, tumour features, disease recurrence nor adjuvant treatment received.

In the following Figure 5.8 showing the matched NvC comparison, there is a trend between clinicopathologic features and differential methylation. Patients with grade 3 tumours and age ≥65 years appear to have hypermethylation in the cancer tissue compared to lower grade

and patients <65years. Tumour size ≥2cm and the presence of LVSI also seem to correlate with greater hypermethylation in cancer specimens. In terms of adjuvant treatment, those who received adjuvant chemotherapy and RT appeared to have tumours with greater hypermethylation, potentially related to these being larger tumours with LVSI. Only one patient had recurrent disease so the difference in methylation between those who did and did not recur cannot be assessed here.

Figure 5.8: Clinicopathological data and Methylation Status of matched Normal Endometrium v Endometrioid Endometrial Cancer samples, A: grade, B: stage and age, C: tumour features, D: primary treatment

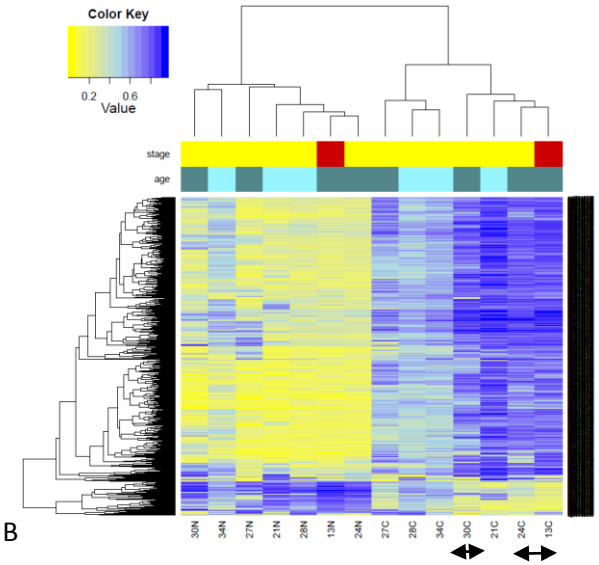
Some grade 3 specimens and those with patients' age ≥65 years showed hypermethylation in the cancer tissue compared to specimens of lower grade and with patients' age <65years (A and B respectively). Specimens with tumour size ≥2cm and LVSI also showed hypermethylation in cancer specimens (C). In terms of adjuvant treatment, those who received adjuvant chemotherapy and RT appeared to have tumours with greater hypermethylation (D). These samples are marked with black arrowheads.



A

Grade 1: green
Grade 3: purple

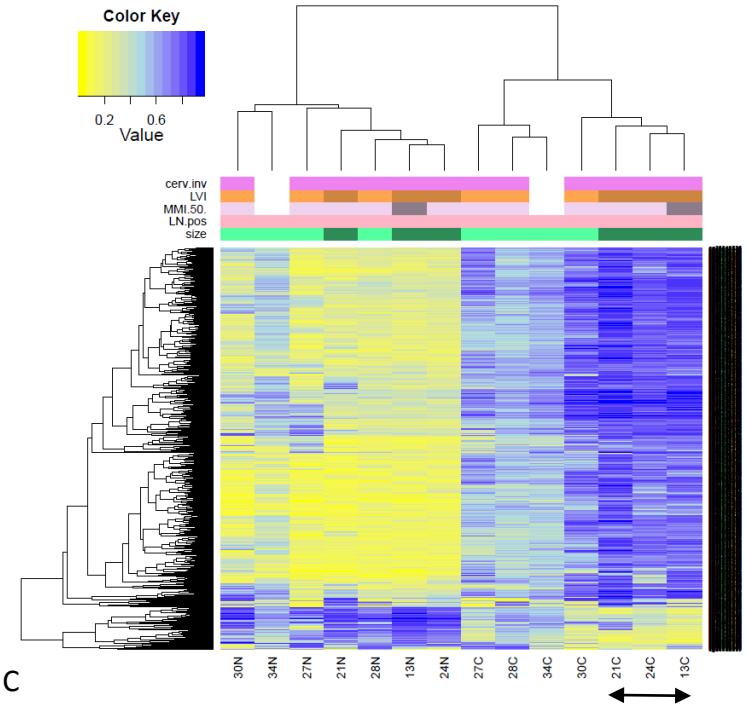
Normal: dark green
Cancer: red



B

Stage I: yellow
Stage II: plum
Stage III: red

Age
<65yrs: light blue
≥65yrs: olive



C

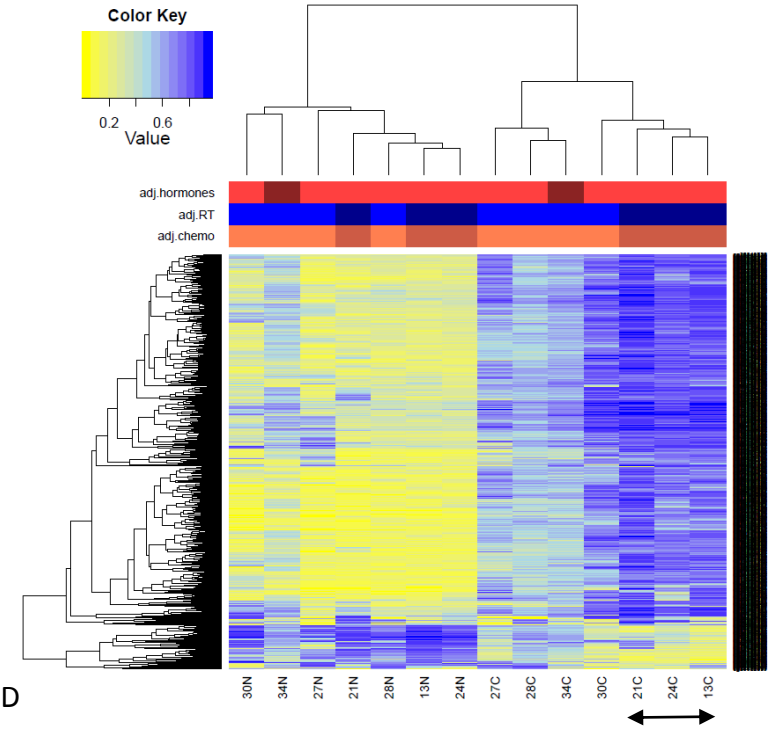
Cervical invasion
No: violet; Yes: violetred

Lymphovascular invasion
No: tan; Yes: dark tan

Myometrial invasion>50%
No: grey; Yes: dark grey

Lymph node positive
No: pink; Yes: dark pink

Size
<2cm: light green; ≥2cm: dark green



D

Adjuvant hormones
No: brown; Yes: dark brown

Adjuvant RT
No: blue; Yes: dark blue

Adjuvant chemotherapy
No: coral; Yes: dark coral

5.2.3. Correlation of Study Data with The Cancer Genome Atlas

The methylation data from the normal endometrium, AEH and EEC samples in this study were compared with what was publicly available through TCGA research network on normal endometrium and cancer tissue. There were no atypical specimens available for comparison on TCGA.

Using a search for healthy endometrial tissue and uterine corpus endometrioid carcinoma, data for 60 healthy and 383 EC samples were obtained from TCGA. For the purpose of the comparison and processing restrictions on the R console with too great a sample number, 50 healthy and 150 EEC samples from TCGA were compared to the study normal, atypical and cancer specimens. Overall, there appeared to be a correlation between the study normal and cancer samples and the TCGA data.

In the normal vs cancer comparison in Figure 5.9, there were 2 separate clusters of cancer specimens from the TCGA data with those on the right side showing greater hypermethylation compared to the normal tissue, than those on the left. There is also a subset of probes between the normal and cancer tissue that are hypomethylated (purple arrow).

The majority of study samples clustered with the corresponding TCGA histological tissue type, except specimens 22C, 33C and 42C that clustered with the normal samples towards the left of the HM, while 36N and 42N clustered with the cancer specimens towards the right of the HM, consistent with earlier assessment of these specimens in this chapter (see Figure 5.2A). All other normal specimens clustered with the TCGA normal specimens in the centre. Specimens 14C, 15C, 20C, 31C, 32C, 37C, 38C lay closer to the first cluster of cancer specimens on the left, while specimens 13C, 21C, 24C, 27C, 28C, 30C, 34C, 35C, 36C, 40C lay within the cluster of cancer specimens on the right.

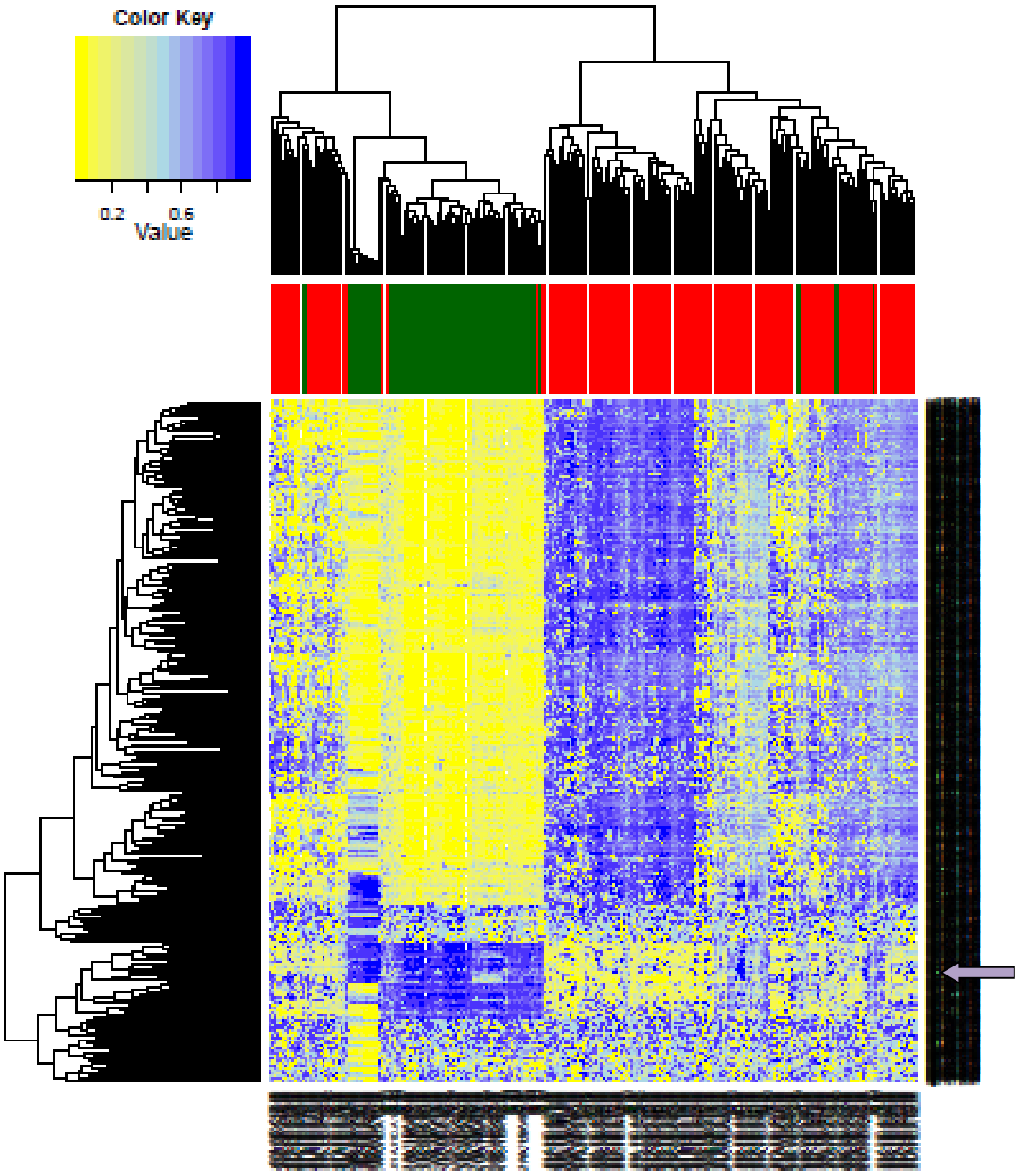
Comparison of normal endometrial, AEH and EEC specimens with TCGA data is detailed in Figure 5.10, and demonstrates that the majority of the AEH specimens cluster with the TCGA cancer samples towards the right side of the HM. Within the TCGA cancer specimens, the spectrum of methylation changes is again demonstrated. Those in the centre show an intermediate level of hypermethylation compared to the normal tissue, while those on the right show greater hypermethylation. Similar to the study data, there is also a set of probes that are hypomethylated in cancer tissue compared to the normal tissue, demonstrated towards the bottom of the HM (purple arrow).

In terms of the atypical samples, 24A, 30A and 34A lay within the normal sample cluster, again consistent with earlier study data in this chapter. Otherwise, 4A, 13A, 21A, 27A, 28A lay within

the cancer cluster of specimens and interestingly, are intermediate in their methylation status, as are many of the TCGA cancer specimens.

This division of atypical specimens between those clustering with normal and those clustering with cancer specimens again may support the hypothesis that there is a spectrum of methylation changes that occur either early or late in endometrial tumorigenesis. As noted in Figure 5.2B, the cases 24A, 30A and 34A may reflect those samples where methylation changes occur later in tumorigenesis as the normal and atypical specimens have a similar methylation pattern. For cases 13A, 27A and 28A, the methylation changes may occur early and thus, they have a similar pattern to cancer tissue, as previously noted in Figure 5.2C.

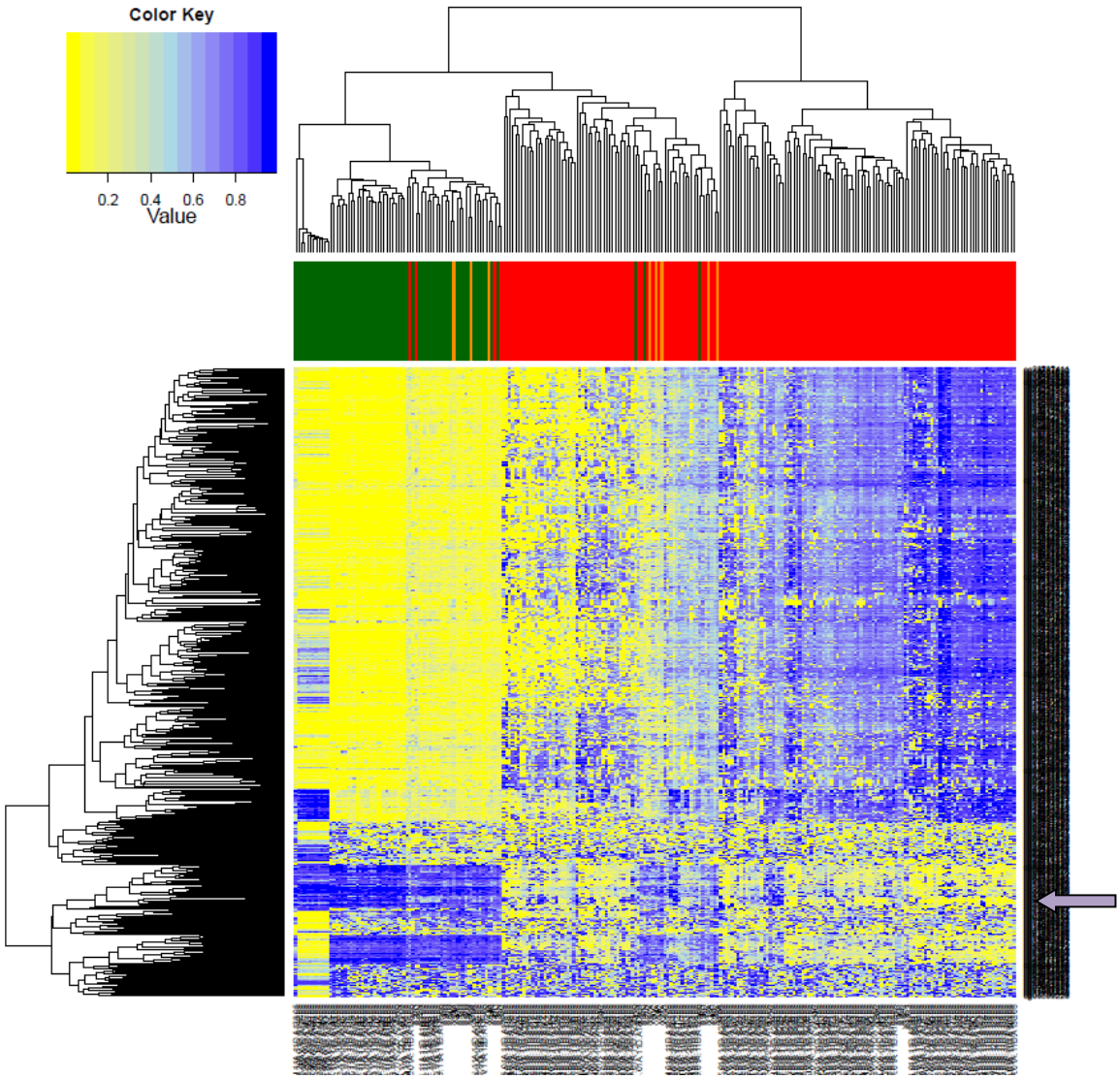
Figure 5.9: Correlation of Differential Methylation for the 500 most variable probes for TCGA and Study Normal Endometrial and Endometrioid Endometrial Cancer specimens



Normal endometrium: dark green; Cancer: red

There are 2 separate clusters of cancer specimens from the TCGA data with those on the right side showing greater hypermethylation compared to the normal tissue, than those on the left. The majority of study samples clustered with the corresponding TCGA histological tissue type, except specimens 22C, 33C and 42C that clustered with the normal samples towards the left of the HM, while 36N and 42N clustered with the cancer specimens towards the right of the HM, consistent with earlier assessment of these specimens in this chapter (see Figure 5.2A). All other normal specimens clustered with the TCGA normal specimens in the centre. Specimens 14C, 15C, 20C, 31C, 32C, 37C, 38C lay closer to the first cluster of cancer specimens on the left, while specimens 13C, 21C, 24C, 27C, 28C, 30C, 34C, 35C, 36C, 40C lay within the cluster of cancer specimens on the right.

Figure 5.10: Correlation of Differential Methylation for the 500 most variable probes for TCGA and Study Normal Endometrial, Atypical Endometrial Hyperplasia and Endometrioid Endometrial Cancer specimens



Normal endometrium: dark green; Cancer: red
Atypical endometrial hyperplasia: orange

The majority of the AEH specimens cluster with the TCGA cancer samples towards the centre/right side of the HM. Within the TCGA cancer specimens, the spectrum of methylation changes is again demonstrated. Those in the centre show an intermediate level of hypermethylation compared to the normal tissue, while those on the right show greater hypermethylation. Similar to the study data, there is also a set of probes that are hypomethylated in cancer tissue compared to the normal tissue, demonstrated towards the bottom of the HM (purple arrow).
In terms of the atypical samples, 24A, 30A and 34A lay within the normal sample cluster, again consistent with earlier study data in this chapter. Otherwise, 4A, 13A, 21A, 27A, 28A lay within the cancer cluster of specimens and interestingly, are intermediate in their methylation status, as are many of the TCGA cancer specimens.

5.2.4. Pathways and Genes affected by Differential Methylation

As only the NvC comparison yielded a significant number of MVPs and DMRs (see Table 5.1), only the associated molecular pathways and genes from this analysis could be identified. From the 23572 MVPs and 706 DMRs generated from the NvC comparison, a list of associated probes and genes was generated from the ChAMP pipeline and analysed through WebGestalt (described in section 2.2.6.1). It was assessed whether genes that were differentially methylated were related to cancer pathways or not. For the NvA and AvC comparisons, although the heatmaps (Figure 5.2) demonstrated that differential methylation occurred between the histological types, there were no significant DMRs that could be analysed in terms of specific molecular pathways.

On WebGestalt, different enrichment analyses are available. KEGG and Pathway Commons analyses were assessed as most appropriate and used for this study. The KEGG analysis generated a list of genes associated with pathways in cancer, as well as other non-cancer related pathways while the Pathway Commons analysis gave greater detail specifically on the cancer pathways.

The results of the KEGG and Pathway Commons analysis are detailed in Table 5.4.

Table 5.4: Results of the KEGG and Pathway Commons analyses on WebGestalt for the DMRs, hypomethylated and hypermethylated probes, along with the associated genes
Cancer and metabolic pathways of interest are shown below.

Analysis	Pathway Names	No. of genes	P value	Examples of genes involved
KEGG analysis for DMRS	Pathways in cancer Cell adhesion molecules Focal adhesion Type I DM	12 11 11 5	0.0008 8.93e-06 0.0003 0.0008	RARA, GLI2, AKT1, COL4A2, DCC, TGFB3, RASSF1, CTBP2, WNT16, FZD5, ETS1, CTBP1
Pathway commons analysis for DMRs (cancer related)	PDGFR SP c-met/HGFR SP EGFR SP PI3K/mTOR/Arf6 pathway Plasma membrane ER SP VEGFR1 & VEGFR2 SP FAK SP E-cadherin SP p53 SP ALK1 SP Notch SP TGFB R SP Trk R SP	36 36 36 36 37 36 36 11 8 10 5 9 4	1.55e-07 1.55e-07 1.55e-07 1.55e-07 1.55e-07 1.55e-07 1.55e-07 0.0004 0.0021 0.0043 0.0073 0.0090 0.0092	CARD11, TP53AIP1, DUSP6, RASSF1, CD8B, ETS1, HSPA1A, SKI, CFLAR, RPTOR, TP73, TJP2, SHC1, GATA2, MMP14, DNM2, H2AFY, RIN1, RIN2, TGIF1, AKT1, PTPN6, COL1A1, KARLN, PENK, RUNX3, HIC1, MEF2D, KIFC3, THY1, GNA12, PRKCZ, SERPINB5, FZD5, CTBP1, SFN
KEGG analysis for hypomethylated probes	MAPK SP Wnt SP Metabolic pathways Focal adhesion	2 2 2 2	0.0431 0.0227 0.0227 0.0289	DUSP6, MAPK8IP3, TBL1XR1, AXIN1, CARD11, PTPN6, NT5C3, HGD, GALNT6, SCLY SPP1, PARVG, PDE6B, HLA-DMB

Pathway commons analysis for hypomethylated probes	c-met/HGFR SP PI3K/mTOR SP EGFR SP PDGFR SP FAK SP E-cadherin SP VEGF & VEGFR SP Plasma membrane ER SP	9 9 9 10 9 5 10 9	3.38e-05 3.38e-05 3.38e-05 3.38e-05 3.38e-05 3.38e-05 3.38e-05 3.38e-05	SPP1, CARD11, DUSP6, TBL1XR1, AXIN1, PTPN6, MAPK8IP3, AQP5, PTPRE, SHB
KEGG analysis for hypermethylated probes	Pathways in cancer Type I DM Cell adhesion molecules Focal adhesion	20 8 13 15	3.19e-05 1.52e-05 1.52e-05 4.59e-05	MMP9, GLI2, DCC, ITGA6, NTRK1, FGF12, RASSF1, FGF5, PRKCA, VEGFA, ETS1, CTNNA2, EGFR, BMP2, WNT4, BMP4, WNT16, WNT3, RUNX1, CTBP
Pathway commons analysis for hypermethylated probes	EGFR SP VEGF & VEGFR SP FAK SP PI3K/mTOR SP PDGFR-β SP HGFR SP ALK1 SP TGFB R SP E-cadherin SP Wnt SP Plasma membrane ER SP	65 64 62 62 62 62 17 16 18 8 63	2.46e-12 5.30e-12 9.88e-12 9.88e-12 9.88e-12 9.88e-12 0.0003 0.0004 1.53e-05 0.0004 9.88e-12	SOX1, WNT4, DAB2, WNT3A, PRKCA, CTBP1, HIPK2, CSNK1E, MMP9, BIN1, ITGA6, ROBO1, RASSF1, PBX1, NDRG2, NEDD4L, ETX1, ITGB5, FOXG1, HSPA1A, HLA-A, SKI, BMP2, TJP2, PRKCE, GSC, RPTOR, TP73, IRF4, ZYX, DAB2, EOMES, SHC1, CTGF, GATA2, DNM2, ACTA2, MDFIC, CXCL12, PRR5, GFRA1, POU4F2, TAOK3, BDNF, HIC1, PENK, OPRM1, VEGFA, EGFR, PITX2, THY1, GNA12, TERT, HIPK2, ACTA1, PRKCZ, TLE1, MYOD1, BMP4, RUNX1, PPP1R13L, CYTH2, FOXA1, RGMA

No.: number, Ag: antigen, R: receptor, DM: diabetes mellitus, SP: signalling pathway.
Other gene/pathway abbreviations are outlined in the Abbreviations section.

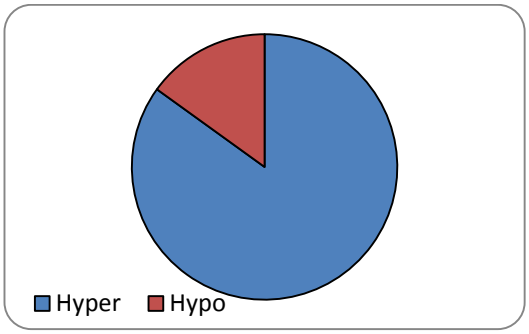
There are a number of cancer related genes and pathways that were identified from the enrichment analyses of the DMRs, as well as pathways involving ER signalling and diabetes that also may be implicated in endometrial carcinogenesis. The pathways demonstrated include those that are well-documented in EC including PI3K/mTOR/Arf6, VEGF, EGFR, PDGFR, TGFB, p53 and the Wnt/E-cadherin signalling pathways. However, there are also a number of pathways that are novel or only recently described in endometrial carcinogenesis including FAK and cell adhesion, ALK, c-met/HGFR, NOTCH/FOXA1 and Trk signalling pathways [264-267]. The genes listed above interact with a number of these pathways rather than being exclusive to one, reflecting the nature of interacting networks that drive tumorigenesis.

From the genes listed in Table 5.4 above, the following parameters were reviewed:

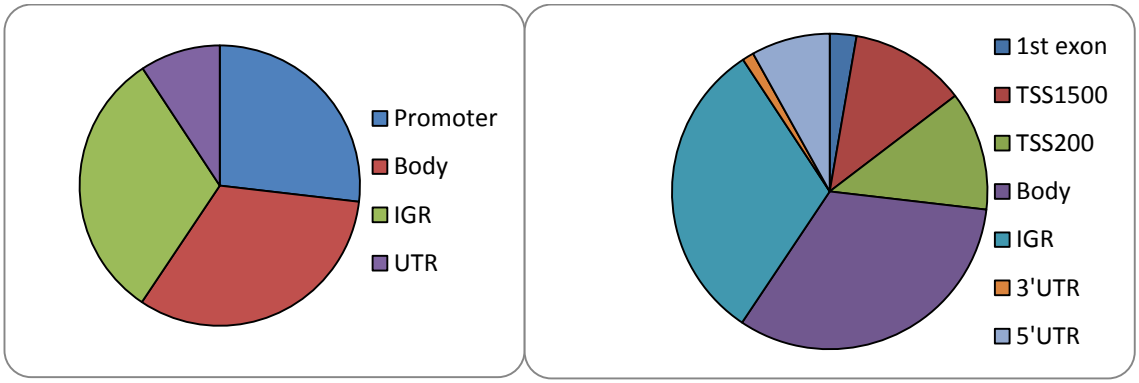
- The frequency of hypermethylated and hypomethylated probes, the genomic distribution and CpG content of the DMRs (detailed in Figure 5.11).
- For those genes with methylation changes in the promoter region or gene body and located in the CGI or CpG shore (as per Table 5.5), the differential changes were reviewed in greater detail in section 5.2.3, along with the gene function where known.

Figure 5.11: The distribution of Differentially Methylated Regions (DMRs) associated with
A: Hypermethylated and hypomethylated probes
B: Genomic distribution
C: CpG content of the DMRs

A: Hypermethylated and hypomethylated probes in DMRS



B: Genomic distribution of the DMRS



IGR: intergenic region, UTR: untranslated region, TSS: transcription start site

C: CpG content of the DMRs

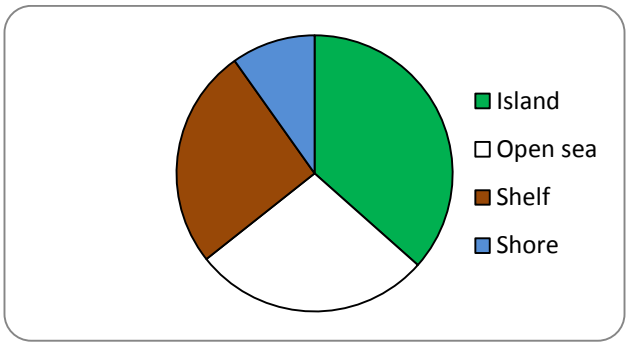


Figure 5.11 above shows that the majority of DMRs involve hypermethylation of the gene promoter or body. There is already evidence in the setting of carcinogenesis that promoter hypermethylation of a tumour suppressor gene may lead to decreased gene expression while promoter hypomethylation of an oncogene may lead to its increased expression [262]. Thus, the function of the implicated genes and their location within the gene and CpG content is further detailed in Table 5.6.

Figure 5.11 also shows that the DMRs were most commonly found in CGIs. CGIs are typically found in the promoter region of genes and CGI promoter methylation is implicated in carcinogenesis [268]. As there is data to suggest that DNA methylation in CpG shores and in the gene body can also be of significance in carcinogenesis [189, 194], changes in both CGIs and the CpG shore within the promoter region and gene body were reviewed further. Table 5.5 details the genomic distribution and CpG content of the genes listed in Table 5.4, generated from the Web Gestalt analysis, particularly if located within a CGI or CpG shore.

Table 5.5: Genomic distribution and CpG content of Selected Genes from the DMRS, hypermethylated and hypomethylated regions

	Promoter (1st exon, TSS200, TSS1500)	Body	5'UTR	3'UTR
Island Hypomethylated	SFN (S)	PTPRE		
Hypermethylated	RASSF1, NDRG2, EOMES, HSPA1A, TP73, GATA2, FGF2 (S), NRG3 (S), MYL9 (S), ITGA8 (S), MADCAM1 (S), NCAM2, PENK, HIC1, SFRP2, SOX1, FZD10, RUNX3	DCC (shelf), FGF12, NTRK1, CTNNA2, MMP9, GSC, RUNX1, PRKCZ, CD8B (S), GNA12 (S), PRR5, POU4F2,	BMP4	
Shore Hypomethylated	AQP5 (S), CLDN4 (I), CFLAR, PTPN6	DUSP6		
Hypermethylated	WNT3A, VEGFA, WNT16, FZD5, H2AFY, MMP2, THY1, TJP2	GLI2, PBX1, FGF5, ETS1, BMP2, CTGF,	CTBP1, CSNK1E	MMP14
Shelf	DCC (S), RIN1 (S), SHC1, TGIF1 (S)	RPTOR, ZYX, SHC1, TGFB3, AXIN1	MEF2D	AKT1
Not specified	COL11A1, SERPINB5, TP53AIP1, KARLN, APC, FGF10, LAMA3, CARD11, SPP1	PRKCA, ITGA6, SKI, EGFR, WNT4, DNM2, HIPK2, HLA- CDH4, RARA, COL4A2, MAPK8IP3	TBL1XR1, PARVG, CTBP2, NLGN1	CNTNAP1

S: shore, I: island. If there is >1 CpG content, the 2nd is indicated in brackets.

5.2.5. Differential Methylation within Candidate Genes

The genes generated through the WebGestalt analysis and located within the promoter region or gene body with CGI or CpG shore content, as detailed in Table 5.5, were then analysed further on the R console, as described in Section 2.2.6.1 (script 4 in appendix 2D).

This further analysis between the normal endometrial, AEH and EEC specimens assessed specific probes within a gene and the spectrum of change in methylation status that occurred between the 3 histological subtypes. Data was generated numerically and diagrammatically detailing where the greatest change in methylation status occurred based on changes in the beta values.

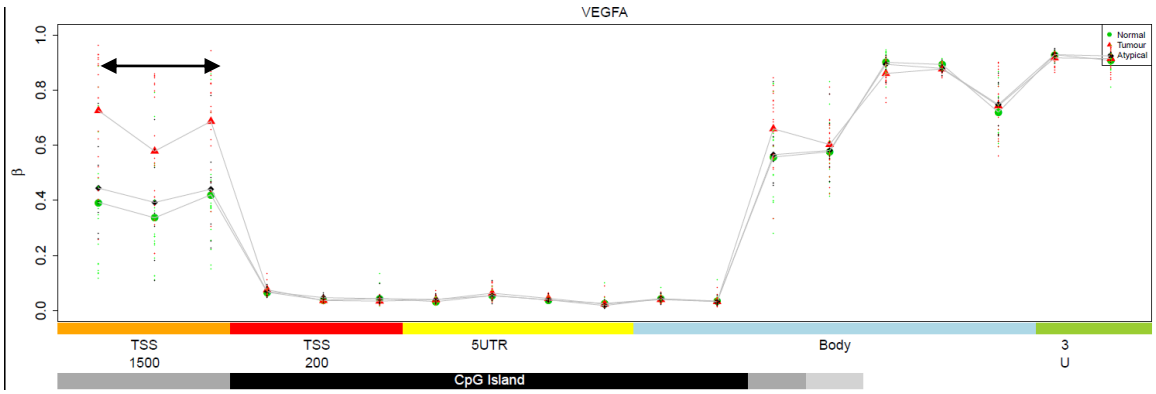
These difference in gene methylation between normal endometrium, AEH and EEC specimens for 10 genes is illustrated diagrammatically in Figure 5.12.

Figure 5.12: Diagrams demonstrating differentially methylated probes (indicated by black arrows), their genomic distribution and CpG content in specific genes involving, A: VEGF, B: MAPK, C: Wnt, D: FGF/FGFR, E: tumour suppressor genes, F: p53 pathways

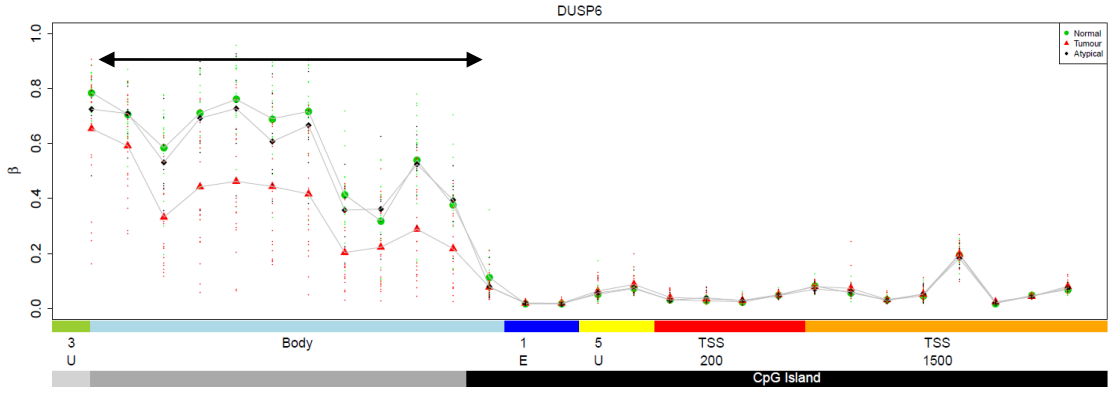
The y-axis shows the degree of methylation of individual probes, as illustrated by the beta value. A beta value of 0 equals non-methylation at the locus and 1 equals total methylation. The x-axis shows both the genomic distribution (promoter region, gene body, intergenic region) and the CpG content (island, shelf, shore, open sea).
For each probe, the beta value of the normal endometrium (green), AEH (black) and EEC (red) specimens is plotted, such that progressive methylation changes can be visualised along with the genomic distribution and CpG content.

All the genes show areas of progressive hypermethylation between normal endometrium, AEH and EEC, except for DUSP6 and TP53AIP1, which show progressive hypomethylation. The hypermethylation within VEGFA, WNT16, FZD5, FGF2, FGF10, DCC, RASSF1 and TP73 all occur within or including the promoter region. The hypomethylation of DUSP6 and TP53AIP1 occurs in the gene body and promoter region respectively.

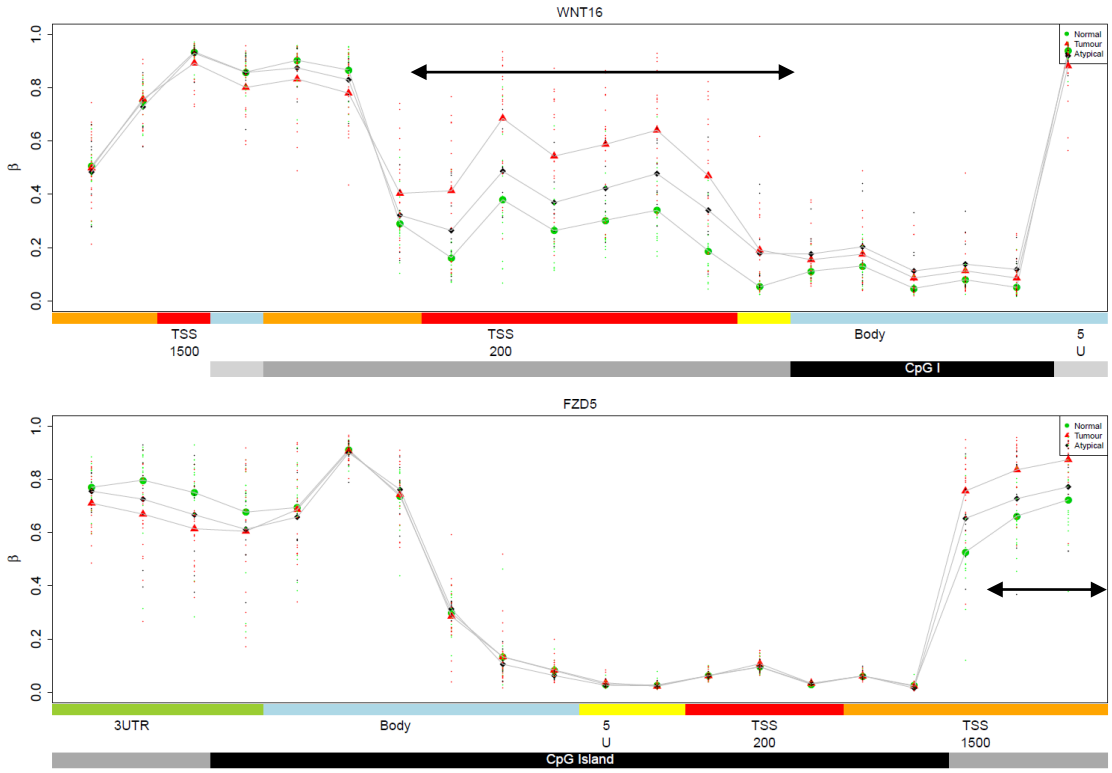
A: VEGF



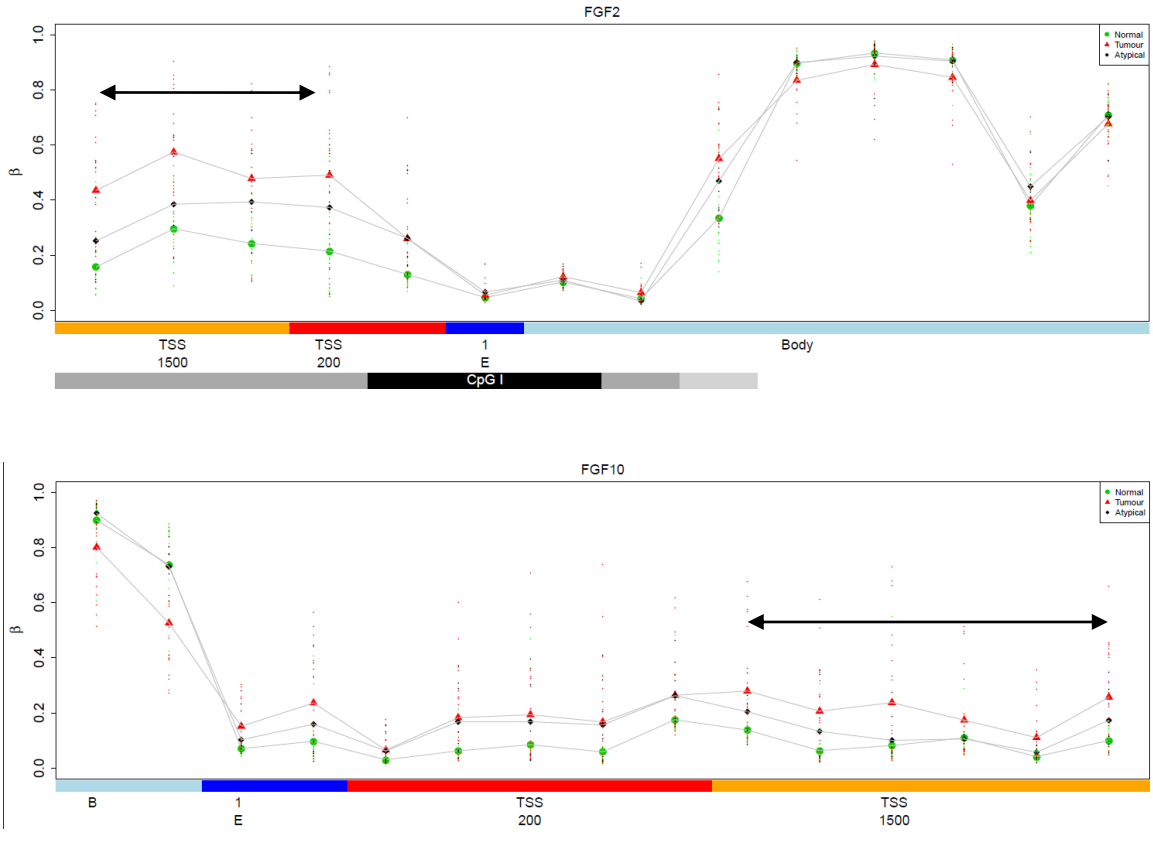
B: MAPK



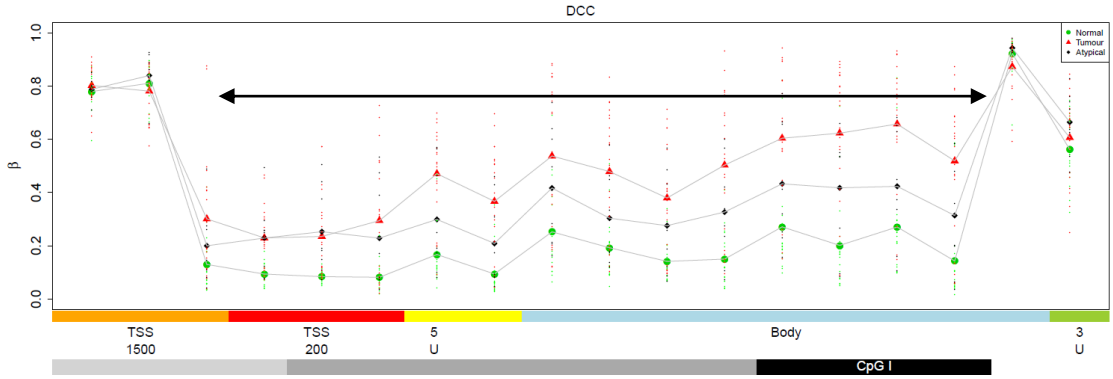
C: Wnt

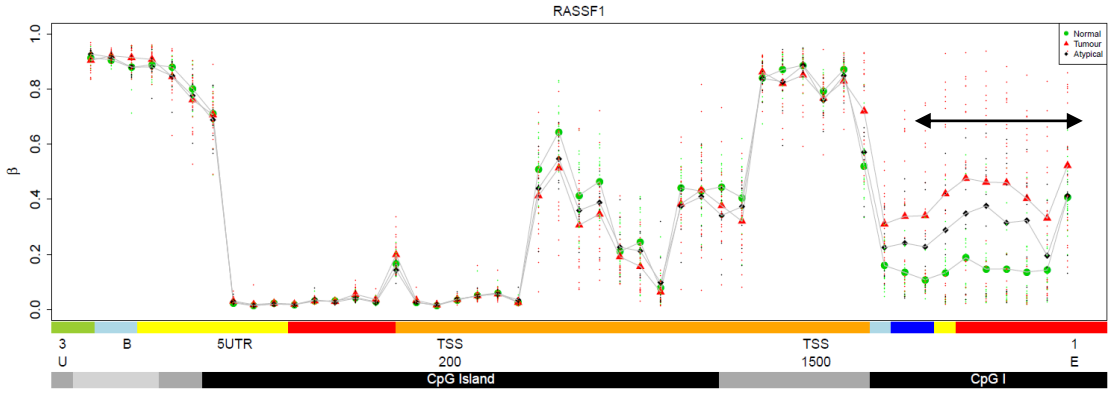


D: FGF/FGFR

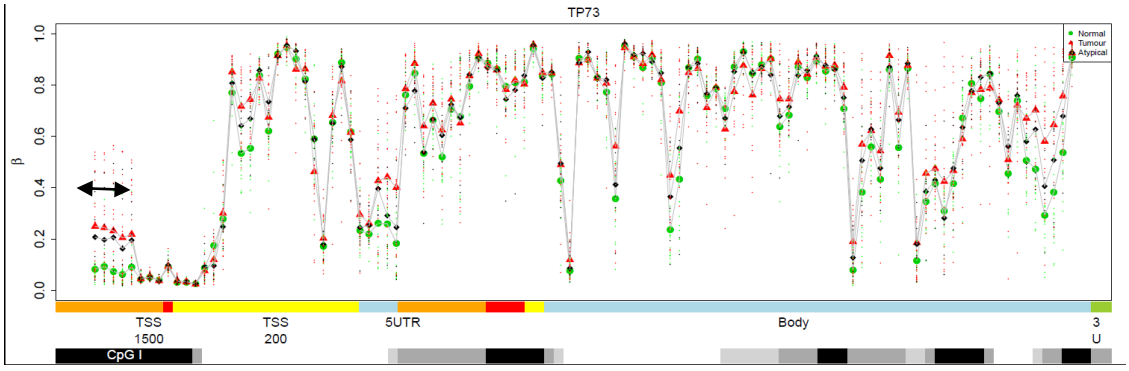


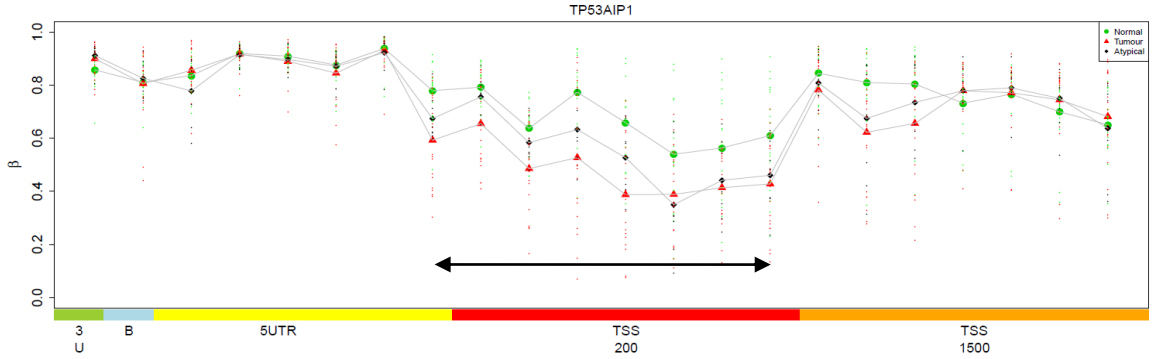
E: Tumour suppressor genes





F: p53





Using the associated numerical data generated in conjunction with the diagrams from Figure 5.12, the probes with differential methylation between normal endometrium, AEH and EEC could be identified and the beta values analysed to determine firstly, whether the methylation changes between normal endometrium and EEC were statistically significant and secondly, whether the changes were more significant between the NvA or between the AvC transitions. The hypermethylated and hypomethylated genes with statistical significance are presented in Tables 5.6 and 5.7 respectively, along with the genomic distribution, CpG content, the number of probes within the gene that are hyper/hypomethylated and the gene function where known.

Table 5.6: Hypermethylated genes, their location and CpG content, associated pathway and significance

Gene	Hyper/ hypo	Genomic distribution	CpG content	Probe no.	Pathway &/or function	NvC p value	NvA p value	AvC p value
Wnt signalling pathway								
WNT3A	hyper	promoter	shore	4	Wnt	0.026	*0.0044*	0.71
SFRP2	hyper	promoter	Island	16	Wnt	0.0014	*0.024*	0.79
SOX1	hyper	promoter	island	23	Wnt TSG [269]	$9.7e^{-07}$	*0.032*	0.052
APC	hyper	promoter	Open sea	12	Wnt TSG	0.0021	*0.22*	0.77
FZD10	hyper	promoter	island	3	Wnt	0.00056	*0.62*	0.95
WNT16	hyper	promoter	shore	7	Wnt	$2.1e^{-06}$	0.046	*0.015*
NDRG2	hyper	promoter	island	3	N-myc Wnt TSG	0.00021	0.16	*0.13*
FZD5	hyper	promoter	shore	3	Wnt	$1.16e^{-05}$	0.23	*0.12*
FGF/FGFR signalling pathway								
FGF10	Hyper	promoter	Open sea	12	FGF	0.0014	*0.12*	0.22
FGF12	hyper	body	Island	15	FGF	0.0014	*0.089*	0.21
FGF2	Hyper	Promoter	Shore & island	4	FGF	$4.6e^{-06}$	0.063	*0.051*
P53 signalling pathway								
TP73	hyper	promoter	island	5	p53 TSG	0.00062	*0.095*	0.49
VEGF signalling pathway								
VEGFA	hyper	promoter	shore	3	VEGF,angio genesis	$9.3e^{-05}$	0.59	*0.0065*
Other signalling pathways								
NRG3	hyper	promoter	Island & shore	9	Neuregulin ERBB4	$4.7e^{-05}$	*0.030*	0.11
MMP2	hyper	promoter	shore	4	MMP	0.00026	*0.084*	0.36
SHC1	hyper	promoter	shelf	3	MAPK	$1.3e^{-06}$	*0.022*	0.055
TGFB3	hyper	body	shelf	3	TGFB	0.00028	*0.11*	0.12
GLI2	hyper	body	shore	3	SHH	$6.6e^{-05}$	0.51	*0.011*

RIN1	hyper	promoter	Shore & shelf	6	MAPK	0.00028	0.23	*0.026*
RARA	hyper	Body	Open sea	10	Retinoic acid receptor	$4.7e^{-06}$	0.15	*0.0012*
Other tumour suppressor genes								
RASSF1A	hyper	promoter	island	10	RASSF1 TSG	0.00039	*0.096*	0.19
DCC	hyper	Promoter body	Shore & shelf Island & shelf	4 8	CAM TSG	$0.39e^{-07}$	*0.044*	0.092
HIC1	hyper	promoter	island	6	TSG Represses transcrip'n	$4.18e^{-05}$	0.11	*0.082*
RUNX3	hyper	promoter	island	27	Regulates transcrip'n TSG	0.00029	0.17	*0.079*
Other pathways (e.g. immune, CAM, transcription regulation)								
COL11A1	hyper	promoter	Open sea	5	Collagen	0.00014	*0.075*	0.18
ITGA8	hyper	promoter	Island & shore	8	Cell-cell interaction	$3.6e^{-05}$	*0.047*	0.066
CD8B	hyper	body Promoter	Island & shore	5	Tcell/immune	$4.6e^{-05}$	0.32	*0.077*
CARD11	hyper	promoter	Open sea	4	CARD/MAGUK	0.00041	0.97	*0.0039*
GATA2	hyper	promoter	island	11	Regulates gene transcrip'n	0.012	0.44	*0.10*
H2AFY	hyper	promoter	shore	9	Represses transcrip'n	0.00035	0.12	*0.10*
KARLN	hyper	Promoter	Open sea	8	Neuronal regulation	$1.9e^{-05}$	0.13	*0.030*
MADCAM1	hyper	promoter	Island & shore	4	Cell adhesion leucocyte receptor	$1.2e^{-05}$	0.21	*0.11*
MYL9	hyper	promoter	Island & shore	9	myosin	$8.6e^{-05}$	0.51	*0.0019*
NCAM2	hyper	promoter	island	6	CAM	$3.1e^{-05}$	0.26	*0.0076*
PENK	hyper	promoter	island	17	Enkephalin	$3.2e^{-07}$	0.077	*0.017*
TGIF1	hyper	promoter	Shore & shelf	4	Regulates transcrip'n	$1.3e^{-06}$	0.14	*0.0033*
THY1	hyper	promoter	Shore	5	Neural cell-cell interaction	0.00055	0.92	*0.0092*

TSG: tumour suppressor gene, CAM: cell adhesion molecule, Transcrip'n: transcription.
Other abbreviations are outlined in the Abbreviations section.
P values are italicised to indicate whether the NvA or AvC p value is smaller

Table 5.7: Hypomethylated genes, their location and CpG content, associated pathway and significance

Gene	Hyper/ hypo	Genomic distribution	CpG content	Probe number	Pathway &/or function	NvC p value	NvA p value	AvC p value
P53 signalling pathway								
TP53AIP1	hypo	promoter	Open sea	10	p53	0.00078	*0.027*	0.22
SFN	hypo	promoter	Island/ shore	11	p53 PI3K	$3.79e^{-05}$	0.12	*0.0076*
Other pathways or TSGs								
SERPINB5	hypo	Promoter	Open sea	8	TSG	$5.87e^{-0.5}$	0.31	*0.0021*
DUSP6	hypo	body	shore	10	MAPK	2.01e-^{05}	0.61	*0.0014*
Other cellular roles								
CFLAR	hypo	promoter	shore	4	Apoptosis regulator	$1.7e^{-05}$	*0.014*	0.15
CLDN4	hypo	promoter	Island/ shore	7	Epithelial cell tight junction	0.0002503	0.069	*0.031*
AQP5	hypo	Promoter and body	Island/ shore	3	Water channel protein	0.00079	0.12	*0.056*
LAMA3	Hypo	Promoter	Open sea	6	CAM	0.00019	0.079	*0.047*
PTPN6	hypo	promoter	shore	7	Protein tyrosine phosphatase	$1.4e^{-05}$	0.12	*0.016*
SPP1	hypo	promoter	Open sea	3	osteopontin	$7.3e^{-05}$	0.033	*0.0091*
TJP2	hypo	promoter	shore	3	Tight junction formation	$5.1e^{-05}$	0.17	*0.028*

TSG: tumour suppressor gene, CAM: cell adhesion molecule.
Other abbreviations are outlined in the Abbreviations section.
P values are italicised to indicate whether the NvA or AvC p value is smaller

The data above illustrate that a number of genes in pathways known to be abnormal in endometrial tumorigenesis demonstrate methylation changes not only in the NvC comparison but also in that early NvA transition. From the WNT pathway, 5 from 8 genes underwent hypermethylation with the NvA comparison having a smaller p value, suggesting this transition may be of greater significance than what occurs in the AvC comparison. Similarly, from the FGF signalling pathway, 2 from 3 genes underwent hypermethylation, with the NvA comparison having a smaller p value.

Of note, promoter hypermethylation in RASSF1A and SFRP2 is consistent with the literature [206, 221]. For both these genes, the p value for hypermethylation between the NvC comparison was significant. In addition, the p value for the NvA comparison was smaller than the p value for the AvC comparison, but was greater than 0.01. Again, this may suggest that important methylation changes in these genes occurs early in endometrial tumorigenesis from the normal endometrium to AEH transition.

Otherwise, the timing of methylation changes, whether the NvA or AvC transition was of greater significance, was similar in other gene pathways. From the p53 pathway, the p value was smaller for the NvA comparison for hypermethylation of the TP73 gene and hypomethylation of the TP53AIP1 gene. For the SFN gene, the p value was smaller for the AvC comparison. Similarly, in other molecular pathways and tumour suppressor genes, there were a similar number of genes where the p value was smaller either in the NvA or AvC comparison.

In addition to the data in Tables 5.6 and 5.7, the difference in methylation status and beta values between normal endometrium, AEH and EEC in individual genes was analysed diagrammatically. A boxplot could be generated for each gene to visualise whether the greater change in methylation and beta value intensity occurred between the NvA or the AvC transition. Figure 5.13 details the associated boxplots with a diagrammatic representation of these relative methylation changes for a number of the genes from Tables 5.6 and 5.7. Further boxplots for the other genes are included in the Appendix section 5C.

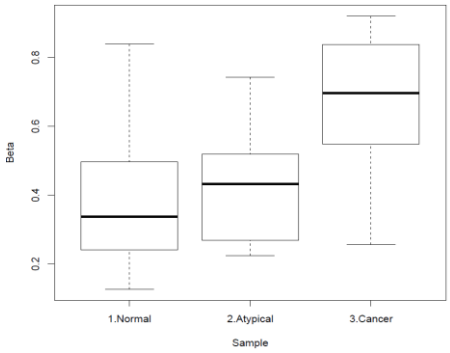
Figure 5.13: Boxplots illustrating 10 hypomethylated or hypermethylated genes between normal endometrium, atypical endometrial hyperplasia and endometrioid endometrial cancer tissue in the pathways involving A: VEGF, B: MAPK, C: Wnt, D: FGF/FGFR, E: tumour suppressor genes, F: p53

As in Figure 5.12, all the genes show progressive hypermethylation between normal endometrium, AEH and EEC, except for DUSP6 and TP53AIP1, which show progressive hypomethylation.

The y-axis shows the degree of methylation, as illustrated by the beta value. A beta value of 0 equals non-methylation at the locus and 1 equals total methylation.

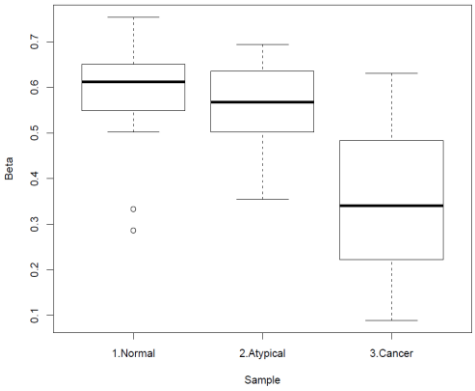
The x-axis shows the histological subtype; normal endometrium, AEH or EEC.

A: VEGF



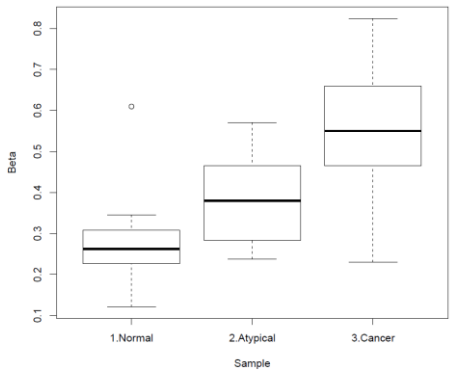
VEGFA

B: MAPK

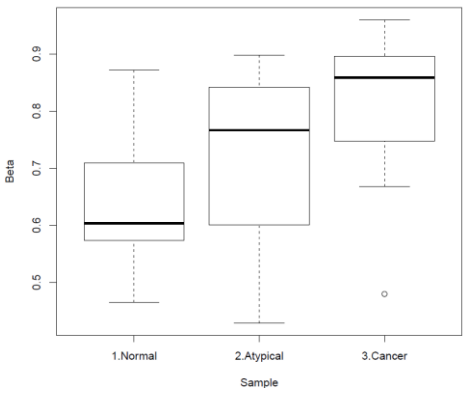


DUSP6

C: Wnt

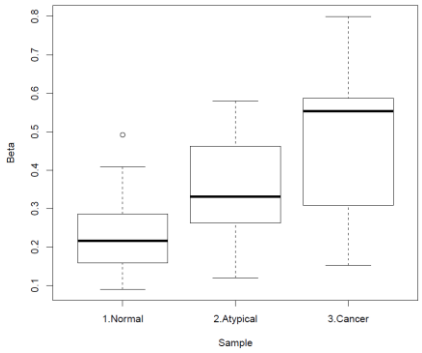


WNT16

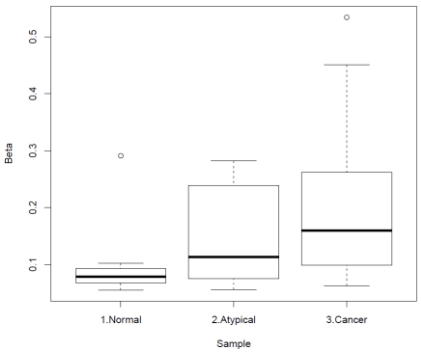


FZD5

D: FGF/FGFR

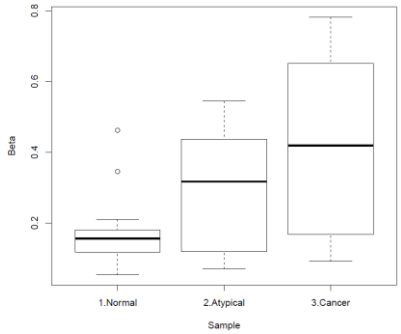


FGF2

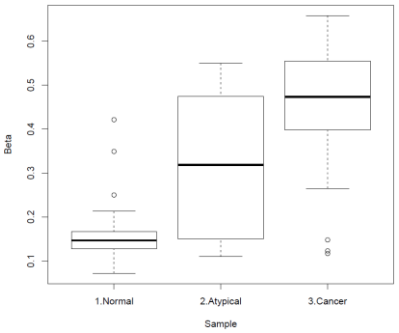


FGF10

E: Tumour suppressor genes

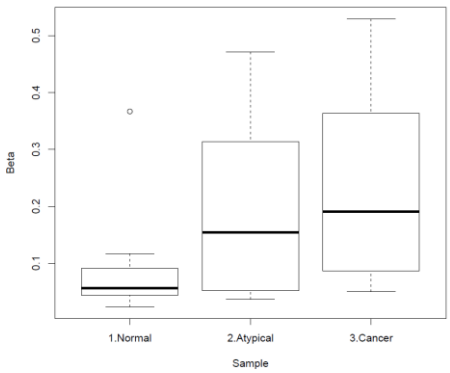


RASSF1

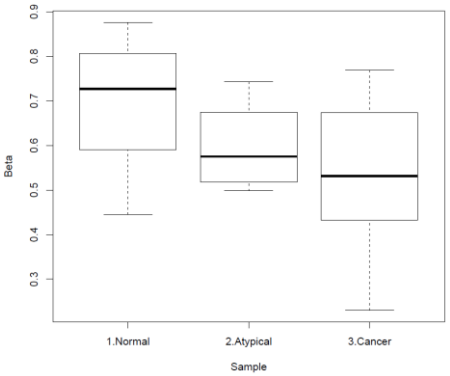


DCC

F: P53



TP73



TP53AIP1

5.3. Results of the Copy Number Variation Analysis

CNV was analysed from the Illumina 450K array data using the ChAMP pipeline on R and the GISTIC2 module, as outlined in section 2.2.6.2. Similar to the methylation analysis for DMRs and MVPs, the CNV analysis through the ChAMP pipeline only allowed 2 groups to be compared at one time and as such, the following comparisons were performed: NvC, NvA and AvC. This allowed an overall assessment of the difference in CNV between normal endometrium and cancer tissue and also looked at whether there was a spectrum of change between the normal endometrial tissue, AEH and EEC.

The number of samples available for analysis after the QC steps in ChAMP was the same as for the methylation analysis and included 39 samples in the NvC, 26 samples in the NvA and 29 samples in the AvC comparisons.

5.3.1. Analysis on ChAMP

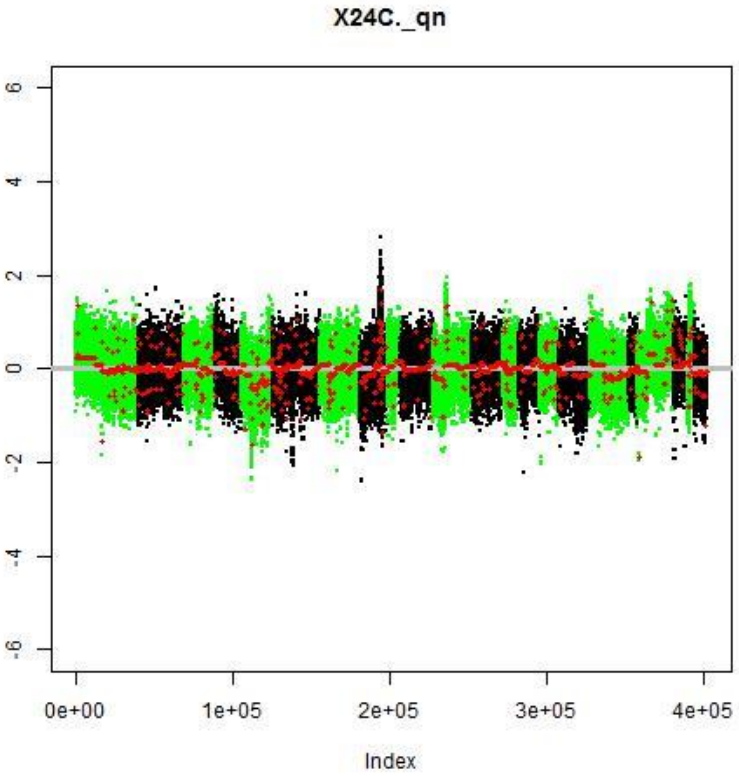
The CNV analysis through the ChAMP pipeline generated figures and raw data based on the CNV per chromosome per specimen in each comparison.

Examples of the individual chromosomal changes generated are illustrated in Figure 5.14.

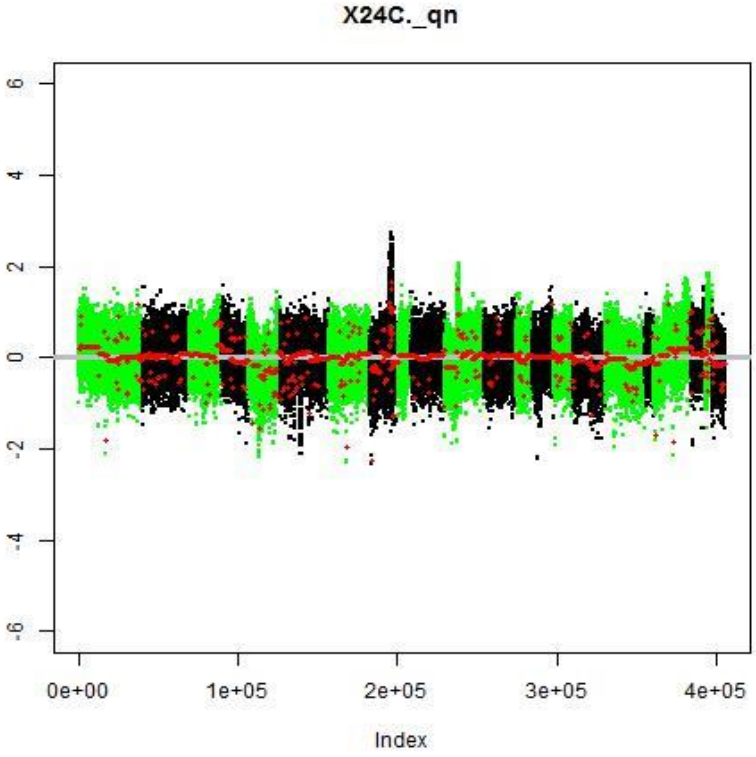
Figure 5.14A. Copy Number Variation per chromosome in sample 24.

For both 5.14A and 5.14B:
The y-axis shows the copy number profile as a $\log^2$ ratio and whether copy number amplification ($\log^2$ ratio >0) or copy number deletion ($\log^2$ ratio <0) occurred.
The x-axis shows each chromosome within a given sample. Odd-numbered chromosomes from 1-21 are shown in green. Even-numbered chromosomes, from 2-22, are shown in black. Segmented copy number is shown in red.

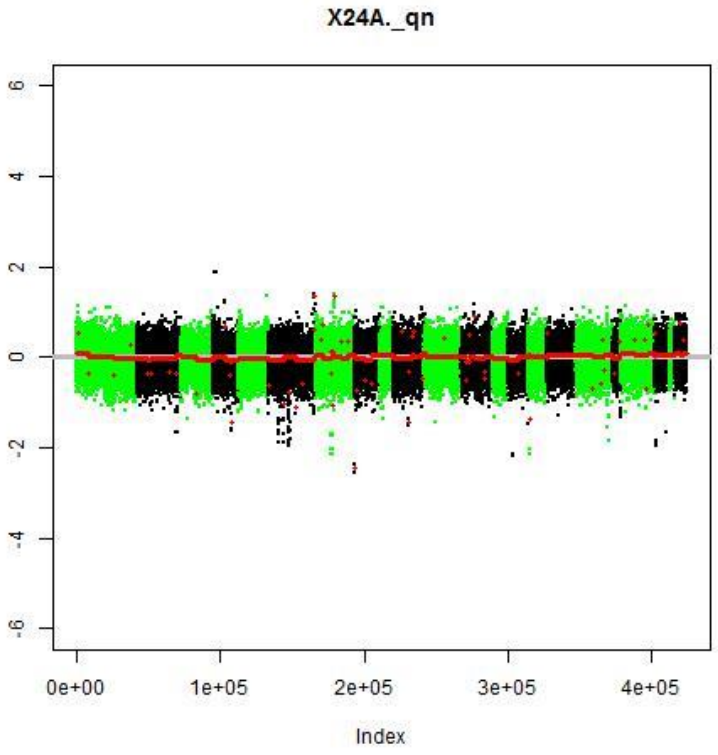
The three figures show the normal endometrium v endometrioid endometrial cancer (NvC), atypical endometrial hyperplasia v endometrioid endometrial cancer (AvC) and normal endometrium v atypical endometrial hyperplasia (NvA) comparisons. The NvC and AvC comparisons show similar copy number changes whereas there is little CNV demonstrated in the NvA comparison, suggesting the CNV occurs later in the normal to atypical to cancer (N-A-C) transition.


X24C._qn
6
4
2
0
-2
-4
-6
0e+00
1e+05
2e+05
3e+05
4e+05
Index

NvC


X24C._qn
6
4
2
0
-2
-4
-6
0e+00
1e+05
2e+05
3e+05
4e+05
Index

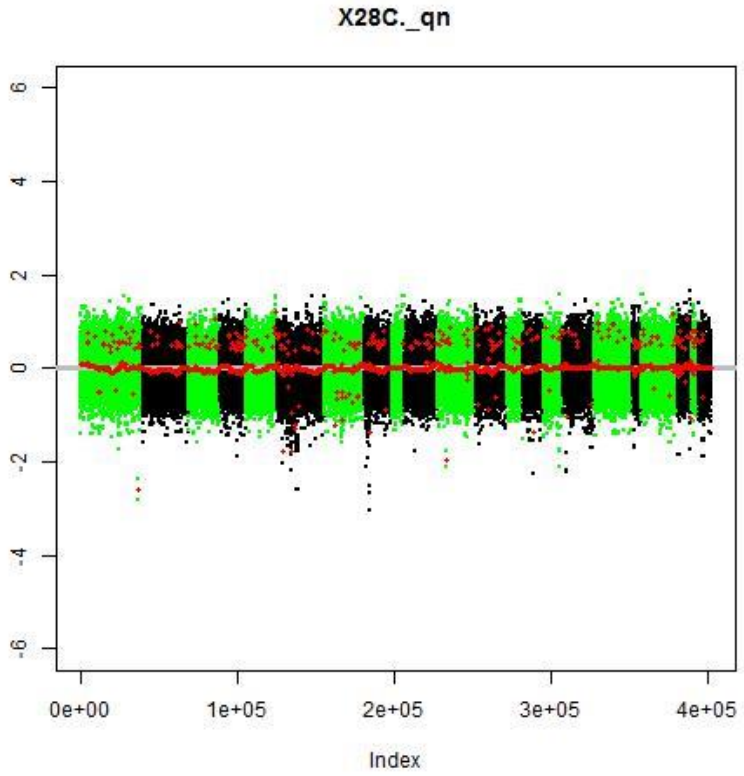
AvC



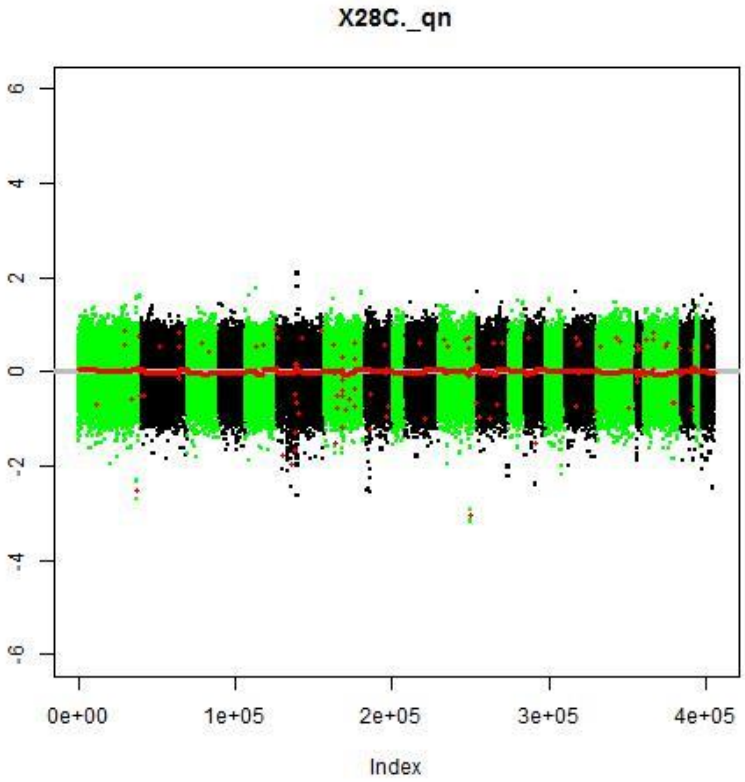
NvA

Figure 5.14B: Copy Number Variation per chromosome in sample 28.

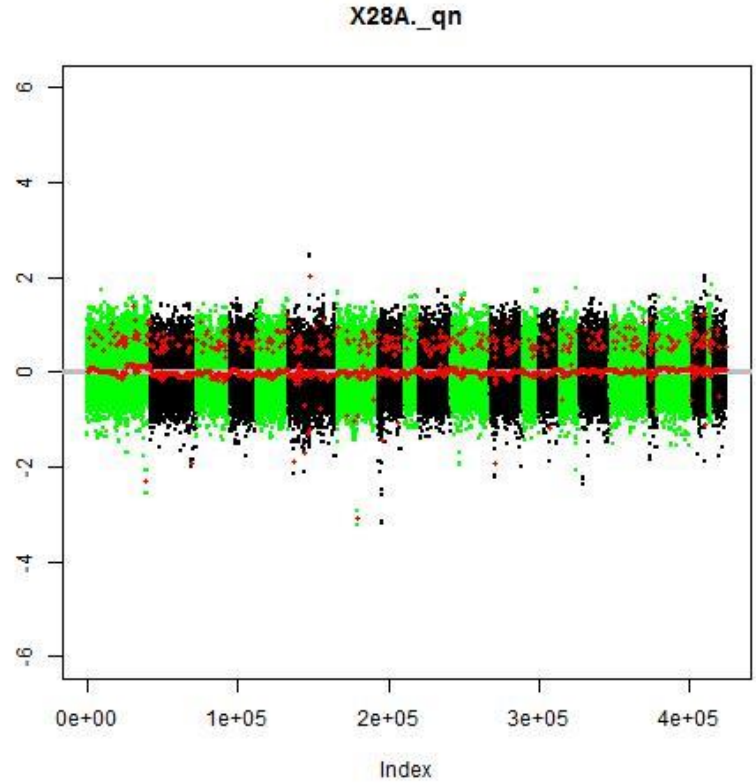
The three figures show the NvC, AvC and NvA comparisons for sample 28. Interestingly, the NvC and NvA comparisons show similar copy number changes whereas there is little CNV demonstrated in the AvC comparison, suggesting the CNV occurs early in the N-A-C transition.



NvC


X28C._qn
6
4
2
0
-2
-4
-6
0e+00
1e+05
2e+05
3e+05
4e+05
Index

AvC


X28A._qn
6
4
2
0
-2
-4
-6
0e+00
1e+05
2e+05
3e+05
4e+05
Index

NvA

5.3.2 Analysis on GISTIC2

The first component of the CNV assessment on the ChAMP pipeline generated both diagrammatic information for each specimen and the copy number changes, as in Figure 5.14, as well as quantitative data. For each specimen, this data included the exact chromosomal location for each CNV, as well as the number of base pairs involved and the degree of separation from the mean.

The second component of the CNV assessment involved processing the data generated through the ChAMP pipeline using the GISTIC2 module on the Broad Institute GenePattern [249] web-based public server. This module could then identify the significance of the CNV per comparison and the implicated genes, as described in section 2.2.6.2 and detailed again here for ease of reference.

Firstly, each aberration is assigned a G-score based on the amplitude of the aberration as well as the frequency of its occurrence across samples. Secondly, the statistical significance of each aberration is defined based on comparing the observed statistic to that expected by chance. Q-values, indicating the false discovery rate and probability that the event is due to a chance fluctuation, are calculated for the aberrant regions and regions with Q-values below a defined threshold are considered significant.

The GISTIC2 module can then report the genomic location and Q-values for each of the regions, along with the genes found in regions of significant amplification or deletion.

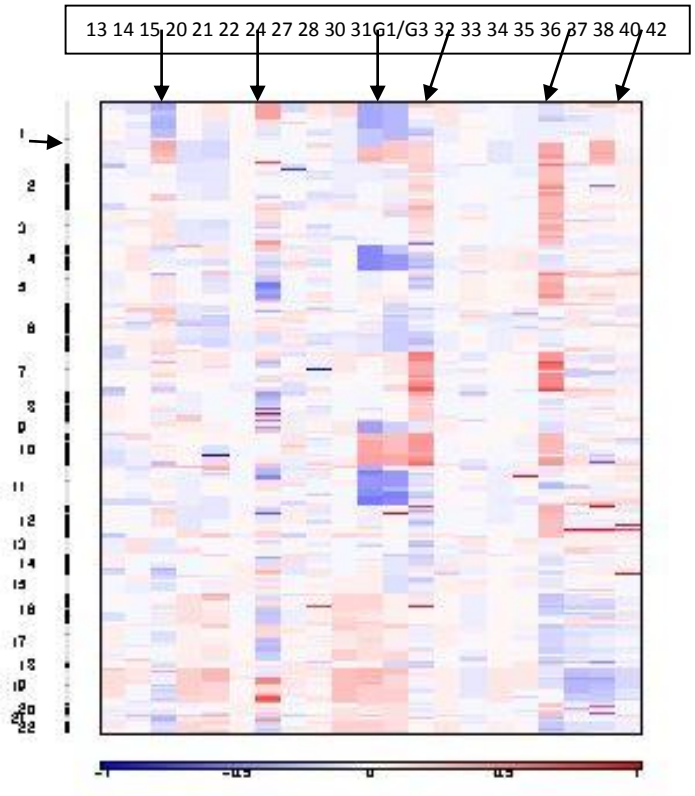
Details of the output files are outlined as follows and are summarised in the following figures and tables:

- Segmented CNV heat map per comparison (Figure 5.15)
- Amplification plot for each comparison (Figures 5.16, 5.17, 5.18)
- CN amplification loci and associated genes (Tables 5.8, 5.9, 5.10)
- Deletion plot for each comparison (Figures 5.19, 5.20, 5.21)
- CN deletion loci and associated genes (Tables 5.11, 5.12)

The third component of the CNV assessment was the analysis of the loci and gene data, similar to the methylation data, through Web Gestalt and the KEGG and Pathway Commons enrichment analyses, with a p-value ≤0.1.

Figure 5.15A: Heatmap showing the copy number amplification (red) and deletion (blue) per chromosome per cancer sample in the Normal Endometrium v Endometrioid Endometrial Cancer (NvC) comparison

The x-axis demonstrates the 21 cancer samples, as compared to the normal samples. Beneath the x-axis is the intensity scale for copy number amplification or deletion.
The y-axis demonstrates each chromosome with p and q segments. The p and q segments within each chromosome are separated by a small dash (as marked on chromosome 1 with a black arrow).



There are certain samples that demonstrate greater CNV (black arrows indicate 15, 24, 31, 32, 37, 40) while other samples show little CNV (13, 14. 20, 21, 22, 27, 28, 30, 33, 34, 35, 36, 38, 42). Note that samples 15, 31, 32 and 37 clustered with an intermediate hypermethylation group within the TCGA data, as per Figure 5.9 (discussed further in section 5.4.3).

Although the specimens that show CNV vary in the intensity of amplification and deletion at different locations within the chromosome, there are some general trends. There is a trend to deletion in Cr1p and amplification in Cr1q, deletion in Cr4, amplification in Cr7 and 10, deletion in Cr11, while there are mixed changes depending on the sample from Cr12-22.
The following two figures demonstrate that the main changes in CNV occur in the atypical to cancer transition, rather than the normal to atypical transition, as the NvC and AvC heatmaps show a similar pattern of change while there are minimal changes in the NvA heatmap. This suggests that these genomic changes occur later in the N-A-C transition.

Figure 5.15B: Heatmap showing the copy number amplification (red) and deletion (blue) per chromosome per cancer sample in the Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC) comparison

The x-axis demonstrates the 21 cancer samples, as compared to the atypical samples.
Beneath the x-axis is the intensity scale for copy number amplification or deletion.
The y-axis demonstrates each chromosome, with p and q segments.
The changes in this heatmap are similar to those in the NvC comparison.

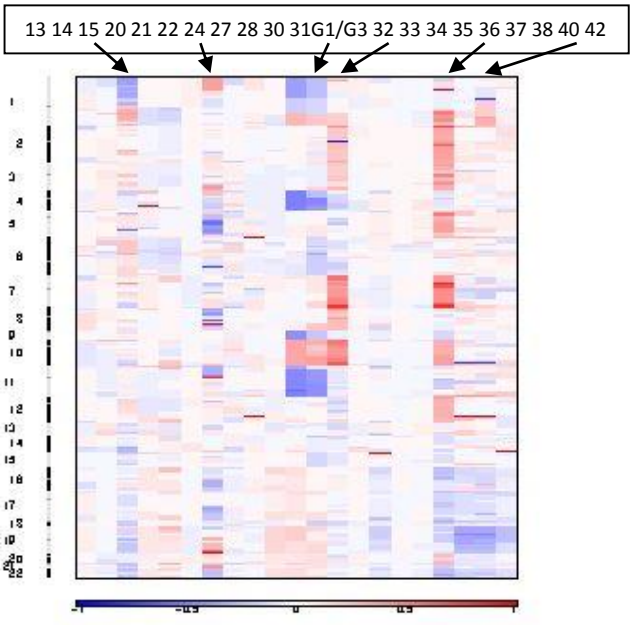


Figure 5.15C: Heatmap showing the copy number amplification (red) and deletion (blue) per chromosome per cancer sample in the Normal Endometrium v Atypical Endometrial Hyperplasia (NvA) comparison

The x-axis demonstrates the 8 atypical samples, as compared to the normal samples.
Beneath the x-axis is the intensity scale for copy number amplification or deletion.
The y-axis demonstrates each chromosome, with p and q segments.
There is very little change in intensity across this heatmap, except in sample 28A, also shown in Figure 5.14B.

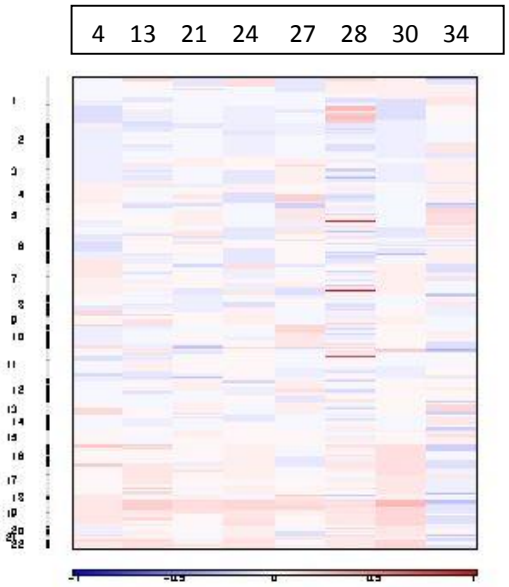
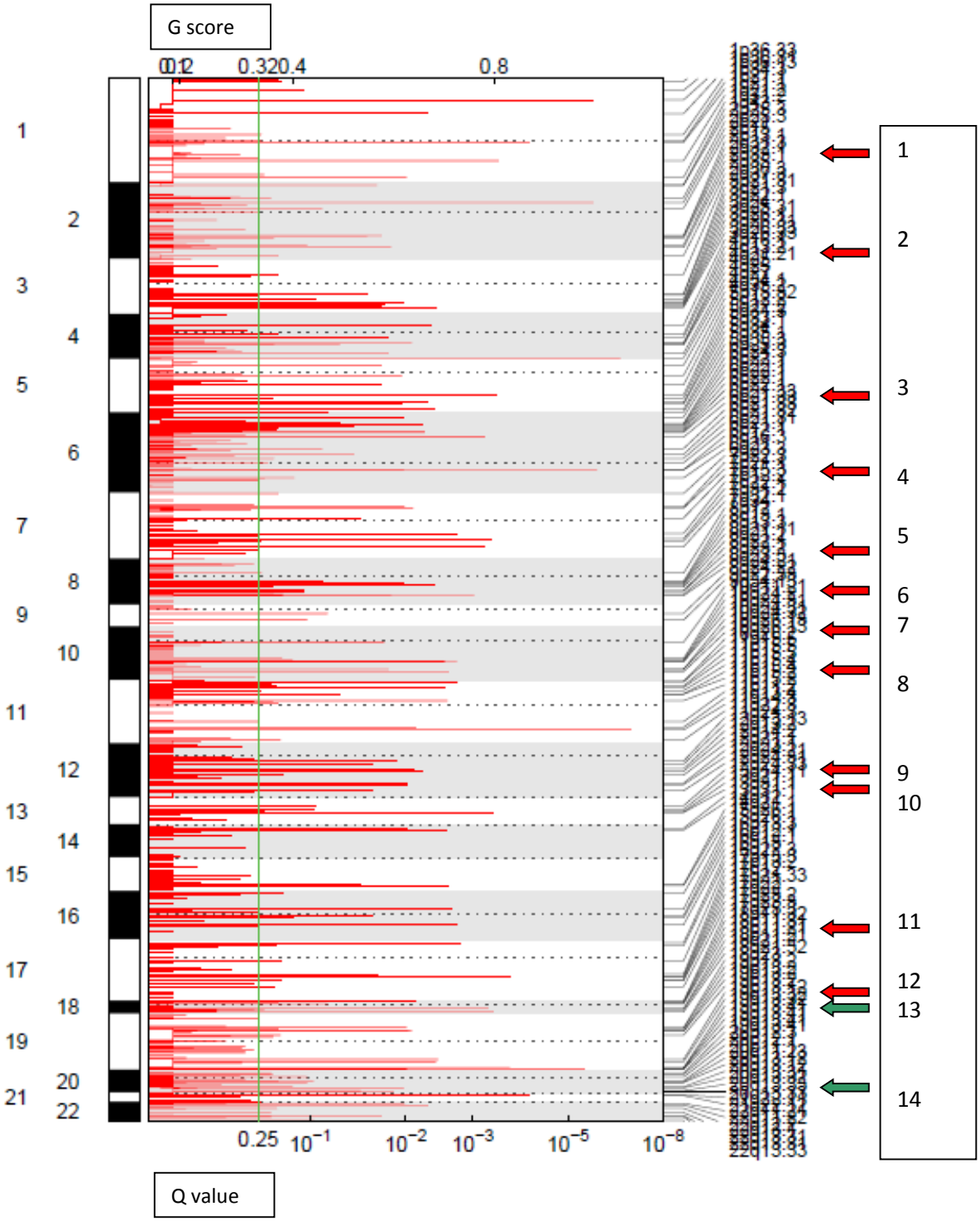


Figure 5.16: Copy number amplification plot for the Normal Endometrium v Endometrioid Endometrial Cancer (NvC) comparison, demonstrated per chromosome and with selected associated genes



For this and the following 2 plots, the chromosome number, with p and q segments are listed down the left hand side. The chromosomal loci are listed down the right hand side for those changes with a Q-value ≥0.25 (vertical green line), as indicated on the bottom x-axis. The Q-value indicates the probability that the event is due to chance.
The top x-axis shows the G-score, reflecting the amplitude of the aberration and frequency across samples (log ratio intensity).
The region with a maximal G-score and minimal q-value is most likely to contain affected genes.
At the furthest right, are indicated the chromosomal loci with genes related to cancer pathways.

Red arrows indicate cancer related genes, listed below and outlined in Table 5.8:
(1) 1q21.1: BCL9 (2) 2q35: PTPRN (3) 5q22.2: APC; 5q35.1: NPM1
(4) 6p21.31: PPARD; 6p21.33: HLA-C; 6p22.1: HLA-G; 6p24.3: TFAP2A; 6p25.3: EXOC2
(5)7q32.1: IRF5 (6) 8p12: NRG1; 8q24.21: MYC (7) 9q34.13: RAPGEF1
(8) 10q24.31: PAX2 (9) 12q13.13: KRT8 (10) 12q24.31: TRIAP1
(11) 16q12.1: NOD2, NKD1 (12)17q22: STXBP4

Green arrows indicate non-cancer related genes:
(13) 17q23.3: ERN1 (14) 20p12.1: PCSK2

Summary of the Amplification Plot Data in the NvC, AvC and NvA comparisons

The NvC and AvC comparisons show greater similarities in CN amplification compared to the NvA comparison, suggesting these genomic changes occur between the AEH to EEC transition. Commonalities in chromosomal and genomic changes between these comparisons occurred with a number of chromosomal loci and genes, including Cr6p21.32/33 (PPARD, HLA-C, TNXB), 7q32.1 (IRF5), 8q24.21 (MYC) and 12q24.31 (TRIAP1, PXN, PLA2GIB).
These 8 genes reflect interactions across a number of cancer related pathways, including the Wnt, VEGF, MAPK, PDGFR-β, EGFR, Arf6, PI3K/mTOR, FAK, c-met and TNFR signalling pathways, as well as being associated with plasma membrane ER signalling and insulin signalling.
MYC amplification was also reported in the TCGA analysis in serous-like EC and in approximately 25% of tumours, classified as high-grade EEC [30].

Table 5.8: Results of the Web Gestalt Enrichment Analyses for the Normal Endometrium v Endometrioid Endometrial Cancer (NvC) copy number amplification analysis including the associated genes and pathways, their location and the Q values

Analysis	Pathway	Gene number	P value	Genes	Location	Q value
Amplified						
KEGG	Wnt SP	4	0.0491	PPARD	6p21.31	0.039
				APC	5q22.2	0.019
Total genes 211				MYC	8q24.21	0.166
				NKD1	16q12.1	0.025
	Endometrial cancer	2	0.0896	APC, MYC		
				HLA-G	6p22.1	0.185
	Type 1DM	3	0.0176	HLA-C	6p21.33	0.00062
				PTPRN	2q35	0.107
Pathway Commons	Wnt SP (W)	4	0.0171	APC (W,p,a)	5q22.2	0.019
	PDGFR SP	11	0.0585	BCL9 (W,a)	1q21.1	0.239
	Arf6 SP	11	0.0585	PAX2 (W)	10q24.31	0.14
	EGFR SP (E)	12	0.0585	TFAP2A (W,T)	6p24.3	0.011
	PI3K/mTOR SP	11	0.0585	RAPGEF1 (a)	9q34.13	0.107
	VEGF & VEGFR SP	11	0.0585	KRT8 (T, a)	12q13.13	0.013
	c-Met/HGFR SP	11	0.0585	MYC (T,a)	8q24.21	0.166
	FAK SP	11	0.0585	TRIAP1 (p,a)	12q24.31	0.010
	TNFR SP (T)	5	0.0694	NOD2 (a)	16q12.1	0.025
	P53 SP (p)	3	0.0759	NPM1 (a)	5q35.1	0.012
	Plasma m/b ER SP	11	0.0585	EXOC2 (a,d)	6p25.3	0.071
	Insulin SP	11	0.0585	IRF5 (a,p)	7q32.1	0.239
	DM pathways (d)	3	0.0949	STXBP4 (a)	17q22	0.0026
				NRG1 (E)	8p12	0.24
				PCSK2 (d)	20p12.1	0.20
				ERN1 (d)	17q23.3	0.20

For this and the following 4 tables:
SP: signalling pathway; DM: diabetes mellitus; m/b: membrane.
Other gene/pathway abbreviations are summarised in the Abbreviations section.
The total gene number generated in the GISTIC analysis is listed in the table, though only the location and Q-value of the cancer-related genes are detailed.
The p value is generated by Web Gestalt, indicating the significance of enrichment for each pathway.
The Q-value is generated by GISTIC2, indicating the probability that the event is due to chance.

There are 11 genes listed in the Pathway Commons section above that are associated with 9 different cancer-related pathways. These genes are listed and marked by (a). Other gene and pathway associations are as indicated.

The 3 pathways at the bottom of the table are of interest for the oestrogenic effects that may play a role in endometrial carcinogenesis.

The Q-value is significant and ≤0.1 for the majority of genes listed above, except BCL9, IRF5, NRG1, PCSK2 and ERN1.

Figure 5.17: Copy number amplification plot of the Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC) comparison, demonstrated per chromosome and with associated genes

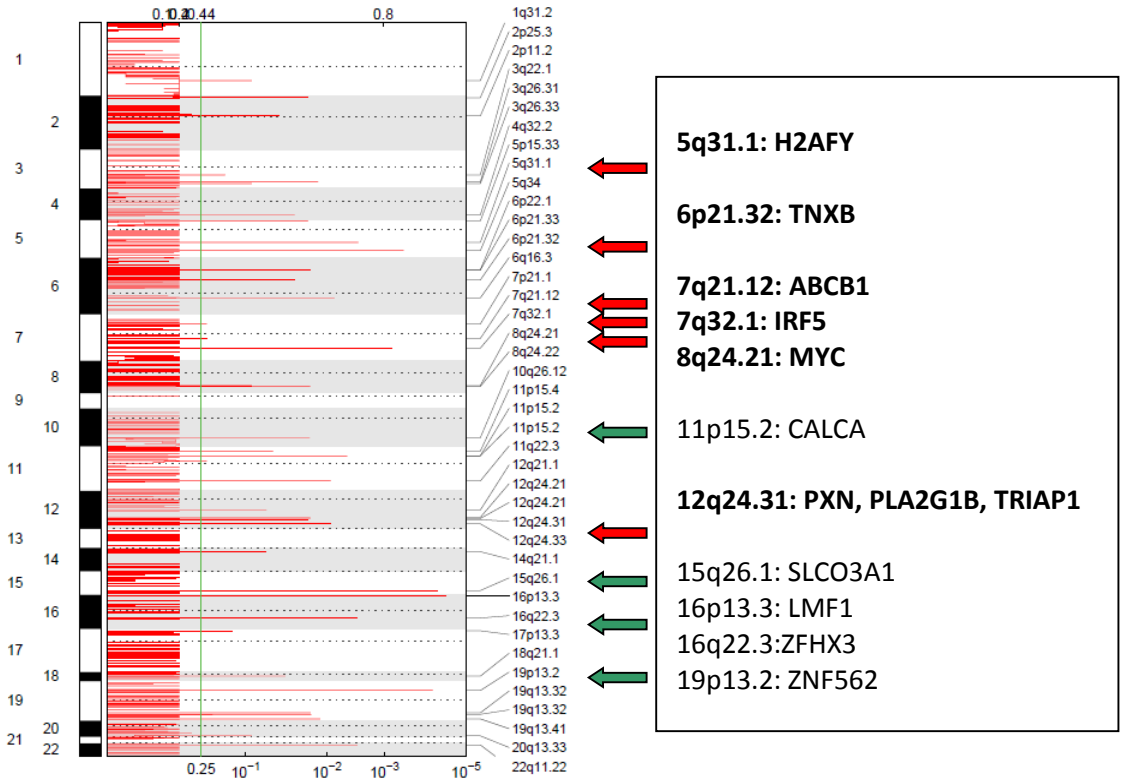
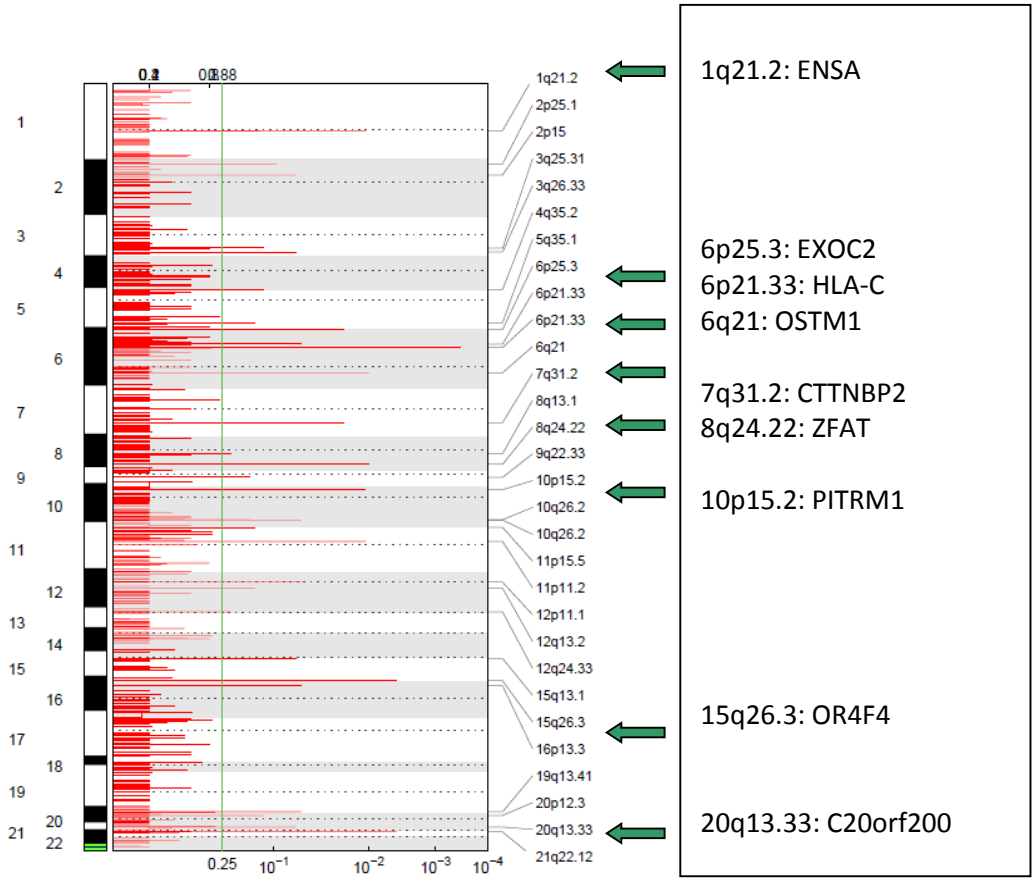


Table 5.9: Results of the Web Gestalt Enrichment Analyses for the Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC) copy number amplification analysis including the associated genes and pathways, their location and the Q values

Analysis	Pathway	Gene number	P value	Genes	Location	Q value
Amplified						
KEGG	VEGF signalling	2	0.0152	PXN	12q24.31	0.021
				PLA2G1B	12q24.31	0.021
Total genes 66	Focal adhesion	2	0.0371	PXN	12q24.31	0.021
				TNXB	6p21.32	0.028
	MAPK signalling	2	0.0538	PLA2G1B	12q24.31	0.021
				MYC	8q24.21	0.22
Pathway Commons	PDGFR-beta SP	6	0.0154	H2AFY (a)	5q31.1	0.0035
	EGFR SP	6	0.0154	PXN (a)	12q24.31	0.021
	ARF6 downstream SP	6	0.0154	MYC (a)	8q24.21	0.22
	PI3K/mTOR SP	6	0.0154	TRIAP1	12q24.31	0.021
	FAK SP	6	0.0154	(a,p)		
	c-MET/HGFR SP	6	0.0154	IRF5 (a,p)	7q32.1	0.00070
	VEGFR SP	6	0.0154	ABCB1 (a)	7q21.12	0.22
	P53 pathway (p)	2	0.0218			
	Plasma m/b ER SP	6	0.0154			

There are 6 genes listed in the Pathway Commons section above that are associated with 9 different cancer-related pathways. These genes are listed and marked by (a). The other cancer-related genes are marked with the indicated pathway abbreviation.
The Q-value is significant and ≤0.1 for the majority of genes listed above, except MYC and ABCB1.

Figure 5.18: Copy number amplification plot of the Normal Endometrium v Atypical Endometrial Hyperplasia (NvA) comparison, demonstrated per chromosome and with associated genes

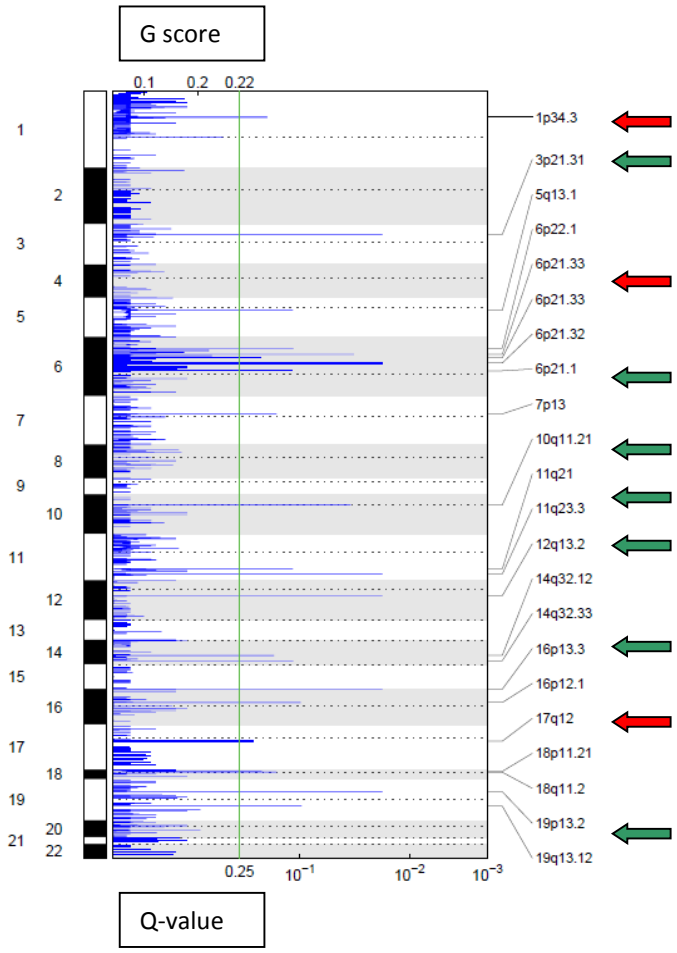


There were no cancer associated genes nor pathways demonstrated through Web Gestalt. The genes identified in the amplification plot as showing a significant change are detailed in table 5.10. Interestingly, this includes genes related to metabolism and diabetes.

Table 5.10: Results of the Web Gestalt Enrichment Analyses for the Normal Endometrium v Atypical Endometrial Hyperplasia (NvA) copy number amplification analysis including the associated genes and pathways, their location and the Q values

Analysis	Pathway	Gene number	P value	Genes	Location	Q value
Amplified						
KEGG Total genes 27	Metabolic pathways	3	0.0329	RRM2 GALNT12 ALG10	2p25.1 9q22.33 12p11.1	0.096 0.16 0.057
Pathway Commons	Signalling by GPCR	3	0.0315	CRH PROKR2 OR4F4	8q13.1 20p12.3 15q26.3	0.22 0.12 0.0042
	Diabetes pathway	2	0.0315	EXOC2 IGFALS	6p25.3 16p13.3	0.020 0.057
	Signal transduction	3	0.0466	CRH, PROKR2, OR4F4	As above	

Figure 5.19: Copy number deletion plot for the Normal Endometrium v Endometrioid Endometrial Cancer (NvC) comparison, demonstrated per chromosome and with associated genes



1p34.3: RPS6KA1, MAP3K6, STMN1, JUN, MKNK1, GADD45A, PIK3R3, PIK3CD, JAK1, PLA2G2C, PLA2G5, WNT4, CSF3R, PTCH2, HDAC1, BMP8A/B
3p21.31: CAMP, CDC25A

6p21.33: RPS6KA2, MAPK14/13, MAP3K4/7, DUSP22, TNF, DAXX, CCND3, PIM1, VEGFA, CDKN1A, DAAM2, LAMA4, HSP90AB1, RXRB, PPARD, HDAC2, BMP5/6, FOXO3
6p21.1: OR2J2, OR2B3P, OR2J3

10q11.21: ZNF33B, BMS1

11q21: PANX1
12q13.2:RDH5

16p13.3: HBA1/2, HBQ1, LUC7L
17q12: TP53, DUSP3/14, MAP3K3/14, MAPK7, PRKCA, MAP2K3/4/6, RAC3, FGF11, TAOK1, PIK3R1/5, STAT5B/A, STAT3, CDC42, ERBB2, FZD2, WNT9B/3, ITGA2B/3,

For this and the following 2 plots, the chromosome number, with p and q segments are listed down the left hand side. The chromosomal loci are listed down the right hand side for those changes with a q-value ≥0.25 (vertical green line), as indicated on the bottom x-axis. The Q-value indicates the probability that the event is due to chance.
The top x-axis shows the G-score, reflecting the amplitude of the aberration and frequency across samples (log ratio intensity).
The region with a maximal G-score and minimal Q-value is most likely to contain affected genes.
At the furthest right, are indicated the chromosomal loci with genes related to cancer pathways.

Cr 1p34.4, 6p21.33 and 17q12 had between 800-1100 genes deleted per chromosomal locus, many of which corresponded to cancer-related pathways as listed in Table 5.11 and demonstrated above in bold font and marked by red arrows.

Cr3p21.31, 6p21.1, 10q11.21, 11q21, 12q13.2, 16p13.3 and 19p13.2 had between 1-3 genes deleted per chromosome but these were not linked to cancer pathways. They are demonstrated in the above figure in non-bold font and marked by green arrows.

Table 5.11: Results of the Web Gestalt Enrichment Analyses for the Normal Endometrium v Endometrioid Endometrial Cancer (NvC) copy number deletion analysis including the associated genes and pathways, their location and the Q values

Analysis	Pathway	Gene n^{o}	P value	Genes	Location	Q value
KEGG	MAPK SP (M)	52	3.89e-12	RPS6KA1/2 (M,m)	1p34.3,6p21.33	0.17,0.19
	VEGF SP (V)	18	4.98e-06	TP53 (M,p,EC)	17q12	0.19
Total genes 3000	Jak-STAT SP (J)	28	2.35e-06	PRKCA (V,F,H,W,C)	17q12	0.19
	Pathways in cancer (C)	44	7.49e-06	MAPK14/13/7 (M)	6p21.33, 17q12	0.19
				MAP3K7/14/3/4/6	6p21.33,17q12,	0.17
	Focal adhesion (F)	30	3.83e-05	(M)	1p34.3	0.19
	Hedgehog SP (H)	11	0.0017	MAP2K3/6/4 (M)	17q12	0.19
	EGFR SP (E)	14	0.0024	DUSP22/14/3 (M)	17q12, 6p21.33	0.17
	Wnt SP (W)	20	0.0026	STMN1 (M)	1p34.3	0.19
	Endometrial cancer (EC)	10	0.0029	RAC3 (M)	17q12	0.17
				JUN (M,E,C)	1p34.3	0.19
	mTOR SP (m)	10	0.0029	FGF11 (M,C)	17q12	0.19
	TGFB SP (T)	13	0.0043	TNF (T, C)	6p21.33	0.19
	P53 SP (p)	11	0.0063	TAOK1 (M)	17q12	0.17
				MKNK1 (M)	1p34.3	0.19
	Type 1 DM	20	3.61e-12	DAXX (M)	6p21.33	0.17
	Insulin signalling	21	0.0003	GADD45A (M)	1p34.3	0.19
	Progesterone-mediated oocyte maturation	15	0.0006	CRK (M,E)	17q12	0.19
				PIK3R1/5/3 (J,V,F,EC,m)	17q12,1p34.3	0.17
				PIK3CD (P,E)	1p34.3	0.17
				STAT5B/A (J,E,C)	17q12	0.19
				STAT3 (J)	17q12	0.19
				JAK1 (J)	1p34.3	0.17
				CCND3 (J,W,p)	6p21.33	0.19
				PIM1 (J)	6p21.33	0.19
				PLA2G2C/E (V)	1p34.3	0.17
				PLA2G5 (M,V)	1p34.3	0.17
				VEGFA (V,m)	6p21.33	0.19
				CDC42 (V)	17q12	0.19
				HER2 (E,F,EC,C)	17q12	0.19
				CDKN1A (E,C,p)	6p21.33	0.19
				FZD2 (W,C)	17q12	0.19
				WNT9B/3 (W,H)	17q12	0.19
				WNT4 (H,W)	1p34.3	0.17
				DAAM2 (W)	6p21.33	0.19
				CSF3R (C)	1p34.3	0.17
				ITGA2B/3 (C, F)	17q12	0.19
				AXIN2 (C,W,EC)	17q12	0.19
				LAMA4 (C, F)	6p21.33	0.19
				PTCH2 (H,C)	1p34.3	0.17
				RARA (C)	17q12	0.19
				HSP90AB1 (C)	6p21.33	0.19
				RXRB (C)	6p21.33	0.19
				PPARD (C)	6p21.33	0.19
				HDAC1/2/5 (C)	1p34.3, 6p21.33,	0.17
					17q12	0.19
				TRAF4 (C)	17q12	0.19
				ROCK1 (F,W,T)	18p11.21	0.18
				BMP5/6/8B/A	6p21.33, 1p34.3	0.19
				(H,T)		0.17
				GRB2 (EC)	17q12	0.19

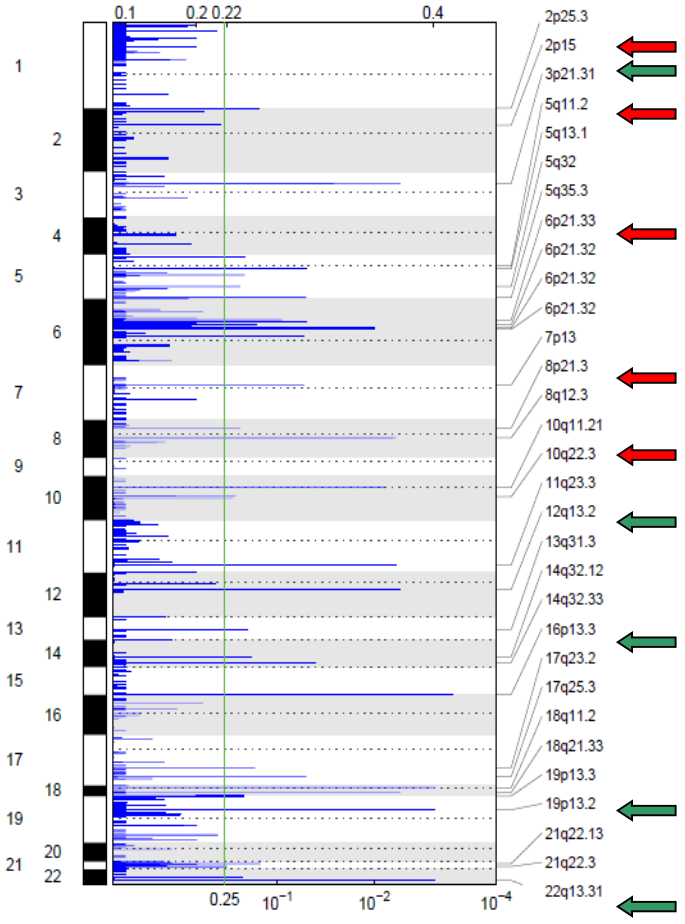
				FOXO3 (EC) CASP9 (EC,V) SMURF2 (T)	6p21.33 1p34.3 17q12	0.19 0.17 0.19
Pathway Common	VEGF&VEGFR SP Arf6 SP PI3K SP PDGFR-β SP EGFR R SP FAK mediated c-MET SP TGFB R SP ALK1 SP TNFR SP c-myc SP HDAC SP MAPK SP E-cadherin SP P53 pathway Wnt SP Notch SP Plasma m/b ER SP DM pathways	215 213 213 213 216 213 213 58 59 53 34 29 38 43 32 32 18 214 38	2.16e-38 2.16e-38 2.16e-38 2.16e-38 2.16e-38 2.16e-38 2.80e-38 3.22e-13 9.00e-13 6.14e-11 3.64e-10 4.85e-10 1.32e-09 1.94e-07 1.54e-06 4.84e-06 6.93e-05 2.59e-38 6.17e-08			

The total gene number generated in the GISTIC2 analysis is listed in the table. Due to the high number of genes identified in both enrichment analyses, the 70 genes with the smallest p values in the Web Gestalt analysis are listed above with their chromosomal location and Q-value. The associated gene pathways are marked next to the gene as indicated in brackets.

The 3 pathways at the bottom of the KEGG analysis row and 2 pathways at the bottom of the Pathway Commons row are of interest for the metabolic and oestrogenic effects that may play a role in endometrial carcinogenesis.

The Q-value is ≤0.1 for all the genes listed above.

Figure 5.20: Copy number deletion plot for the Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC) comparison, demonstrated per chromosome and with associated genes



2p15: MAP4K4/3, GLI2, FZD5/7, WNT10A/6, PPP3R1, MSH2/6, SOS1, STAT1/4, COL4A4, TGFA, PAX8, ITGAV, DUSP2, RASGRP3, CACNB4, ATF2, ERBB4, IHH, CALM2

3p21.31: CAMP, CDC25A

5p15.33: MAP3K1/2, ITGA2, MSH3, FGF10, RASA1, RASGRF2

5q13.1: PIK3R1

6p21.33: MAPK13/14, DAAM2, PPARD, CCND3, CDKN1A, RXRB, HSPA1A

8p21.3: FGF17/20, FGFR1, FZD3, DUSP4, PPR3CC

10q22.2: MAPK8/3K8, CTNNA3, FZD8, WNT8B, SFRP5, ITGA8, RET, CHUK, NFKB2, FGF8, FGFR2, FAS, SUFU, ITGB1, DUSP5, PPP3CB, CACNB2, NRG3, PIP4K2A, PI4K2A, CALML3/5

11q23.3 HYOU1, 12q13.2 RDH5

16p13.3: HBA1/2, HBQ1, LUC7L

19p13.2: ZNF823

22q13.31: ATXN10

Cr 2p15, 5p15.33, 6p21.33, 8p21.33 and 10q22.2 had between 180-1000 genes deleted per chromosomal locus, many of which corresponded to cancer-related pathways as listed in table 5.12 and demonstrated above in bold and marked by red arrows.

Cr3p21.31, 11q23.3, 12q13.2, 16p13.3, 19p13.2 and 22q13.31 had between 1-5 genes deleted per chromosome but these were not linked to cancer pathways. They are demonstrated in the above figure as non-bolded and with green arrows.

Table 5.12: Results of the Web Gestalt Enrichment Analyses for the Atypical Endometrial Hyperplasia v Endometrioid Endometrial Cancer (AvC) copy number deletion analysis including the associated genes and pathways, their location and the Q values

Analysis	Pathway	Gene number	P value	genes	location	Q value
Deleted						
KEGG	Pathways in cancer (C)	53	1.30e-10	MAPK8. 3K8	10q22.2	0.21
	MAPK SP (M)	42	2.10e-08	(M,F,N,E)		
Total	Wnt SP (W)	28	1.80e-07	MAPK13,14	6p21.33	0.056
genes	TGFB signalling	16	6.06e-05	MAP3K1,2 (M)	5p15.33	0.18
2712	Focal adhesion (F)	27	7.95e-05	MAP4K4, 4K3	2p15	0.27
	Phosphatidylinositol	15	8.99e-05	GLI2 (H)	2p15	0.27
	SP (P)			CTNNA3 (J)	10q22.2	0.21
	Jak-STAT SP (J)	20	0.0011	FZD8 (W)	10q22.2	0.21
	Hedgehog SP (H)	10	0.0023	FZD3 (W)	8p21.3	0.19
	P53 SP	11	0.0028	FZD5,7 (W)	2p15	0.27
	NOD-like R SP (N)	10	0.0028	WNT10A, 6	2p15	0.27
	EGFR SP (E)	12	0.0060	(W,H)		
				DAAM2 (W)	6p21.33	0.056
	Metabolic pathways	191	2.31e-41	WNT8B (W,H)	10q22.2	0.21
	Type 1 DM	21	1.42e-13	SFRP5 (W)	10q22.2	0.21
	Oocyte meiosis	20	2.20e-05	APC (W)	5p15.33	0.18
	Insulin signalling	22	3.73e-05	PPARD (W)	6p21.33	0.056
				CCND3 (W)	6p21.33	0.056
				PPP3R1, 3CC	2p15, 8p21.3	0.27
				(M,W)		
				ITGA2,6, 8 (F)	5p15.33,2p15,	0.18
					10q22.2	0.21
				RET (C)	10q22.2	0.21
				CHUK (M)	10q22.2	0.21
				CDKN1A (C)	6p21.33	0.056
				MSH2 3,6 (C)	2p15, 5p15.33	0.27
				NFKB2 (M)	10q22.2	0.21
				FGF10 (M)	5p15.33	0.18
				FGF8 (M)	10q22.2	0.21
				FGF17,20 (M)	8p21.3	0.19
				FGFR1 (M)	8p21.3	0.19
				FGFR2 (M)	10q22.2	0.21
				SOS1 (M,F,J,E)	2p15	0.27
				STAT1,4 (J)	2p15	0.27
				COL4A4 (F)	2p15	0.27
				TGFA (E)	2p15	0.27
				FAS (M)	10q22.2	0.21
				RXRB (C)	6p21.33	0.056
				SUFU (H)	10q22.2	0.21
				PAX8 (C)	2p15	0.27
				ITGB1 (F)	10q22.2	0.21
				ITGAV (F)	2p15	0.27
				DUSP2, P5.P4	2p15, 10q22.2,	0.27
				(M)	8p21.3	0.19
				RASA1 (M)	5p15.33	0.18
				RASGRF2 (M)	5p15.33	0.18
				RASGRP3 (M)	2p15	0.27
				PPP3CB (M)	10q22.2	
				CACNB2,B4	10q22.2, 2p15	0.27
				(M)		
				HSPA1A (M)	6p21.33	0.056
				ATF2 (M)	2p15	0.27

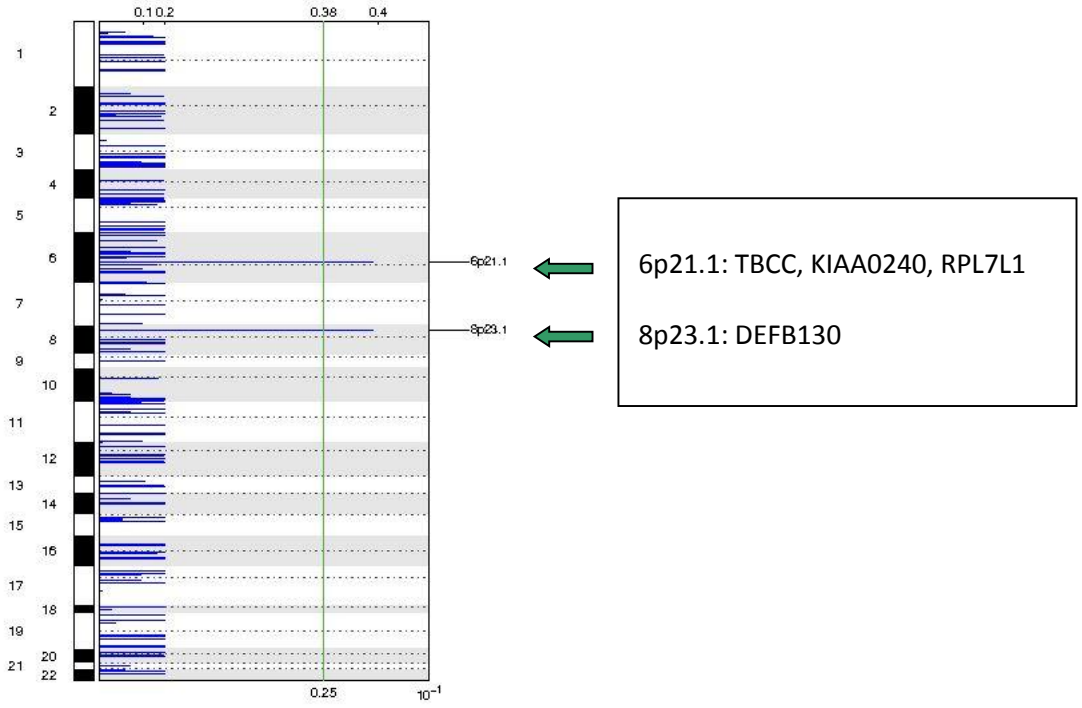
				NRG3 (E) ERBB4 (E) PTEN (P,F) PIK3R1 (P,F) CALM2, L5, L3 (P) PIP4K2A (P) PI4K2A (P) IHH (H)	10q22.2 2p15 10q22.2 5q13.1 2p15, 10q22.2, 2p15 10q22.2 10q22.2 2p15	0.21 0.27 0.21 0.056 0.27 0.27 0.21 0.21 0.27
Pathway Commons	EGFR receptor SP VEGF & VEGFR SP FAK mediated SP PI3K/mTOR SP Arf6 SP PDGFR-β SP c-MET/HGFR SP c-myc SP P53 SP TNFR SP ALK1 SP TGFB R SP Wnt SP FGFR SP Plasma m/b ER SP	202 200 196 196 196 196 197 35 38 46 46 41 30 18 197	1.61e-37 3.52e-37 1.24e-36 1.24e-36 1.24e-36 1.24e-36 1.24e-36 5.08e-12 8.07e-11 7.13e-09 6.57e-08 2.01e-06 7.01e-06 3.50e-05 $1.52e^{-36}$			

The Q-value is ≤0.1 for 21 of the genes listed above, corresponding to those located on Cr 5p15.33, 6p21.33 and 8p21.3. The Q-value is between 0.2-0.3 for the remaining genes.

The total gene number generated in the GISTIC2 analysis is listed in the table. Due to the high number of genes identified in both enrichment analyses, the top 70 genes are listed above with their chromosomal location and Q-value. The associated gene pathways are marked next to the gene as indicated in brackets.

The 4 pathways at the bottom of the KEGG analysis row and 1 pathway at the bottom of the Pathway Commons row are of interest for the metabolic and oestrogenic effects that may play a role in endometrial carcinogenesis.

Figure 5.21: Copy number deletion plot for the Normal Endometrium v Atypical Endometrial Hyperplasia (NvA) comparison, demonstrated per chromosome and with associated genes



On KEGG and Pathway Commons enrichment analyses on WebGestalt, along with reviewing the gene information on GeneCards, no significant cancer association was found.

Summary of the Deletion Plot Data in the NvC, AvC and NvA comparisons

As with the amplification plots, the NvC and AvC comparisons show greater similarities in CNV compared to the NvA comparison, suggesting these genomic changes occur in the AEH to EEC transition, rather than between the normal endometrium to AEH transition.

Only Cr6p21.33 deletion is common between the NvC and AvC comparison for cancer-related pathways. Deletion in 3p21.31, 16p13.3 and 19p13.2 were also common between the NvC and AvC comparison, but no cancer-related genes were identified at these loci. The other deletions of Cr1p34.4 and 17q12 occurred between the NvC comparison, while Cr2p15, 5p15.33, 8p21.33 and 10q22.2 deletion occurred between AvC, in terms of cancer-related pathways.

Although there was only one chromosomal locus in common with CN deletion, there were a number of genes and cancer-related pathways implicated at this site. The deleted genes include MAPK13, MAPK14, DAAM2, PPARD, CCND3, CDKN1A, RXRB, HSPA1A, RPS6KA2, DUSP22, TNF, DAXX, VEGFA, PIM1, LAMA4, HSP90AB1, HDAC2, BMP5, BMP6 and FOXO3. These genes are associated not only with EC but also with signalling pathways implicated in endometrial tumorigenesis including the Wnt, VEGF, EGFR, MAPK, PI3K/mTOR, Jak-STAT, Hedgehog, focal adhesion, TGFB and p53 signalling pathways.

5.4. Discussion

The methylation and CNV data generated from the Illumina 450K array in this study demonstrate the feasibility of this platform for FFPE endometrial tissue. Not only were the FFPE and FF data consistent within the study specimens, but the FFPE endometrial data here were also consistent with publicly available TCGA FF endometrial data of normal and cancer tissue. Importantly, a spectrum of methylation and CNV changes was demonstrated between normal endometrium, AEH and EEC, which may provide a foundation for the development of a methylation signature for prognostic and predictive purposes. In addition, the reproducibility of these findings in FFPE samples allows a far greater bank of tissue to be interrogated for epigenetic and genetic studies.

5.4.1. Evidence of Differential Methylation and Clinical Correlation

The methylation data analysis illustrated the feasibility of FFPE endometrial tissue processing on the Illumina 450K array. In addition, it demonstrated that differential methylation occurs in endometrial carcinogenesis from normal endometrium to AEH to EEC. A significant part of this progression from normal endometrium to EEC involved gene hypermethylation but there were also a smaller number of genes that were hypomethylated. The methylation changes occurred predominantly in the gene promoter region and gene body and involved CGIs and CpG shores. Between the normal endometrium to EC transition, there was evidence of gene hyper- and hypomethylation in the PI3K/mTOR/Arf6, VEGF, FGF/FGFR, EGFR, PDGFR, TGFB, p53 and the Wnt/E-cadherin signalling pathways, for which molecular aberrations are well documented. There were also methylation changes associated with FAK and cell adhesion, ALK, c-met/HGFR, NOTCH and Trk signalling pathways, for which data in EC is limited and these may represent novel pathways involved in endometrial carcinogenesis.

In addition, the methylation changes appeared to differ in timing between patients. There were some cases where the methylation changes occurred early in tumorigenesis between normal endometrium and AEH, while other cases demonstrated later methylation changes, between AEH and EEC. Whether the timing of these changes impacts on clinical outcomes and response to therapy could not be established from this analysis but may be part of a methylation signature that could give prognostic and predictive information. Further clinical correlation was limited here with the small sample size and available follow-up. Thus, the possible trend between grade, tumour size and older age with greater hypermethylation needs further clarification.

Although molecular aberrations could be identified between the normal endometrium to EEC transition, it was not possible to determine which pathway and gene changes were most significant between the normal endometrium to AEH transition compared with the AEH to EEC transition. Again, this was due to limitations in sample size, particularly in the number of AEH specimens available.

However, for a number of genes in the Wnt, FGF/FGFR, MAPK and VEGF pathways, progressive methylation changes between the histological types was evident. Interestingly, there were a number of genes in the Wnt and FGFR signalling pathways as well as the TSG RASSF1A that demonstrated greater hypermethylation between the normal endometrium to AEH transition. It is known that hypermethylation in the promoter region of TSGs can inactivate the gene and promote carcinogenesis [262] and these changes may reflect this early progression.

A larger study is required to validate these findings and may include where possible, longitudinal tissue analysis with repeat biopsies at diagnosis, metastasis and progression and with longer clinical follow-up. This would allow correlation of molecular findings in early and late stage disease with clinical parameters to help define a methylation signature with prognostic and predictive potential. Identifying a methylation signature to define the epigenetic transition between normal endometrium, AEH and EC may then improve treatment stratification and strategies.

5.4.2. Evidence for CNV analysis in EEC and Correlation with Methylation Changes

The analysis of CNV on FFPE EC tissue was also feasible and further demonstrated a spectrum of changes between normal endometrium, AEH and EEC. CN changes varied across the samples with 28% (6 from 21 samples) demonstrating more prominent CN alterations. For both CN amplification and deletion, the overall NvC comparison was similar to the AvC comparison, indicating that the majority of changes occurred later in tumorigenesis. Thus, for the purpose of finding a molecular signature early in tumorigenesis, methylation analysis rather than CNV may be of the most utility.

The implicated cancer-related pathways for both CN amplification and deletion included the Wnt, VEGF, PI3K/mTOR/Arf6, MAPK, EGFR, PDGFR-β, TGFB and p53 signalling pathways. As with the methylation data, there were also a number of novel pathways that were implicated including focal adhesion, c-met/HGFR, Jak-STAT and Hedgehog signalling, though further validation of these results is required.

Individual genes were mapped to the chromosomal loci that demonstrated significant CN alterations and there were commonalities between those that were affected by methylation changes as well as those that underwent CN alterations. In the Wnt pathway for example, APC promoter hypermethylation occurred early in tumorigenesis between the normal endometrium and AEH transition, while CN amplification of APC occurred as a later event. Other FZD genes and Wnt genes either underwent promoter hypermethylation or demonstrated CN deletion, reflecting a possible interplay between the genomic and epigenomic aberrations.

There were also changes within the FGF/FGFR signalling pathway that included promoter hypermethylation of FGF2, 12 and 10, as well as CN deletion in FGF10 and 11. VEGFA also showed promoter hypermethylation and CN deletion, both of which tended to occur in the AEH to EEC transition.

There are multiple methods of CNV analysis which utilise other platforms, but using the Illumina 450K array and ChAMP pipeline was particularly amenable here, as data for both the methylation and CNV analyses could be generated from the one platform. This is of importance when tissue availability is limited, particularly for AEH in this analysis, but could be extended to other tissue types where the maximum data needs to be generated from limited DNA.

5.4.3. Correlation of Epigenetic and CNV data with TCGA data

The data presented here was consistent with that in the wider literature and TCGA public database. As detailed in chapter 1, TCGA [30] identified four categories of EC based on integrated genomic and epigenomic analysis of FF tissue: POLE ultramutated, MSI/hypermutated, CN low and CN high. They found most EEC had few CN changes but frequent mutations in PTEN, CTNNB1, PIK3CA, ARID1A and KRAS, while a subset of EEC showed greater similarity to serous tumours and were CN high with few methylation changes.

In terms of the methylation changes, as discussed in section 5.2.3, the normal endometrium, AEH and EC clustered similarly to the TCGA endometrial tissue types. As TCGA did not include AEH, it was interesting to see how some of the AEH samples clustered with normal tissue and some clustered with cancer tissue, indicating that there may be some methylation changes that occur early in tumorigenesis and some that occur later. Identifying a methylation signature for those AEH samples that cluster with the cancer samples may identify those cases that require more aggressive treatment.

In the TCGA analysis of EC and the CN low and high groups, the CN low tumours were mainly EEC and demonstrated 1q amplification. The CN high tumours were serous-like, heavily methylated and with amplification in Cr8q24.12 (MYC), 17q12 (HER2), 19q12 (CCNE1), 4p16.3

(FGFR3) and 8q11.23 (SOX17). Although the sample analysis here in that 28% that demonstrated CNV only identified the amplification in Cr8q24.12 (MYC) in common with TCGA data, other loci were demonstrated where FGFR1 and 2 were deleted (Cr8p21.3 and 10q22.2 respectively) as well as where FGF8, 10, 17 and 20 were deleted (Cr10q22.2, 5p15.33 and 8p21.3). PI3K pathway changes were also identified with CN deletions in Cr10q22.2 and 17q12, as well as CN deletion in STMN in Cr1p34.4. Correlating these results at the protein and receptor level would also be key in showing biologic and clinical relevance, and this again would require further validation in larger studies. As the PI3K and FGFR pathways are readily targetable with available agents, validation of these findings would be particularly important and may present an additional predictive biomarker of these agents.

The number of samples with significant CNV were small in this study but consistent with TCGA findings. TCGA found 25% of EEC were CN high and it was these samples that behaved more like serous cancers. From this analysis, 28% appeared to have greater CNV. Interestingly, 4 of these samples (15, 31, 32 ,37 as in Figure 5.15A) clustered with a group of TCGA cancer samples that had less intense methylation changes, as evidenced in Figure 5.9. This may correspond to the subset of EEC tumours that are CN high with few methylation changes. Such tumours may behave more like serous tumours with a worse prognosis, though there was inadequate follow-up data to comment on that here.

Figure 5.22 outlines the key findings from the epigenetic and CNV analysis across normal endometrium, AEH and EEC tissue. Importantly, TCGA data and this analysis show commonalities in molecular alterations in a number of pathways, particularly with Wnt, PI3K, FGFR and VEGF signalling. While agents targeting the Wnt pathway remain in development, there are a number of available agents or those already in clinical trials to address the other targets. Applying these targeted therapies to EC with concurrent biomarker development is urgently required to improve on current patient outcomes. The incorporation of novel biomarker strategies is key and the ability to do this from FFPE tissue provides a far greater repository for expansion of research in this area.

Further analysis of these pathways and the described research techniques in a greater sample size would better validate the frequency of these aberrations and determine whether targeting these pathways may show activity in EEC.

Figure 5.22. Key Findings in the Epigenetic and CNV analysis in EEC

- ❖ Epigenetic and CNV analysis of FFPE and FF EEC tissue is feasible on the Illumina 450K array
- ❖ Differential methylation patterns, predominantly hypermethylation, are evident between normal endometrium, AEH and EEC
- ❖ Study data from the epigenetic and CNV analysis were consistent with that available from TCGA
- ❖ Differential methylation of specific genes was demonstrated between normal endometrium, AEH and EEC
- ❖ Further validation in a larger sample size for both EEC and NEEC at early and advanced stage is needed to determine the clinical significance of these findings with the aim of developing an epigenetic and genetic signature to stratify patients for prognostic and predictive purposes.

CHAPTER 6: Conclusions and Future Directions

There is an ongoing need in the treatment of EC to establish novel treatment strategies and improve patient selection for existing drugs and those in development. The parameters used to guide therapy of EC have not changed for many years and current treatment strategies do not reflect the known molecular aberrations, being limited to a "one-size-fits-all" surgery/chemotherapy/radiotherapy approach.

In this context, development of biomarkers that are prognostic or predictive for treatment outcome may be helpful. Such markers may potentially accelerate drug development and subsequent approval through selection of patients that are most likely to benefit from a given treatment.

The aims of this work were two-fold. The first was to assess CTC enumeration and molecular profiling with stathmin in advanced EC and how it might correlate with clinical course and clinicopathologic factors. The second was to conduct epigenetic and CNV analysis from FFPE and FF normal endometrium, AEH and EEC, to establish if this was feasible, whether there were novel differential changes between normal, atypical and cancerous endometrial tissue and the histopathological grades, to characterise the molecular pathways involved and finally, to correlate with data already in the public domain.

The results presented in Chapters 3, 4 and 5 demonstrate the feasibility of these two techniques for the first time in EC and the potential to identify novel predictive and prognostic markers. The overall findings from these studies are discussed below and highlight areas of future investigation that might enable their use in the clinical management of patients.

Sections 6.1-6.4 discuss novel technologies for CTC enumeration and molecular profiling to develop in EC, and is followed by discussion of the use of a methylation signature as a predictive or prognostic biomarker in EC in sections 6.5-6.8.

6.1. CTC Enumeration and Molecular Profiling

To our knowledge, this is the first documentation of CTC enumeration in EC, with 60% of patients with advanced EC showing detectable levels of CTCs. Furthermore, in a subset of patients receiving first-line chemotherapy for recurrent, metastatic disease, longitudinal assessment of CTCs could be a predictive biomarker of response and correlate with clinical outcome.

There were a number of histopathological factors, including LVSI, myometrial and cervical invasion that currently determine post-surgical primary treatment, which did not show an association with CTC enumeration. This may be a reflection of the limited utility of these histopathological factors or that CTC enumeration needs further clarification in studies with a larger number of patients at different stages of treatment.

Stathmin overexpression was also demonstrated for the first time in EC CTCs using the 4th channel on the CellSearch platform. This was reproducible and time-efficient as it could be performed at the same time as CTC enumeration. As outlined in section 1.4.1.3, stathmin overexpression is prognostic in EC [122, 164] and may also have predictive value for patients treated with taxane-based chemotherapy [168]. Although the small patient number here limits interpretation, the potential to correlate this with outcomes with both chemotherapy and targeted agents is worthy of further exploration. If validated in larger studies, longitudinal evaluation of CTCs, stathmin expression and correlation with treatment outcome may provide the first predictive biomarker for treatment response in EC in those patients who are CTC+. Moreover, as stathmin more accurately reflects PI3K pathway activation compared to other available markers [122, 162], CTCs and their stathmin expression should be incorporated into clinical trials that assess PI3K pathway targeting agents as well as with taxane-based chemotherapy combinations.

Supporting the use of EpCAM-guided CTC enumeration techniques in EC, EpCAM overexpression was reported in 86% of archival FFPE tissue here. Detection of stathmin overexpression on FFPE tissue was also reproducible and reported at 80%. Correlation of FFPE EpCAM and stathmin IHC with CTC findings, though not demonstrated here, is worth further investigation in conjunction with CTC validation studies to delineate the relationship between IHC and CTC findings [163].

As the UCL Cancer Institute has both validated stathmin expression on CTCs on the CellSearch platform and has novel technologies in development including the Gilupi nanodetector, ISET and DepARRAY platforms, there is a novel opportunity to conduct further research in EC here. Further details on these techniques and molecular characterisation of CTCs is outlined in the following three sections.

6.2. Methods to Optimise CTC collection in Endometrial Cancer

There are a number of novel technologies in CTC enumeration and molecular characterisation that might optimise their utility in the clinical setting and be used for future research in EC. The enumeration technologies are broadly based on either indirect immune-mediated methods, such as EpCAM detection, or direct methods based on physical properties, such as size, density or conductivity.

Another EpCAM based approach utilises the Gilupi nanodetector, a seldinger guidewire coated with anti-EpCAM [270] that is placed in the cubital vein of patients for 30 minutes, allowing a blood volume of 1.5-3 litres to come in contact, and increasing cell exposure to EpCAM based detection. Once the collection is performed, the cells are stained for cytokeratins, CD45 antibody and DAPI to differentiate CTCs from other cellular material. This method has the

potential both to increase detection of viable cells and allow isolation of cells for profiling, though further optimisation at the UCL Cancer Institute is still required.

Another CTC detection method that has been described and will be used at the UCL Cancer Institute is the ISET (isolation by size of epithelial tumour cells) system based on cell size rather than EpCAM enrichment. Larger CTCs are separated out through filtering membranes with calibrated pores 8 micron in diameter [98, 271] and can then be characterized morphologically and by protein expression. This type of method may avoid the concerns regarding sensitivity when EpCAM or other panels of markers are used for detection, particularly in regards to EMT and tumour heterogeneity. Early comparative studies of the ISET and CellSearch platforms demonstrate mixed findings in breast, prostate and lung cancer with higher counts on ISET in prostate and lung cancer and lower counts in breast cancer [272]. ISET may also have its limitations in enumeration as CTCs can be lost through the ISET detection process. It is also more time consuming, blood samples have to be processed within 4 hours and it has not yet been standardised for routine clinical application.

Two other detection methods that may warrant comparison with the CellSearch platform are the DepARRAY system and ImageStream (Amnis). The DepARRAY system is a semi-automated system that uses application of dielectrophoretic field flow fractionation (depFFF) to isolate CTCs from blood specimens [273]. The depFFF method relies on differences in cell phenotype and membrane capacitance to isolate cells so that labelling is unnecessary. On the DepARRAY system, isolation and visualisation of CTCs is based on the capture of cells by 'electric cages' created on a microelectronic chip [274] with the ability to recover the cells for further molecular analysis. It is however a time intensive system to screen and isolate relevant cells. Another depFFF based platform, Apostream (Apocell, Houston, USA) reports CTC detection rates up to 95% in metastatic prostate, breast and colorectal cancer [275].

Imagestream (Amnis) is an imaging cytometry device that combines flow cytometry and fluorescent microscopy in a single platform [276]. Studies to date have used similar immunomagnetic positive selection as CellSearch, including anti-EpCAM, anti-CD45 and anti-cytokeratin. A potential benefit is the ability to perform impartial analysis of a large number of cells in a short period of time and it can be fully automated if required. In a comparison to CellSearch, detected CTC counts were similar though there are limitations at lower CTC counts with ImageStream [276].

There are certainly limitations of the CellSearch platform and EpCAM-based technologies, with their reduced sensitivity for detecting CTCs due to variation in expression of EpCAM and EMT. However, this study demonstrated the presence and early clinical correlation patterns with the current technology platform. Other technologies and markers for detection may be more sensitive but identification of the most appropriate markers and development of new

techniques in individual cancers requires further validation [277]. At this time point, an EpCAM detection system may still be preferred in EC, particularly with its high rate of EpCAM overexpression. Within the UCL Cancer Institute, further evaluation of CellSearch, the Gelupi nanodetector, ImageStream and DepARRAY may be developed in EC, to determine the optimum method for this cancer type and whether a certain platform is better for enumeration or isolation for molecular profiling.

6.3. Molecular Profiling Techniques for CTCs in Endometrial Cancer

Techniques for molecular characterisation of CTCs have developed rapidly and while earlier reports were based on cell extraction from the CellSearch magnetic cartridge [278-280], more recent characterisation can now incorporate isolation from other technologies and include next generation sequencing (NGS) platforms [281].

The most time efficient method for CTC molecular profiling when using the CellSearch platform is the use of an antibody within the 4th channel. Although this only allows analysis for a single antibody per blood sample and each antibody must be appropriately validated for use on CTCs, results are automatically generated at the time of enumeration and can potentially be correlated with more in-depth molecular analysis by other means.

In addition to the 4th channel analysis, techniques have been described to extract CTCs from the CellSearch magnetic cartridge and perform further molecular characterisation. For example, it is possible to identify EGFR expression, androgen receptor (AR) gene amplification [278], ERG, AR and PTEN gene rearrangement status in prostate cancer [279] as well as EGFR gene amplification, KRAS, BRAF and PIK3CA mutation status in colorectal cancer [280]. In lung cancer, CTCs isolated on the CellSearch system have been subjected to ultra-deep NGS with the identification of EGFR mutations corresponding to those present in matching tumour tissue [281]. Similarly, this could be applied to CTCs isolated from EC patients, particularly for changes in the PI3K, FGFR and Wnt pathways that have the most frequent aberrations.

The choice of methodology for isolating CTCs for molecular characterisation will be guided by operator expertise as well as time and cost constraints, and should be taken into account when planning further studies for CTC enumeration, such that the maximum amount of information can be gained from each study. Ideally, a broad range of techniques would be used where possible including IHC, ISH for overexpression and gene translocation/fusion analysis, NGS for mutation analysis, as well as RNA and protein expression analysis to correlate upstream findings with downstream effects.

There are already planned trials with targeted therapies in EC that will utilise CTCs and molecular profiling both on CTCs and tumour tissue, including a poly ADP-ribose polymerase (PARP) inhibitor study and a PI3K inhibitor study [282].

Correlating longitudinal CTC findings with archival tumour tissue at diagnosis and metastasis/progression, where available, would also be of great value for further studies in EC. As tumour tissue is the gold standard for molecular characterisation of tumours, CTC molecular profiling needs to be closely correlated to justify its use in the clinical setting for assessment of molecular aberrations. Particularly with arguments for and against longitudinal tissue biopsies for molecular analysis of patients' tumours during treatment, CTC assessment would be an attractive option as a so-called 'liquid biopsy' that is readily accessible and non-invasive.

6.4. Future Directions beyond CTCs

Only 20-50% of patients, depending on the cancer type, will demonstrate detectable CTCs at baseline, though novel technologies report detection rates higher than this [275]. Although the detection rate in EC here was 60%, this was in the setting of advanced disease and extending analysis to earlier stage disease will potentially result in a lower detection rate. As well as the other enumeration techniques as described above, the choice of collection site may also warrant further testing to account for lower detection peripherally. Other tumour types have compared peripheral collection with collection from a central, draining vein intra-operatively [283], and this may be another approach to consider. Otherwise, focusing only on CTCs in monitoring and molecular profiling may not benefit a significant proportion of the patient population currently requiring treatment.

The feasibility of analysing circulating free DNA (cfDNA) in EC is another area of interest, and along with CTC detection and profiling, is an attractive option for biomarker development due to readily available sampling that can be repeated longitudinally. CfDNA fragments are thought to originate from apoptotic or necrotic tumour cells that release their DNA into the blood circulation. It is now possible to apply NGS technologies to cfDNA and studies have demonstrated genetic mutations relevant to drug resistance in metastatic cancer [284]. Detection of cfDNA has also been associated with tumour burden, prognosis [285, 286], prediction of response [287, 288], detection of recurrence and with the ability to identify genetic aberrations including HER2 amplification in breast cancer and KRAS, PIK3CA and BRAF mutations in colorectal cancer [289-291]. Like CTC enumeration, cfDNA analysis relies on there being adequate concentrations in the peripheral blood and requires standardization and validation of laboratory and preanalytic conditions. Normal DNA from dying cells after blood collection may contaminate the specimen and there is the question as to whether the DNA from dying tumour cells gives the best information on therapy-resistant cancer cells that remain viable.

Combined CTC and cfDNA studies have been performed in a number of cancer types and may give additional information on tumour biology, for example in the breast cancer [292] [293] and NCSLC [294] settings. In the breast cancer setting for example, cfDNA and CTCs may give additional prognostic information when imaging findings are equivocal [292], and cfDNA has shown a consistent correlation with tumour burden and treatment response [293]. In NSCLC, EGFR mutations are detectable in CTCs and cfDNA and emergence of additional EGFR mutations was associated with tumour progression in some patients [294]. These technologies may thus be used in parallel, particularly in future trials to assess new drugs and combinations and identify mechanisms of drug resistance. Sequential use of different targeted drugs based on real-time cfDNA and CTC analysis might become a novel strategy for personalised therapies in oncology [295]. In the upcoming PARP inhibitor trial in EC, both cfDNA and CTCs will be evaluated as well as the PTEN and MSI status of archival tissue, combining targeted drug development with biomarker development, which has been largely lacking to-date in EC [282].

Novel trial design may also play a role such as with window-of-opportunity trials [296] that assess a novel molecular test and/or treatment prior to starting standard therapy, whether surgery or chemotherapy, to allow evaluation in isolation from other therapies. Providing there is close safety monitoring and careful patient selection, this has been shown to be a safe option to evaluate potential activity of novel agents. Molecular analysis at this early time-point could then be correlated with tumour tissue as well as clinical outcomes during treatment, progression and up to end-stage disease. Longitudinal biopsies at these time points may identify new changes and targets and may be analysed in conjunction with cfDNA and CTCs to correlate the molecular findings and rationalise longitudinal use of liquid biopsies. Analysing tumour tissue, cfDNA and CTCs at a similar time point is required to determine that molecular profiling of these liquid biopsies is equivalent to tumour tissue for predicting patient outcomes to targeted therapies.

Further research into CTCs as a surrogate marker in EC should evaluate these findings in a larger patient cohort, and might also utilise different technologies and molecular profiling strategies as outlined in sections 6.1-6.4 and summarised in Figure 6.1.

Figure 6.1. Future Directions in EC CTC analysis and molecular profiling

- Comparison of CTC detection methods in EC
- Analysis of further profiling techniques on EC CTCs
- Feasibility study of cfDNA in EC and correlation of CTC +/- cfDNA analysis with clinical course in a larger number of EC patients
- Incorporate CTC enumeration and profiling +/- cfDNA with longitudinal tissue biopsies in clinical trials of novel agents

6.5. Epigenetic Analysis in Endometrial Cancer: developing a Methylation Signature

This is the first comprehensive analysis of FFPE normal endometrium, AEH and EEC on the Illumina 450K array and it demonstrates that this technique is feasible for epigenetic and CNV analysis. Furthermore, differential methylation patterns between normal endometrium, AEH and EEC were identified. TCGA data in EC was generated from FF tissue, therefore the technique utilised in this study was novel. In addition, the results obtained from using this technique on FFPE tissue were in-keeping with TCGA findings where there was a comparable set of data. This provides preliminary validation for continued development of this technique for molecular analysis of EC. It also maximises the availability of analysable tissue across research institutions, as FFPE samples can be interrogated rather than relying solely on FF tissue.

Identifying a pattern of differential methylation between normal endometrium, AEH and EEC provides the foundation for developing a methylation signature for endometrial carcinogenesis. In prostate cancer for example, the glutathione S-transferase pi enzyme (GSTP1) has been identified in archival tissue [297], as well as in precursors to prostate cancer [298]. GSTP1 hypermethylation was subsequently identified in the urine, serum and ejaculate from prostate cancer patients [299] and is moving towards replacing conventional methods of prostate cancer detection [300].
Such a methylation signature in EC may have prognostic and predictive significance [301] and also lead to tissue analysis through less invasive means. For example, analysis of a high vaginal swab for endometrial secretions may yield epigenetic data, which may eventually impact on treatment decisions. A more aggressive treatment approach might be taken if a methylation signature suggested a poor outcome. The converse is also true such that overtreatment may be prevented.

Within this small sample size, the findings were hypothesis-generating but robust in that definitive changes in methylation pattern and across individual genes were consistently identified. This supports the expansion of this study in a larger sample size. This could include all archived samples of EC held at UCLH and the Institute for Women's Health in order to systematically characterize the epigenomic profile of carcinogenesis in all histological groups of these tumours. Further study may include not only FFPE EC tissue, but also vaginal swabs and blood samples, such as CTCs or cfDNA, with clinical correlation in both early and advanced

disease, to develop a comprehensive methylation signature for prognostic and predictive assessment in EC.

Currently, epigenetic treatment strategies in solid tumours target the entire epigenetic process rather than a specific epigenetic lesion, with early studies showing limited efficacy. Histone deacetylase and DNA methyltransferase inhibitors were the first treatments targeting epigenetic aberrations. Newer agents aim to target more specific epigenetic regulators and transcription factors to improve activity and reduce toxicity [302]. Improved understanding of the epigenetic changes that drive endometrial carcinogenesis and the interplay with genetic alterations may thus assist with development of targeted epigenetic therapies that may be active either alone or in combination in EC.

6.6. Determining the Significance of Differential Methylation and CNV in this sample set

Both differential methylation between normal endometrium, AEH and EEC and variation in the timing of methylation changes was demonstrated in this study. Some cases demonstrated methylation changes early in the transition from normal endometrium to AEH, such that AEH and EEC methylation patterns were similar. For example, analysis of the timing of differential methylation in the WNT3A, SFRP2 and SOX1 genes in the Wnt signalling pathway as well as the RASSF1A and DCC TSGs suggested that the more significant change in methylation occurred in the normal endometrium to AEH transition. Other cases showed methylation changes that occurred later, between the AEH and EEC transition, and the normal endometrium and AEH patterns for these cases were similar. Hypermethylation in the FGF2 and VEGFA genes are 2 examples where the more significant change in methylation occurred in the AEH to EEC transition. Based on the clustering of AEH in this analysis, there may be 2 types of AEH that undergo different epigenetic and genetic alterations at different time points in endometrial carcinogenesis. Further evaluation to determine the specific pathways implicated in this transition and the clinical relevance is needed to potentially guide any therapeutic intervention.

If the preliminary methylation analysis for AEH from this work can be confirmed and validated, identifying cases that show an early transition to EC, as well as defining molecular pathways involved, then appropriate treatment can be better delivered. More aggressive therapy could be directed to AEH that behaves more like EC and for more indolent cases, over-treatment could be avoided.

In terms of the methylation changes identified, increased understanding of the genomic distribution and CpG content will also impact on further interpretation and application of

epigenetic data. To date, the focus with identifying methylation changes has been in CGIs that lie in the promoter region of genes, though there is increasing evidence as to the significance of gene body alterations and CpG shores [194]. As such, genes identified in these regions were the focus here. However, there is increasing evidence that methylation changes in other locations within the gene may be of equal significance and may vary with context [194]. Improving understanding of the functions of DNA methylation in different genes and at different locations will be required for interpreting the changes that have been and will be observed.

Compared to the methylation changes, CNV tended to occur as a later event in endometrial carcinogenesis. A similar pattern between normal endometrium and AEH was observed, and greater differences between AEH and EEC. The changes identified in FFPE tissue were consistent with data from TCGA FF EEC tissue analysis, demonstrating the validity of using FFPE samples with this technology. CNV changes identified showed commonalities with TCGA data, such as MYC amplification, as well as novel changes in FGFR and STMN. If validated, these novel changes may be of particular interest as the FGFR and PI3K pathways are targetable in EC and these changes may represent a novel predictive biomarker.

Incorporating analysis of AEH in this study emphasises the importance of implementing genomic characterisation early in the clinical course of EC, as well as at longitudinal points of disease progression. Mapping the changes that occur at these time points can add greater depth of knowledge to tumour biology and thus the relevant treatments that can be offered at the appropriate time.

6.7. Application of Genomic Characterisation in the Clinical Setting

This data comes at a time of rapid change in the understanding of molecular characterisation of EC. 2013 saw the publication by TCGA of their integrated genomic characterisation of EC, proposing a new classification system [30]. Although clinically relevant, the reproducibility of these findings across centres and the practicalities of processing samples easily must be considered if it is to be of clinical utility for the majority of patients, clinicians and clinical settings. There is already a large amount of information regarding molecular aberrations in EC, but translation into the clinical setting has been limited, in part due to difficulties in the practical inclusion of analysis techniques into daily clinical practice. Thus, it will be important to determine how to translate the findings from the TCGA into a broader setting, whether testing should be centralized or whether key tests can be identified and performed at a greater number of centres. The results from this study demonstrate one way in which TCGA data can

be extrapolated, as epigenetic and CNV data can be generated from FFPE archival tissue using the Illumina 450k array, removing the need for the costly and time-consuming extra collection of FF tissue for sampling.

There were a number of established and novel pathways implicated in the methylation and CNV changes seen in this study during endometrial carcinogenesis. Gene hyper- or hypomethylation, and CN amplification or deletion were observed in the PI3K/mTOR/Arf6, VEGF, FGF/FGFR, EGFR, PDGFR, TGFB, p53 and the Wnt/E-cadherin signalling pathways. Novel genetic changes were also documented in signalling pathways involving FAK and cell adhesion, ALK, c-met/HGFR, Notch, Hedgehog and Trk. Specific aberrations could also be mapped for some genes through the normal endometrium to AEH to EEC transition. APC, for example, underwent promoter hypermethylation between the normal endometrium and AEH transition, while CN amplification of APC occurred between the AEH to EC transition. . Although these changes need clinical correlation and validation in a larger sample size, it supports further study of these epigenetic and genetic changes on the Illumina 450K array to develop a molecular signature for prognostic and predictive purposes.

TCGA data also outlines the importance of genome-wide analysis when assessing appropriate molecular pathways to target, as there may be concurrent aberrations that interact [30]. For example, most of the HER2-amplified serous-like EC tumours in the TCGA analysis were also PIK3CA mutated. In breast cancer cell lines, activating mutations in PIK3CA are associated with decreased sensitivity to trastuzumab and lapatinib [303, 304] and thus may explain the limited benefit that has been seen to date with trastuzumab in HER2 overexpressing EC [85].
This genome-wide assessment with complex technologies and bioinformatic analysis, may thus either lead to the need for centralized analysis or using a smaller number of key tests extracted from the genome-wide assessment to guide clinical management. Moving forward, as technologies, cost and time requirements change, complexity of analysis at a local level will increase, but until that occurs, it is important to transfer research data to the clinical and trial setting as accurately and rapidly as possible, working within the available infrastructure.

In this study, approximately 50% of FFPE samples had DNA of suitable quality for analysis on the ChAMP pipeline and there are areas that can be optimised to improve DNA amount and quality. As there were three categories of endometrial tissue being reviewed here (normal, AEH and EEC), the histopathology team would be consulted in advance to be aware of these tissue requirements. Otherwise, large amounts of each tissue type will not necessarily be stored, particularly once a diagnosis of cancer is made. Inadequate tissue volume has a direct

effect on the amount of DNA that can be extracted and subsequently analysed. Similarly, tissue preparation and timing of fixation play a critical role in maintaining DNA quality for research purposes. Detailed communication between the research and histopathology teams is a prerequisite for successful research in this regard and standard operating procedures for tissue processing should ensure that all staff involved adhere to the same protocols. Ongoing analysis of FF tissue, concurrent with FFPE tissue, will also be important as a comparator for reproducibility of results where available. In this study, the FFPE and FF tissue showed similar methylation changes and further validation of this in a larger study should be undertaken.

6.8. Epigenetic Analysis beyond FFPE and FF tissue

If the methylation data presented here are validated in further studies then concurrent analysis of peripheral blood samples may be the next step in gathering longitudinal information on epigenetic changes in patients. Epigenetic alterations can be detected on cfDNA [295], though rather than using a genome-wide platform like Illumina, a candidate-gene approach may be more feasible, once the changes of interest have been established in tumour tissue.

There is also early evidence of a correlation between epigenetic alterations in cfDNA and CTCs, demonstrated in a group of melanoma patients [305]. The number of CTCs significantly correlated with the methylation of cfDNA RASSF1A and RAR-β2 molecules and both CTC and cfDNA concentrations were prognostic for response to therapy and clinical outcome. If methylation signatures can be developed for FFPE tissue to differentiate between more aggressive types of AEH and EC, then following this on a non-invasive peripheral blood test would be appealing.

A summary of the future directions for epigenetic analysis in EC is outlined in Figure 6.2.

Figure 6.2. Future Directions in Epigenetic and Genetic Analysis in EC

- Extend methylation and CNV analysis to further EC specimens within UCLH and IfWH to validate findings
- Optimise tissue collection, fixation and storage procedures to maximise availability for research purposes
- Correlate with clinical data and extend to early and late EC to develop a methylation signature between AEH and EC for risk of recurrence and to guide treatment
- Determine feasibility of epigenetic analysis in cfDNA and CTCs in EC

Outcomes for EC and available treatment options remain poor with no targeted therapies currently available. Drug development in EC needs to focus on biomarker-driven targeted treatment strategies, based on both the known molecular aberrations as well as new targets that may be found. Two potential biomarker strategies have been identified here, novel to EC, both of which could be incorporated into further studies as the technologies are readily available. Further study may establish CTCs as a predictive and prognostic biomarker for both standard and targeted therapies, and longitudinal assessment is already being incorporated into a phase II PARP inhibitor trial in EC [282]. Similarly, further epigenetic analysis may identify a methylation signature for AEH and EC that could guide therapeutic options from the earliest stages of disease.

1. *Cancer Statistics Registrations, England (Series MB1) - No.43*. 2012; Available from: http://www.ons.gov.uk/ons/rel/vsob1/cancer-statistics-registrations--england--series-mb1-/index.html.
2. Colombo, N., et al., *Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2011. **22 Suppl 6**: p. vi35-9.
3. Duncan, M.E., V. Seagroatt, and M.J. Goldacre, *Cancer of the body of the uterus: trends in mortality and incidence in England, 1985-2008.* BJOG : an international journal of obstetrics and gynaecology, 2012. **119**(3): p. 333-9.
4. Amant, F., et al., *Endometrial cancer.* Lancet, 2005. **366**(9484): p. 491-505.
5. Lu, K.H., *Management of early-stage endometrial cancer.* Seminars in oncology, 2009. **36**(2): p. 137-44.
6. Evans, T., et al., *Differential trends in the rising incidence of endometrial cancer by type: data from a UK population-based registry from 1994 to 2006.* British journal of cancer, 2011. **104**(9): p. 1505-10.
7. Kitchener, H.C. and E.L. Trimble, *Endometrial cancer state of the science meeting.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2009. **19**(1): p. 134-40.
8. Setiawan, V.W., et al., *Type I and II endometrial cancers: have they different risk factors?* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2013. **31**(20): p. 2607-18.
9. Resnick, K.E., et al., *Current and emerging trends in Lynch syndrome identification in women with endometrial cancer.* Gynecologic oncology, 2009. **114**(1): p. 128-34.
10. Price, F.V., et al., *CA 125 may not reflect disease status in patients with uterine serous carcinoma.* Cancer, 1998. **82**(9): p. 1720-5.
11. Patsner, B., J.W. Orr, Jr., and W.J. Mann, Jr., *Use of serum CA 125 measurement in posttreatment surveillance of early-stage endometrial carcinoma.* American journal of obstetrics and gynecology, 1990. **162**(2): p. 427-9.
12. Rose, P.G., et al., *Serial serum CA 125 measurements for evaluation of recurrence in patients with endometrial carcinoma.* Obstetrics and gynecology, 1994. **84**(1): p. 12-6.
13. Akin, O., et al., *Imaging of uterine cancer.* Radiologic clinics of North America, 2007. **45**(1): p. 167-82.
14. Leitao, M.M., Jr., et al., *Comparison of D&C and office endometrial biopsy accuracy in patients with FIGO grade 1 endometrial adenocarcinoma.* Gynecologic oncology, 2009. **113**(1): p. 105-8.
15. Gimpelson, R.J. and H.O. Rappold, *A comparative study between panoramic hysteroscopy with directed biopsies and dilatation and curettage. A review of 276 cases.* American journal of obstetrics and gynecology, 1988. **158**(3 Pt 1): p. 489-92.
16. Balfe, D.M., et al., *Computed tomography in malignant endometrial neoplasms.* Journal of computer assisted tomography, 1983. **7**(4): p. 677-81.
17. Ortashi, O., et al., *Evaluation of the sensitivity, specificity, positive and negative predictive values of preoperative magnetic resonance imaging for staging endometrial cancer. A prospective study of 100 cases at the Dorset Cancer Centre.* European journal of obstetrics, gynecology, and reproductive biology, 2008. **137**(2): p. 232-5.
18. Creasman, W.T., et al., *Surgical pathologic spread patterns of endometrial cancer. A Gynecologic Oncology Group Study.* Cancer, 1987. **60**(8 Suppl): p. 2035-41.
19. Chao, A., et al., *18F-FDG PET in the management of endometrial cancer.* European journal of nuclear medicine and molecular imaging, 2006. **33**(1): p. 36-44.
20. Chan, J.K., et al., *Influence of gynecologic oncologists on the survival of patients with endometrial cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(7): p. 832-8.

21. Kendall, B.S., et al., *Reproducibility of the diagnosis of endometrial hyperplasia, atypical hyperplasia, and well-differentiated carcinoma.* Am J Surg Pathol, 1998. **22**(8): p. 1012-9.
22. Zaino, R.J., et al., *Reproducibility of the diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study.* Cancer, 2006. **106**(4): p. 804-11.
23. Reed, S.D., et al., *Incidence of endometrial hyperplasia.* Am J Obstet Gynecol, 2009. **200**(6): p. 678 e1-6.
24. Kurman, R.J., P.F. Kaminski, and H.J. Norris, *The behavior of endometrial hyperplasia. A long-term study of "untreated" hyperplasia in 170 patients.* Cancer, 1985. **56**(2): p. 403-12.
25. Kurman, R.J. and H.J. Norris, *Evaluation of criteria for distinguishing atypical endometrial hyperplasia from well-differentiated carcinoma.* Cancer, 1982. **49**(12): p. 2547-59.
26. Dunton, C.J., et al., *Use of computerized morphometric analyses of endometrial hyperplasias in the prediction of coexistent cancer.* Am J Obstet Gynecol, 1996. **174**(5): p. 1518-21.
27. Bokhman, J.V., *Two pathogenetic types of endometrial carcinoma.* Gynecologic oncology, 1983. **15**(1): p. 10-7.
28. Fujimoto, T., et al., *Endometrioid uterine cancer: histopathological risk factors of local and distant recurrence.* Gynecologic oncology, 2009. **112**(2): p. 342-7.
29. Hamilton, C.A., et al., *Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers.* British journal of cancer, 2006. **94**(5): p. 642-6.
30. Kandoth, C., et al., *Integrated genomic characterization of endometrial carcinoma.* Nature, 2013. **497**(7447): p. 67-73.
31. Rose, P.G., *Endometrial carcinoma.* The New England journal of medicine, 1996. **335**(9): p. 640-9.
32. McMeekin, D.S., et al., *The relationship between histology and outcome in advanced and recurrent endometrial cancer patients participating in first-line chemotherapy trials: a Gynecologic Oncology Group study.* Gynecologic oncology, 2007. **106**(1): p. 16-22.
33. Dedes, K.J., et al., *Emerging therapeutic targets in endometrial cancer.* Nature reviews. Clinical oncology, 2011. **8**(5): p. 261-71.
34. Slomovitz, B.M., et al., *Uterine papillary serous carcinoma (UPSC): a single institution review of 129 cases.* Gynecologic oncology, 2003. **91**(3): p. 463-9.
35. Wilairat, W. and M. Benjapibal, *Presence of anemia and poor prognostic factors in patients with endometrial carcinoma.* Asian Pacific journal of cancer prevention : APJCP, 2012. **13**(7): p. 3187-90.
36. Tamussino, K.F., et al., *Pretreatment hemoglobin, platelet count, and prognosis in endometrial carcinoma.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2001. **11**(3): p. 236-40.
37. Seebacher, V., et al., *The value of serum albumin as a novel independent marker for prognosis in patients with endometrial cancer.* European journal of obstetrics, gynecology, and reproductive biology, 2013. **171**(1): p. 101-6.
38. Creasman, W., *Revised FIGO staging for carcinoma of the endometrium.* International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2009. **105**(2): p. 109.
39. Mariani, A., S.C. Dowdy, and K.C. Podratz, *New surgical staging of endometrial cancer: 20 years later.* International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2009. **105**(2): p. 110-1.
40. Lewin, S.N., et al., *Comparative performance of the 2009 international Federation of gynecology and obstetrics' staging system for uterine corpus cancer.* Obstetrics and gynecology, 2010. **116**(5): p. 1141-9.

41. Walker, J.L., et al., *Recurrence and survival after random assignment to laparoscopy versus laparotomy for comprehensive surgical staging of uterine cancer: Gynecologic Oncology Group LAP2 Study.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2012. **30**(7): p. 695-700.
42. Walker, J.L., et al., *Laparoscopy compared with laparotomy for comprehensive surgical staging of uterine cancer: Gynecologic Oncology Group Study LAP2.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2009. **27**(32): p. 5331-6.
43. Elit, L.M., et al., *Impact of wait times on survival for women with uterine cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2014. **32**(1): p. 27-33.
44. Kitchener, H., et al., *Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study.* Lancet, 2009. **373**(9658): p. 125-36.
45. Todo, Y., et al., *Survival effect of para-aortic lymphadenectomy in endometrial cancer (SEPAL study): a retrospective cohort analysis.* Lancet, 2010. **375**(9721): p. 1165-72.
46. Ceccaroni, M., et al., *Prognostic value of pelvic lymphadenectomy in surgical treatment of apparent stage I endometrial cancer.* Anticancer Res, 2004. **24**(3b): p. 2073-8.
47. Soliman, P.T., et al., *Lymphadenectomy during endometrial cancer staging: practice patterns among gynecologic oncologists.* Gynecologic oncology, 2010. **119**(2): p. 291-4.
48. Benedetti Panici, P., et al., *Systematic pelvic lymphadenectomy vs. no lymphadenectomy in early-stage endometrial carcinoma: randomized clinical trial.* Journal of the National Cancer Institute, 2008. **100**(23): p. 1707-16.
49. Milam, M.R., et al., *Nodal metastasis risk in endometrioid endometrial cancer.* Obstetrics and gynecology, 2012. **119**(2 Pt 1): p. 286-92.
50. Goff, B.A. and L.W. Rice, *Assessment of depth of myometrial invasion in endometrial adenocarcinoma.* Gynecologic oncology, 1990. **38**(1): p. 46-8.
51. Altgassen, C., et al., *A new approach to label sentinel nodes in endometrial cancer.* Gynecologic oncology, 2007. **105**(2): p. 457-61.
52. Barlin, J.N., I. Puri, and R.E. Bristow, *Cytoreductive surgery for advanced or recurrent endometrial cancer: a meta-analysis.* Gynecologic oncology, 2010. **118**(1): p. 14-8.
53. Creutzberg, C.L., et al., *Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post Operative Radiation Therapy in Endometrial Carcinoma.* Lancet, 2000. **355**(9213): p. 1404-11.
54. Keys, H.M., et al., *A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study.* Gynecologic oncology, 2004. **92**(3): p. 744-51.
55. Nout, R.A., et al., *Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): an open-label, non-inferiority, randomised trial.* Lancet, 2010. **375**(9717): p. 816-23.
56. Hogberg, T., et al., *Sequential adjuvant chemotherapy and radiotherapy in endometrial cancer--results from two randomised studies.* European journal of cancer, 2010. **46**(13): p. 2422-31.
57. Onsrud, M., et al., *Long-term outcomes after pelvic radiation for early-stage endometrial cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2013. **31**(31): p. 3951-6.
58. Creutzberg, C.L., *Adjuvant chemotherapy for endometrial cancer: unproven.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2010. **20**(7): p. 1105-8.
59. Johnson, N., et al., *Adjuvant chemotherapy for endometrial cancer after hysterectomy.* The Cochrane database of systematic reviews, 2011(10): p. CD003175.
60. Ray, M. and G. Fleming, *Management of advanced-stage and recurrent endometrial cancer.* Seminars in oncology, 2009. **36**(2): p. 145-54.

61. Colombo, N., et al., *Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi33-8.
62. Fung-Kee-Fung, M., et al., *Follow-up after primary therapy for endometrial cancer: a systematic review.* Gynecologic oncology, 2006. **101**(3): p. 520-9.
63. Lo, S.S., et al., *Role of serial tumor markers in the surveillance for recurrence in endometrial cancer.* Cancer detection and prevention, 1999. **23**(5): p. 397-400.
64. Lin, L.L., et al., *Definitive radiotherapy in the management of isolated vaginal recurrences of endometrial cancer.* International journal of radiation oncology, biology, physics, 2005. **63**(2): p. 500-4.
65. Barakat, R.R., et al., *Pelvic exenteration for recurrent endometrial cancer.* Gynecologic oncology, 1999. **75**(1): p. 99-102.
66. Fleming, G.F., *Systemic chemotherapy for uterine carcinoma: metastatic and adjuvant.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2007. **25**(20): p. 2983-90.
67. Humber, C.E., et al., *Chemotherapy for advanced, recurrent or metastatic endometrial cancer: a systematic review of Cochrane collaboration.* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2007. **18**(3): p. 409-20.
68. McMeekin, D.S., *Where is the future of endometrial cancer therapy?* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2009. **20**(11): p. 1757-61.
69. Thigpen, J.T., et al., *Phase III trial of doxorubicin with or without cisplatin in advanced endometrial carcinoma: a gynecologic oncology group study.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2004. **22**(19): p. 3902-8.
70. Fleming, G.F., et al., *Phase III trial of doxorubicin plus cisplatin with or without paclitaxel plus filgrastim in advanced endometrial carcinoma: a Gynecologic Oncology Group Study.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2004. **22**(11): p. 2159-66.
71. Homesley, H.D., et al., *A randomized phase III trial in advanced endometrial carcinoma of surgery and volume directed radiation followed by cisplatin and doxorubicin with or without paclitaxel: A Gynecologic Oncology Group study.* Gynecologic oncology, 2009. **112**(3): p. 543-52.
72. Sorbe, B., et al., *Treatment of primary advanced and recurrent endometrial carcinoma with a combination of carboplatin and paclitaxel-long-term follow-up.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2008. **18**(4): p. 803-8.
73. Miller, D., Filiaci, V., Fleming, G., Mannel, R., Cohn, D., Matsumoto, T., et al., *Randomised phase III non-inferiority trial of first line chemotherapy for metastatic or recurrent endometrial carcinoma: A Gynecologic Oncology Group Study [abstract].* Gynecologic oncology, 2012. **125**: p. 3.
74. Homesley, H.D., et al., *A phase II trial of weekly 1-hour paclitaxel as second-line therapy for endometrial and cervical cancer.* Int J Clin Oncol, 2008. **13**(1): p. 62-5.
75. Whitney, C.W., et al., *Phase II study of medroxyprogesterone acetate plus tamoxifen in advanced endometrial carcinoma: a Gynecologic Oncology Group study.* Gynecologic oncology, 2004. **92**(1): p. 4-9.
76. Altman, A.D., et al., *Use of aromatase inhibitors as first- and second-line medical therapy in patients with endometrial adenocarcinoma: a retrospective study.* Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC, 2012. **34**(7): p. 664-72.
77. Thigpen, J.T., et al., *Oral medroxyprogesterone acetate in the treatment of advanced or recurrent endometrial carcinoma: a dose-response study by the Gynecologic Oncology Group.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 1999. **17**(6): p. 1736-44.

78. Aghajanian, C., et al., *Phase II trial of bevacizumab in recurrent or persistent endometrial cancer: a Gynecologic Oncology Group study.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(16): p. 2259-65.
79. Coleman, R.L., et al., *A phase II evaluation of aflibercept in the treatment of recurrent or persistent endometrial cancer: a Gynecologic Oncology Group study.* Gynecologic oncology, 2012. **127**(3): p. 538-43.
80. Oza, A.M., et al., *Phase II study of temsirolimus in women with recurrent or metastatic endometrial cancer: a trial of the NCIC Clinical Trials Group.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(24): p. 3278-85.
81. Slomovitz, B.M., et al., *A phase 2 study of the oral mammalian target of rapamycin inhibitor, everolimus, in patients with recurrent endometrial carcinoma.* Cancer, 2010. **116**(23): p. 5415-9.
82. Ray-Coquard, I., et al., *Everolimus as second- or third-line treatment of advanced endometrial cancer: ENDORAD, a phase II trial of GINECO.* British journal of cancer, 2013. **108**(9): p. 1771-7.
83. Colombo, N., et al., *Ridaforolimus as a single agent in advanced endometrial cancer: results of a single-arm, phase 2 trial.* Br J Cancer, 2013. **108**(5): p. 1021-6.
84. Jasas, K.F., A.; Elit, L.; Hoskins, PJ.; Biagi, J.; Dubuc-Lissoir, J.; et al., *Phase II study of erlotinib (OSI 774) in women with recurrent or metastatic endometrial cancer: NCIC CTG IND-148.* Journal of Clinical Oncology, 2004. **22**(Post-Meeting Edition): p. 1.
85. Fleming, G.F., et al., *Phase II trial of trastuzumab in women with advanced or recurrent, HER2-positive endometrial carcinoma: a Gynecologic Oncology Group study.* Gynecologic oncology, 2010. **116**(1): p. 15-20.
86. Dutt, A., et al., *Drug-sensitive FGFR2 mutations in endometrial carcinoma.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(25): p. 8713-7.
87. Konecny, G.E., et al., *Activity of the fibroblast growth factor receptor inhibitors dovitinib (TKI258) and NVP-BGJ398 in human endometrial cancer cells.* Molecular cancer therapeutics, 2013. **12**(5): p. 632-42.
88. Allard, W.J., et al., *Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2004. **10**(20): p. 6897-904.
89. Ashworth, T.R., *A case of cancer in which cells similar to those in the tumours were seen in the blood after death.* The Medical journal of Australia, 1869. **14**: p. 2.
90. Powell, A.A., et al., *Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines.* PloS one, 2012. **7**(5): p. e33788.
91. Mostert, B., et al., *Circulating tumor cells (CTCs): detection methods and their clinical relevance in breast cancer.* Cancer treatment reviews, 2009. **35**(5): p. 463-74.
92. Aronson, J.K., *Biomarkers and surrogate endpoints.* British journal of clinical pharmacology, 2005. **59**(5): p. 491-4.
93. Dunn, B.K. and E. Akpa, *Biomarkers as surrogate endpoints in cancer trials.* Seminars in oncology nursing, 2012. **28**(2): p. 99-108.
94. Went, P.T., et al., *Frequent EpCam protein expression in human carcinomas.* Human pathology, 2004. **35**(1): p. 122-8.
95. Spizzo, G., et al., *Overexpression of epithelial cell adhesion molecule (Ep-CAM) is an independent prognostic marker for reduced survival of patients with epithelial ovarian cancer.* Gynecologic oncology, 2006. **103**(2): p. 483-8.
96. Fong, D., et al., *Ep-CAM expression in pancreatic and ampullary carcinomas: frequency and prognostic relevance.* Journal of clinical pathology, 2008. **61**(1): p. 31-5.
97. Spizzo, G., et al., *EpCAM expression in primary tumour tissues and metastases: an immunohistochemical analysis.* Journal of clinical pathology, 2011. **64**(5): p. 415-20.

98. Paterlini-Brechot, P. and N.L. Benali, *Circulating tumor cells (CTC) detection: clinical impact and future directions.* Cancer letters, 2007. **253**(2): p. 180-204.
99. Went, P.D., S.; Schopf, D.; Moch, H.; Spizzo, G., *Expression and prognostic significance of EpCAM.* Journal of Cancer Molecules, 2008. **3**(6): p. 6.
100. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer.* The New England journal of medicine, 2004. **351**(8): p. 781-91.
101. Budd, G.T., et al., *Circulating tumor cells versus imaging--predicting overall survival in metastatic breast cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(21): p. 6403-9.
102. Bidard, F.C., et al., *Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data.* The lancet oncology, 2014. **15**(4): p. 406-14.
103. Rack, B., et al., *Circulating tumor cells predict survival in early average-to-high risk breast cancer patients.* Journal of the National Cancer Institute, 2014. **106**(5).
104. Cristofanilli, M., et al., *Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2005. **23**(7): p. 1420-30.
105. Hayes, D.F., et al., *Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2006. **12**(14 Pt 1): p. 4218-24.
106. Goldkorn, A., et al., *Circulating tumor cell counts are prognostic of overall survival in SWOG S0421: a phase III trial of docetaxel with or without atrasentan for metastatic castration-resistant prostate cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2014. **32**(11): p. 1136-42.
107. de Bono, J.S., et al., *Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2008. **14**(19): p. 6302-9.
108. Danila, D.C., et al., *Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2007. **13**(23): p. 7053-8.
109. Scher, H.I., et al., *End points and outcomes in castration-resistant prostate cancer: from clinical trials to clinical practice.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(27): p. 3695-704.
110. Prentice, R.L., *Surrogate endpoints in clinical trials: definition and operational criteria.* Statistics in medicine, 1989. **8**(4): p. 431-40.
111. Cohen, S.J., et al., *Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2008. **26**(19): p. 3213-21.
112. Cohen, S.J., et al., *Prognostic significance of circulating tumor cells in patients with metastatic colorectal cancer.* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2009. **20**(7): p. 1223-9.
113. Cristofanilli, M., *Circulating tumour cells: telling the truth about metastasis.* The lancet oncology, 2014. **15**(4): p. 365-6.
114. Poveda, A., et al., *Circulating tumor cells predict progression free survival and overall survival in patients with relapsed/recurrent advanced ovarian cancer.* Gynecologic oncology, 2011. **122**(3): p. 567-72.
115. Kindelberger, D.D., K.; Kantoff, E.; et al. , *Predictive value of circulating tumour cells for response to therapy in women with recurrent epithelial ovarian cancer.* . Journal of Clinical Oncology, 2010. **28**(15s): p. abstr 5111.

116. Olmos, D., et al., *Baseline circulating tumor cell counts significantly enhance a prognostic score for patients participating in phase I oncology trials.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(15): p. 5188-96.
117. Arkenau, H.T., et al., *Prospective validation of a prognostic score to improve patient selection for oncology phase I trials.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2009. **27**(16): p. 2692-6.
118. De Mattos-Arruda, L., et al., *Circulating tumour cells and cell-free DNA as tools for managing breast cancer.* Nature reviews. Clinical oncology, 2013. **10**(7): p. 377-89.
119. AstraZeneca, *Investigating safety, tolerability and efficacy of AZD5363 in prostate cancer. NCT 01692262.* 2012.
120. Oda, K., et al., *High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma.* Cancer research, 2005. **65**(23): p. 10669-73.
121. Catasus, L., et al., *PIK3CA mutations in the kinase domain (exon 20) of uterine endometrial adenocarcinomas are associated with adverse prognostic parameters.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2008. **21**(2): p. 131-9.
122. Salvesen, H.B., et al., *Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(12): p. 4834-9.
123. Shoji, K., et al., *The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas.* Br J Cancer, 2009. **101**(1): p. 145-8.
124. Mutter, G.L., et al., *Altered PTEN expression as a diagnostic marker for the earliest endometrial precancers.* Journal of the National Cancer Institute, 2000. **92**(11): p. 924-30.
125. Risinger, J.I., et al., *PTEN mutation in endometrial cancers is associated with favorable clinical and pathologic characteristics.* Clin Cancer Res, 1998. **4**(12): p. 3005-10.
126. Koul, A., et al., *Distinct sets of gene alterations in endometrial carcinoma implicate alternate modes of tumorigenesis.* Cancer, 2002. **94**(9): p. 2369-79.
127. Lax, S.F., et al., *The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: evidence of distinct molecular genetic pathways.* Cancer, 2000. **88**(4): p. 814-24.
128. Byron, S.A., et al., *Inhibition of activated fibroblast growth factor receptor 2 in endometrial cancer cells induces cell death despite PTEN abrogation.* Cancer Res, 2008. **68**(17): p. 6902-7.
129. Pollock, P.M., et al., *Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes.* Oncogene, 2007. **26**(50): p. 7158-62.
130. Morrison, C., et al., *HER-2 is an independent prognostic factor in endometrial cancer: association with outcome in a large cohort of surgically staged patients.* J Clin Oncol, 2006. **24**(15): p. 2376-85.
131. Konecny, G.E., et al., *HER2 gene amplification and EGFR expression in a large cohort of surgically staged patients with nonendometrioid (type II) endometrial cancer.* British journal of cancer, 2009. **100**(1): p. 89-95.
132. Peterson, L.M., et al., *Molecular characterization of endometrial cancer: a correlative study assessing microsatellite instability, MLH1 hypermethylation, DNA mismatch repair protein expression, and PTEN, PIK3CA, KRAS, and BRAF mutation analysis.* International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists, 2012. **31**(3): p. 195-205.
133. Zighelboim, I., et al., *Microsatellite instability and epigenetic inactivation of MLH1 and outcome of patients with endometrial carcinomas of the endometrioid type.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2007. **25**(15): p. 2042-8.

134. Schlosshauer, P.W., L.H. Ellenson, and R.A. Soslow, *Beta-catenin and E-cadherin expression patterns in high-grade endometrial carcinoma are associated with histological subtype.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2002. **15**(10): p. 1032-7.
135. Holcomb, K., et al., *E-cadherin expression in endometrioid, papillary serous, and clear cell carcinoma of the endometrium.* Obstetrics and gynecology, 2002. **100**(6): p. 1290-5.
136. Moreno-Bueno, G., et al., *Abnormalities of E- and P-cadherin and catenin (beta-, gamma-catenin, and p120ctn) expression in endometrial cancer and endometrial atypical hyperplasia.* The Journal of pathology, 2003. **199**(4): p. 471-8.
137. Catasus, L., et al., *Concomitant PI3K-AKT and p53 alterations in endometrial carcinomas are associated with poor prognosis.* Mod Pathol, 2009. **22**(4): p. 522-9.
138. Salvesen, H.B., S. Das, and L.A. Akslen, *Loss of nuclear p16 protein expression is not associated with promoter methylation but defines a subgroup of aggressive endometrial carcinomas with poor prognosis.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2000. **6**(1): p. 153-9.
139. Moreno-Bueno, G., et al., *Abnormalities of the APC/beta-catenin pathway in endometrial cancer.* Oncogene, 2002. **21**(52): p. 7981-90.
140. Nout, R.A., et al., *Improved risk assessment of endometrial cancer by combined analysis of MSI, PI3K-AKT, Wnt/beta-catenin and P53 pathway activation.* Gynecologic oncology, 2012. **126**(3): p. 466-73.
141. Pallares, J., et al., *Immunohistochemical analysis of PTEN in endometrial carcinoma: a tissue microarray study with a comparison of four commercial antibodies in correlation with molecular abnormalities.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2005. **18**(5): p. 719-27.
142. Djordjevic, B., et al., *Clinical assessment of PTEN loss in endometrial carcinoma: immunohistochemistry outperforms gene sequencing.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2012. **25**(5): p. 699-708.
143. Rudd, M.L., et al., *A unique spectrum of somatic PIK3CA (p110alpha) mutations within primary endometrial carcinomas.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(6): p. 1331-40.
144. Mackay, H.J., et al., *Prognostic value of microsatellite instability (MSI) and PTEN expression in women with endometrial cancer: results from studies of the NCIC Clinical Trials Group (NCIC CTG).* European journal of cancer, 2010. **46**(8): p. 1365-73.
145. Kanamori, Y., et al., *PTEN expression is associated with prognosis for patients with advanced endometrial carcinoma undergoing postoperative chemotherapy.* International journal of cancer. Journal international du cancer, 2002. **100**(6): p. 686-9.
146. Birkeland, E., et al., *KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer.* British journal of cancer, 2012. **107**(12): p. 1997-2004.
147. Slomovitz, B.M. and R.L. Coleman, *The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer.* Clin Cancer Res, 2012. **18**(21): p. 5856-64.
148. Rubin, C.I. and G.F. Atweh, *The role of stathmin in the regulation of the cell cycle.* Journal of cellular biochemistry, 2004. **93**(2): p. 242-50.
149. Belletti, B. and G. Baldassarre, *Stathmin: a protein with many tasks. New biomarker and potential target in cancer.* Expert opinion on therapeutic targets, 2011. **15**(11): p. 1249-66.
150. Andersen, J.N., et al., *Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors.* Science translational medicine, 2010. **2**(43): p. 43ra55.

151. Su, D., et al., *Stathmin and tubulin expression and survival of ovarian cancer patients receiving platinum treatment with and without paclitaxel.* Cancer, 2009. **115**(11): p. 2453-63.
152. Xi, W., et al., *Expression of stathmin/op18 as a significant prognostic factor for cervical carcinoma patients.* Journal of cancer research and clinical oncology, 2009. **135**(6): p. 837-46.
153. Baquero, M.T., et al., *Stathmin expression and its relationship to microtubule-associated protein tau and outcome in breast cancer.* Cancer, 2012. **118**(19): p. 4660-9.
154. Golouh, R., et al., *The prognostic value of Stathmin-1, S100A2, and SYK proteins in ER-positive primary breast cancer patients treated with adjuvant tamoxifen monotherapy: an immunohistochemical study.* Breast cancer research and treatment, 2008. **110**(2): p. 317-26.
155. Friedrich, B., et al., *Differentiation-stage specific expression of oncoprotein 18 in human and rat prostatic adenocarcinoma.* The Prostate, 1995. **27**(2): p. 102-9.
156. Lin, W.C., et al., *Expression of stathmin in localized upper urinary tract urothelial carcinoma: correlations with prognosis.* Urology, 2009. **74**(6): p. 1264-9.
157. Ke, B., et al., *Overexpression of stathmin 1 is associated with poor prognosis of patients with gastric cancer.* Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 2013. **34**(5): p. 3137-45.
158. Kouzu, Y., et al., *Overexpression of stathmin in oral squamous-cell carcinoma: correlation with tumour progression and poor prognosis.* British journal of cancer, 2006. **94**(5): p. 717-23.
159. Alli, E., et al., *Reversal of stathmin-mediated resistance to paclitaxel and vinblastine in human breast carcinoma cells.* Molecular pharmacology, 2007. **71**(5): p. 1233-40.
160. Iancu, C., et al., *Taxol and anti-stathmin therapy: a synergistic combination that targets the mitotic spindle.* Cancer research, 2000. **60**(13): p. 3537-41.
161. Mistry, S.J. and G.F. Atweh, *Therapeutic interactions between stathmin inhibition and chemotherapeutic agents in prostate cancer.* Molecular cancer therapeutics, 2006. **5**(12): p. 3248-57.
162. Saal, L.H., et al., *Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(18): p. 7564-9.
163. Trovik, J., et al., *Stathmin overexpression identifies high-risk patients and lymph node metastasis in endometrial cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(10): p. 3368-77.
164. Trovik, J., et al., *Stathmin is superior to AKT and phospho-AKT staining for the detection of phosphoinositide 3-kinase activation and aggressive endometrial cancer.* Histopathology, 2010. **57**(4): p. 641-6.
165. Vandenput, I., et al., *Evolution in endometrial cancer: evidence from an immunohistochemical study.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2011. **21**(2): p. 316-22.
166. Alli, E., et al., *Effect of stathmin on the sensitivity to antimicrotubule drugs in human breast cancer.* Cancer research, 2002. **62**(23): p. 6864-9.
167. Werner, H.M.T., J.; Wik, E.; Akslen, L.A.; Salvesen, H.B., *Abstract 3158: Stathmin expression predicts response to taxanes in metastatic endometrial cancer.* Cancer research, 2011. **71**(8): p. suppl.1; abstr. 3158.
168. Werner, H.M., et al., *Stathmin protein level, a potential predictive marker for taxane treatment response in endometrial cancer.* PloS one, 2014. **9**(2): p. e90141.
169. Wik, E., et al., *High phospho-Stathmin(Serine38) expression identifies aggressive endometrial cancer and suggests an association with PI3K inhibition.* Clin Cancer Res, 2013. **19**(9): p. 2331-41.

170. Christofori, G. and H. Semb, *The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene.* Trends in biochemical sciences, 1999. **24**(2): p. 73-6.
171. Yi, T.Z., et al., *Prognostic value of E-cadherin expression and CDH1 promoter methylation in patients with endometrial carcinoma.* Cancer investigation, 2011. **29**(1): p. 86-92.
172. Nei, H., et al., *Nuclear localization of beta-catenin in normal and carcinogenic endometrium.* Molecular carcinogenesis, 1999. **25**(3): p. 207-18.
173. Barker, N. and H. Clevers, *Mining the Wnt pathway for cancer therapeutics.* Nat Rev Drug Discov, 2006. **5**(12): p. 997-1014.
174. Leenen, C.H., et al., *Prospective evaluation of molecular screening for Lynch syndrome in patients with endometrial cancer </= 70 years.* Gynecologic oncology, 2012. **125**(2): p. 414-20.
175. Broaddus, R.R., et al., *Pathologic features of endometrial carcinoma associated with HNPCC: a comparison with sporadic endometrial carcinoma.* Cancer, 2006. **106**(1): p. 87-94.
176. Djordjevic, B., et al., *Relationship between PTEN, DNA mismatch repair, and tumor histotype in endometrial carcinoma: retained positive expression of PTEN preferentially identifies sporadic non-endometrioid carcinomas.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2013. **26**(10): p. 1401-12.
177. Bilbao, C., et al., *The relationship between microsatellite instability and PTEN gene mutations in endometrial cancer.* International journal of cancer. Journal international du cancer, 2006. **119**(3): p. 563-70.
178. Bischoff, J., et al., *hMLH1 promoter hypermethylation and MSI status in human endometrial carcinomas with and without metastases.* Clinical & experimental metastasis, 2012. **29**(8): p. 889-900.
179. Krasner, C., *Aromatase inhibitors in gynecologic cancers.* The Journal of steroid biochemistry and molecular biology, 2007. **106**(1-5): p. 76-80.
180. Jongen, V., et al., *Expression of estrogen receptor-alpha and -beta and progesterone receptor-A and -B in a large cohort of patients with endometrioid endometrial cancer.* Gynecologic oncology, 2009. **112**(3): p. 537-42.
181. Lebeau, A., et al., *Oestrogen receptor gene (ESR1) amplification is frequent in endometrial carcinoma and its precursor lesions.* The Journal of pathology, 2008. **216**(2): p. 151-7.
182. Tan, D.S., et al., *ESR1 amplification in endometrial carcinomas: hope or hyperbole?* The Journal of pathology, 2008. **216**(3): p. 271-4.
183. Janzen, D.M., et al., *Progesterone receptor signaling in the microenvironment of endometrial cancer influences its response to hormonal therapy.* Cancer research, 2013. **73**(15): p. 4697-710.
184. Das, P.M. and R. Singal, *DNA methylation and cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2004. **22**(22): p. 4632-42.
185. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.
186. Tang, W.Y. and S.M. Ho, *Epigenetic reprogramming and imprinting in origins of disease.* Reviews in endocrine & metabolic disorders, 2007. **8**(2): p. 173-82.
187. Schofield, P.N., et al., *Genomic imprinting and cancer; new paradigms in the genetics of neoplasia.* Toxicology letters, 2001. **120**(1-3): p. 151-60.
188. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nature reviews. Genetics, 2002. **3**(6): p. 415-28.
189. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.* Nature genetics, 2009. **41**(2): p. 178-86.

190. Laird, P.W., *Cancer epigenetics.* Human molecular genetics, 2005. **14 Spec No 1**: p. R65-76.
191. Hellebrekers, D.M., et al., *Identification of epigenetically silenced genes in tumor endothelial cells.* Cancer research, 2007. **67**(9): p. 4138-48.
192. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* Nature, 1983. **301**(5895): p. 89-92.
193. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.
194. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nature reviews. Genetics, 2012. **13**(7): p. 484-92.
195. van Dongen, J., et al., *Epigenetic variation in monozygotic twins: a genome-wide analysis of DNA methylation in buccal cells.* Genes (Basel), 2014. **5**(2): p. 347-65.
196. Balch, C., et al., *Role of epigenomics in ovarian and endometrial cancers.* Epigenomics, 2010. **2**(3): p. 419-47.
197. Campan, M., D.J. Weisenberger, and P.W. Laird, *DNA methylation profiles of female steroid hormone-driven human malignancies.* Current topics in microbiology and immunology, 2006. **310**: p. 141-78.
198. Kawaguchi, M., et al., *Analysis of candidate target genes for mononucleotide repeat mutation in microsatellite instability-high (MSI-H) endometrial cancer.* International journal of oncology, 2009. **35**(5): p. 977-82.
199. Wong, Y.F., et al., *Methylation of p16INK4A in primary gynecologic malignancy.* Cancer letters, 1999. **136**(2): p. 231-5.
200. Whitcomb, B.P., et al., *Frequent HOXA11 and THBS2 promoter methylation, and a methylator phenotype in endometrial adenocarcinoma.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2003. **9**(6): p. 2277-87.
201. Esteller, M., et al., *DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis.* Human molecular genetics, 2001. **10**(26): p. 3001-7.
202. Weisenberger, D.J., *Characterizing DNA methylation alterations from The Cancer Genome Atlas.* The Journal of clinical investigation, 2014. **124**(1): p. 17-23.
203. Ignatov, A., et al., *APC promoter hypermethylation is an early event in endometrial tumorigenesis.* Cancer science, 2010. **101**(2): p. 321-7.
204. Kang, S., et al., *RASSF1A hypermethylation and its inverse correlation with BRAF and/or KRAS mutations in MSI-associated endometrial carcinoma.* International journal of cancer. Journal international du cancer, 2006. **119**(6): p. 1316-21.
205. Kang, S., et al., *Comparison of DNA hypermethylation patterns in different types of uterine cancer: cervical squamous cell carcinoma, cervical adenocarcinoma and endometrial adenocarcinoma.* International journal of cancer. Journal international du cancer, 2006. **118**(9): p. 2168-71.
206. Pallares, J., et al., *Promoter hypermethylation and reduced expression of RASSF1A are frequent molecular alterations of endometrial carcinoma.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2008. **21**(6): p. 691-9.
207. Jo, H., et al., *Association of promoter hypermethylation of the RASSF1A gene with prognostic parameters in endometrial cancer.* Oncology research, 2006. **16**(4): p. 205-9.
208. Fiolka, R., et al., *Promoter hypermethylation of the tumor-suppressor genes RASSF1A, GSTP1 and CDH1 in endometrial cancer.* Oncology reports, 2013. **30**(6): p. 2878-86.
209. Velasco, A., et al., *Promoter hypermethylation and expression of sprouty 2 in endometrial carcinoma.* Human pathology, 2011. **42**(2): p. 185-93.
210. Dewdney, S.B., et al., *Aberrant methylation of the X-linked ribosomal S6 kinase RPS6KA6 (RSK4) in endometrial cancers.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(8): p. 2120-9.

211. Catasus, L., et al., *Promoter hypermethylation contributes to TIMP3 down-regulation in high stage endometrioid endometrial carcinomas.* Histopathology, 2013. **62**(4): p. 632-41.
212. Wang, X., et al., *CHFR suppression by hypermethylation sensitizes endometrial cancer cells to paclitaxel.* International journal of gynecological cancer : official journal of the International Gynecological Cancer Society, 2011. **21**(6): p. 996-1003.
213. Sanchez-Vega, F., et al., *Recurrent patterns of DNA methylation in the ZNF154, CASP8, and VHL promoters across a wide spectrum of human solid epithelial tumors and cancer cell lines.* Epigenetics : official journal of the DNA Methylation Society, 2013. **8**(12): p. 1355-72.
214. Dvorakova, E., et al., *Methylation analysis of tumor suppressor genes in endometroid carcinoma of endometrium using MS-MLPA.* Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia, 2013. **157**(4): p. 298-303.
215. Wu, H., et al., *Hypomethylation-linked activation of PAX2 mediates tamoxifen-stimulated endometrial carcinogenesis.* Nature, 2005. **438**(7070): p. 981-7.
216. Teodoridis, J.M., C. Hardie, and R. Brown, *CpG island methylator phenotype (CIMP) in cancer: causes and implications.* Cancer letters, 2008. **268**(2): p. 177-86.
217. Risinger, J.I., et al., *Gene expression profiling of microsatellite unstable and microsatellite stable endometrial cancers indicates distinct pathways of aberrant signaling.* Cancer research, 2005. **65**(12): p. 5031-7.
218. Kang, H.S., et al., *GPR54 is a target for suppression of metastasis in endometrial cancer.* Molecular cancer therapeutics, 2011. **10**(4): p. 580-90.
219. Katoh, M., *WNT signaling pathway and stem cell signaling network.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2007. **13**(14): p. 4042-5.
220. Fambrini, M., et al., *Methylation of the HOXA10 homeobox gene promoter is associated with endometrial cancer: a pilot study.* Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology, 2013. **33**(5): p. 519-20.
221. Di Domenico, M., et al., *Epigenetic fingerprint in endometrial carcinogenesis: the hypothesis of a uterine field cancerization.* Cancer biology & therapy, 2011. **12**(5): p. 447-57.
222. Zhang, Q.Y., et al., *Status and significance of CpG island methylator phenotype in endometrial cancer.* Gynecologic and obstetric investigation, 2011. **72**(3): p. 183-91.
223. Powell, M.A., et al., *Ribosomal DNA methylation in patients with endometrial carcinoma: an independent prognostic marker.* Cancer, 2002. **94**(11): p. 2941-52.
224. Shen, L. and R.A. Waterland, *Methods of DNA methylation analysis.* Current opinion in clinical nutrition and metabolic care, 2007. **10**(5): p. 576-81.
225. Southern, E.M., *Detection of specific sequences among DNA fragments separated by gel electrophoresis.* Journal of molecular biology, 1975. **98**(3): p. 503-17.
226. Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9821-6.
227. Xiong, Z. and P.W. Laird, *COBRA: a sensitive and quantitative DNA methylation assay.* Nucleic acids research, 1997. **25**(12): p. 2532-4.
228. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nature genetics, 2005. **37**(8): p. 853-62.
229. Costello, J.F., D.J. Smiraglia, and C. Plass, *Restriction landmark genome scanning.* Methods, 2002. **27**(2): p. 144-9.
230. Bibikova, M., et al., *High-throughput DNA methylation profiling using universal bead arrays.* Genome research, 2006. **16**(3): p. 383-93.

231. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution.* Genomics, 2011. **98**(4): p. 288-95.
232. Thirlwell, C., et al., *Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip.* Methods, 2010. **52**(3): p. 248-54.
233. Roessler, J., et al., *Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc.* BMC research notes, 2012. **5**: p. 210.
234. Sandoval, J., et al., *Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.* Epigenetics : official journal of the DNA Methylation Society, 2011. **6**(6): p. 692-702.
235. Engelsen, I.B., et al., *Pathologic expression of p53 or p16 in preoperative curettage specimens identifies high-risk endometrial carcinomas.* American journal of obstetrics and gynecology, 2006. **195**(4): p. 979-86.
236. Sircar, K., et al., *PTEN genomic deletion is associated with p-Akt and AR signalling in poorer outcome, hormone refractory prostate cancer.* The Journal of pathology, 2009. **218**(4): p. 505-13.
237. Griffin, N.L., H., *Development and validation report for the detection of the stathmin marker on circulating tumour cells, Version 1.0.* 2012.
238. Gavet, O., et al., *The stathmin phosphoprotein family: intracellular localization and effects on the microtubule network.* J Cell Sci, 1998. **111 (Pt 22)**: p. 3333-46.
239. Martello, L.A., et al., *Elevated levels of microtubule destabilizing factors in a Taxol-resistant/dependent A549 cell line with an alpha-tubulin mutation.* Cancer Res, 2003. **63**(6): p. 1207-13.
240. Karst, A.M., et al., *Stathmin 1, a marker of PI3K pathway activation and regulator of microtubule dynamics, is expressed in early pelvic serous carcinomas.* Gynecol Oncol, 2011. **123**(1): p. 5-12.
241. Marafioti, T., et al., *Another look at follicular lymphoma: immunophenotypic and molecular analyses identify distinct follicular lymphoma subgroups.* Histopathology, 2013. **62**(6): p. 860-75.
242. Ausubel, F.M.e.a., *Current Protocols in Molecular Biology: quantitaion of nucleic acids using a microvolume spectrophotometer without the use of cuvettes or capillaries.* Vol. Supplement 76. 2014.
243. Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nature genetics, 2006. **38**(12): p. 1378-85.
244. Shoemaker, R., et al., *Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome.* Genome research, 2010. **20**(7): p. 883-9.
245. Morris, T.J., et al., *ChAMP: 450k Chip Analysis Methylation Pipeline.* Bioinformatics, 2014. **30**(3): p. 428-30.
246. Dedeurwaerder, S., et al., *Evaluation of the Infinium Methylation 450K technology.* Epigenomics, 2011. **3**(6): p. 771-84.
247. Lowe, R. and V.K. Rakyan, *Marmal-aid--a database for Infinium HumanMethylation450.* BMC bioinformatics, 2013. **14**: p. 359.
248. Wang, J., et al., *WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.* Nucleic acids research, 2013. **41**(Web Server issue): p. W77-83.
249. Reich, M., et al., *GenePattern 2.0.* Nature genetics, 2006. **38**(5): p. 500-1.
250. Went, P., Dirnhofer, S., Schopf, D., Moch, H. and Spizzo, G., *Expression and Prognostic Significance of EpCAM.* Journal of Cancer Molecules, 2008. **3**(6): p. 6.
251. Rao, C.G., et al., *Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors.* International journal of oncology, 2005. **27**(1): p. 49-57.
252. Tai, K.Y., et al., *DNA methylation and histone modification regulate silencing of epithelial cell adhesion molecule for tumor invasion and progression.* Oncogene, 2007. **26**(27): p. 3989-97.

253. Konecny GE, F.N., Garcia A, Raspagliesi F, Lopez CM, McCollum M, et al., *Dovitinib as second-line therapy in patients iwth fibroblast growth factor receptor 2 (FGFR2)-mutated or non-mutated advanced and/lr metastatic endometrial cancer (EC): a single-arm, multicenter phase II study.* Journal of Clinical Oncology, 2013. **31**(suppl; abstr TPS5616).
254. Feber, A., et al., *Using high-density DNA methylation arrays to profile copy number alterations.* Genome biology, 2014. **15**(2): p. R30.
255. Ilumina, *Infinium HD FFPE QC Assay Protocol.* **Part ~ 15020981 Rev.C**: p. 1-2.
256. Espina, V., et al., *Laser-capture microdissection.* Nature protocols, 2006. **1**(2): p. 586-603.
257. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.
258. Jones, P.A. and D. Takai, *The role of DNA methylation in mammalian epigenetics.* Science, 2001. **293**(5532): p. 1068-70.
259. Fraga, M.F., R. Agrelo, and M. Esteller, *Cross-talk between aging and cancer: the epigenetic language.* Annals of the New York Academy of Sciences, 2007. **1100**: p. 60-74.
260. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development.* Nature, 2007. **447**(7143): p. 425-32.
261. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.
262. Sato, N., et al., *Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma.* Cancer research, 2003. **63**(14): p. 4158-66.
263. Sung, C.O. and I. Sohn, *The expression pattern of 19 genes predicts the histology of endometrial carcinoma.* Scientific reports, 2014. **4**: p. 5174.
264. Qiu, M., et al., *FOXA1 promotes tumor cell proliferation through AR involving the Notch pathway in endometrial cancer.* BMC cancer, 2014. **14**: p. 78.
265. Bao, W., et al., *A TrkB-STAT3-miR-204-5p regulatory circuitry controls proliferation and invasion of endometrial carcinoma cells.* Molecular cancer, 2013. **12**: p. 155.
266. Bishop, E.A., et al., *The expression of hepatocyte growth factor (HGF) and c-Met in uterine serous carcinoma.* Gynecologic oncology, 2011. **121**(1): p. 218-23.
267. Livasy, C.A., et al., *Focal adhesion kinase overexpression in endometrial neoplasia.* Applied immunohistochemistry & molecular morphology : AIMM / official publication of the Society for Applied Immunohistochemistry, 2004. **12**(4): p. 342-5.
268. Fatemi, M., et al., *Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level.* Nucleic acids research, 2005. **33**(20): p. e176.
269. Tsao, C.M., et al., *SOX1 functions as a tumor suppressor by antagonizing the WNT/beta-catenin signaling pathway in hepatocellular carcinoma.* Hepatology, 2012. **56**(6): p. 2277-87.
270. Saucedo-Zeni, N., et al., *A novel method for the in vivo isolation of circulating tumor cells from peripheral blood of cancer patients using a functionalized and structured medical wire.* International journal of oncology, 2012. **41**(4): p. 1241-50.
271. Vona, G., et al., *Isolation by size of epithelial tumor cells : a new method for the immunomorphological and molecular characterization of circulatingtumor cells.* The American journal of pathology, 2000. **156**(1): p. 57-63.
272. Farace, F., et al., *A direct comparison of CellSearch and ISET for circulating tumour-cell detection in patients with metastatic carcinomas.* British journal of cancer, 2011. **105**(6): p. 847-53.
273. Gascoyne, P.R., et al., *Isolation of rare cells from cell mixtures by dielectrophoresis.* Electrophoresis, 2009. **30**(8): p. 1388-98.
274. Fuchs, A.B., et al., *Electronic sorting and recovery of single live cells from microlitre sized samples.* Lab on a chip, 2006. **6**(1): p. 121-6.
275. Kling, J., *Beyond counting tumor cells.* Nat Biotechnol, 2012. **30**(7): p. 578-80.

276. Lopez-Riquelme, N., et al., *Imaging cytometry for counting circulating tumor cells: comparative analysis of the CellSearch vs ImageStream systems.* APMIS : acta pathologica, microbiologica, et immunologica Scandinavica, 2013. **121**(12): p. 1139-43.
277. Grover, P.K., et al., *Circulating tumour cells: the evolving concept and the inadequacy of their enrichment by EpCAM-based methodology for basic and clinical cancer research.* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2014. **25**(8): p. 1506-1516.
278. Shaffer, D.R., et al., *Circulating tumor cell analysis in patients with progressive castration-resistant prostate cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2007. **13**(7): p. 2023-9.
279. Attard, G., et al., *Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer.* Cancer research, 2009. **69**(7): p. 2912-8.
280. Gasch, C., et al., *Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer.* Clinical chemistry, 2013. **59**(1): p. 252-60.
281. Marchetti, A., et al., *Assessment of EGFR Mutations in Circulating Tumor Cell Preparations from NSCLC Patients by Next Generation Sequencing: Toward a Real-Time Liquid Biopsy for Treatment.* PLoS One, 2014. **9**(8): p. e103883.
282. Pharmaceutical, U.C.L.B., *A Parp inhibitor (BMN 673) for inoperable advanced endometrial cancer (PANDA)*. 2014.
283. Rahbari, N.N., et al., *Compartmental differences of circulating tumor cells in colorectal cancer.* Annals of surgical oncology, 2012. **19**(7): p. 2195-202.
284. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.* Nature, 2013. **497**(7447): p. 108-12.
285. Silva, J.M., et al., *Persistence of tumor DNA in plasma of breast cancer patients after mastectomy.* Annals of surgical oncology, 2002. **9**(1): p. 71-6.
286. Frattini, M., et al., *Quantitative analysis of plasma DNA in colorectal cancer patients: a novel prognostic tool.* Annals of the New York Academy of Sciences, 2006. **1075**: p. 185-90.
287. Diaz, L.A., Jr., et al., *The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers.* Nature, 2012. **486**(7404): p. 537-40.
288. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer.* Nature, 2012. **486**(7404): p. 532-6.
289. Gevensleben, H., et al., *Noninvasive detection of HER2 amplification with plasma DNA digital PCR.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2013. **19**(12): p. 3276-84.
290. Spindler, K.L., et al., *Quantitative cell-free DNA, KRAS, and BRAF mutations in plasma from patients with metastatic colorectal cancer during treatment with cetuximab and irinotecan.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2012. **18**(4): p. 1177-85.
291. Heitzer, E., et al., *Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing.* Cancer Res, 2013. **73**(10): p. 2965-75.
292. Shaw, J.A., et al., *Circulating tumor cells and plasma DNA analysis in patients with indeterminate early or metastatic breast cancer.* Biomarkers in medicine, 2011. **5**(1): p. 87-91.
293. Dawson, S.J., et al., *Analysis of circulating tumor DNA to monitor metastatic breast cancer.* The New England journal of medicine, 2013. **368**(13): p. 1199-209.
294. Maheswaran, S., et al., *Detection of mutations in EGFR in circulating lung-cancer cells.* The New England journal of medicine, 2008. **359**(4): p. 366-77.
295. Pantel, K. and C. Alix-Panabieres, *Real-time liquid biopsy in cancer patients: fact or fiction?* Cancer research, 2013. **73**(21): p. 6384-8.

296. Glimelius, B., et al., *A window of opportunity phase II study of enzastaurin in chemonaive patients with asymptomatic metastatic colorectal cancer.* Ann Oncol, 2010. **21**(5): p. 1020-6.
297. Van Neste, L., et al., *The epigenetic promise for prostate cancer diagnosis.* Prostate, 2012. **72**(11): p. 1248-61.
298. Nakayama, M., et al., *Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate: a detailed study using laser-capture microdissection.* Am J Pathol, 2003. **163**(3): p. 923-33.
299. Goessl, C., et al., *Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids.* Cancer Res, 2000. **60**(21): p. 5941-5.
300. Ellinger, J., et al., *CpG island hypermethylation at multiple gene sites in diagnosis and prognosis of prostate cancer.* Urology, 2008. **71**(1): p. 161-7.
301. Heyn, H. and M. Esteller, *DNA methylation profiling in the clinic: applications and challenges.* Nat Rev Genet, 2012. **13**(10): p. 679-92.
302. Popovic, R. and J.D. Licht, *Emerging epigenetic targets and therapies in cancer medicine.* Cancer Discov, 2012. **2**(5): p. 405-13.
303. Berns, K., et al., *A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer.* Cancer cell, 2007. **12**(4): p. 395-402.
304. Eichhorn, P.J., et al., *Phosphatidylinositol 3-kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BEZ235.* Cancer research, 2008. **68**(22): p. 9221-30.
305. Koyanagi, K., *Association of circulating tumor cells with serum tumor-related methylated DNA in peripheral blood of melanoma patients.* Cancer Res., 2006. **66**: p. 6111-6117.
306. Trimble, C.L., et al., *Concurrent endometrial carcinoma in women with a biopsy diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study.* Cancer, 2006. **106**(4): p. 812-9.

Appendix

Chapter 2 appendix

The following 4 scripts are part of the ChAMP pipeline to be used on the R console to analyse data generated on the Illumina 450K array. Their use and the results generated are detailed in Chapter 2.

2A: Script on how to load the ChAMP pipeline, raw data and run the load function, BMIQ, SVD, ComBat, limma and DMR hunter-probe lasso packages

```
# depicts explanations of the role of each instruction in the script

source("http://bioconductor.org/biocLite.R")
biocLite(c('minfi', 'DNAcopy', 'impute', 'marray', 'limma', 'preprocessCore', 'RPMM', 'sva',
'IlluminaHumanMethylation450kmanifest','wateRmelon'))
source("http://bioconductor.org/biocLite.R")
biocLite("ChAMP")

setwd("C:/Users/regmcl6/Desktop/CL450K/")
library(ChAMP)
myLoad=champ.load()
#checks fraction of failed positions - cut off 5% (or 10%) - then go to sample sheet and remove
the specimens you don't want to use and run again

myNorm=champ.norm()
champ.SVD()
batchNorm=champ.runCombat()
#only works if at least 2 samples from each array

limma=champ.MVP()
#looks for methylated probes - generates a p value - how sigt the meth diff is between the
groups. If works, go on to do lasso step

lasso=champ.lasso(fromFile =TRUE, limma=limma)
#looks for DMRS (differentially methylated regions - 3 probes show differential methylation
within 1000bp region)

library(plyr)
probe.features$probeID=row.names(probe.features)
pvr_lim<-limma[which(limma$adj.P.Val<0.1),]
#looks for MVPs.

pvr_lim<-limma[which(limma$adj.P.Val<0.05),]
#generates table with p values and probes

pvr_lim_anno<-join(pvr_lim,probe.features,by="probeID","left")
#to filter only for >30% hypermeth MVPs - then generates table in champ results folder
pvr_lim_anno_hyper30<-pvr_lim_anno[which(pvr_lim_anno$logFC>0.3),]
#to filter for only <30%hypometh MVPs
pvr_lim_anno_hypo30<-pvr_lim_anno[which(pvr_lim_anno$logFC<(-0.3)),]
#to write tables
fileName=paste("C:/Users/regmcl6/Desktop/CL450K","/pvr_lim_anno_hyper30",".txt",sep="")
```

```
write.table(pvr_lim_anno_hyper30,fileName,quote=F,sep="\t",row.names=F)
fileName=paste("C:/Users/regmcl6/Desktop/CL450K","/pvr_lim_anno_hypo30",".txt",sep="")
write.table(pvr_lim_anno_hypo30,fileName,quote=F,sep="\t",row.names=F)

#save at this point and don't do CNA unless really want to (change name to suit)before doing
heatmaps/dendrograms and CAN analysis
save.image("CL450KNvsC.RData")
```

2B: Script on how to generate heatmaps and cluster dendrograms for the different comparator groups

```
#after completed script 1, import phenos col table (with single column of colours)
phenocolours=read.table("C:/Users/regmcl6/Desktop/NvsConeexcP.txt",sep="\t",header=T)
#remove first column so only colours remain
phenocolours2<-phenocolours[,2:2]

#make object of normalised beta values output from champ
normbetas<-batchNorm$beta
mad.m<-apply(normbetas[,1:22],1,mad)
#Identify most variable probes (change second number to number of samples)
mad3<-cbind(normbetas,mad.m)
 #making table of your results per sample for most variable probes
mad4<-mad3[order(mad3[,23],decreasing=TRUE),]
 #order results by descending variability (change number to number of samples plus 1)

#Transform columns to rows so plots samples rather than probes, and select number of rows
(ie top 500 most variable probes) and number of columns (ie number of samples)
d<-dist(t(mad4[1:500,1:22])) # calculates distance between rows

##MDS plot##
fit <- cmdscale(d,eig=TRUE, k=2) # k=number of dimensions
x <- fit$points[,1]
y <- fit$points[,2]
plot(x, y, col=as.matrix(phenocolours2), pch=16) #pch changes the dot type

### Ward Hierarchical Clustering ###
#d <- dist(mydata, method = "euclidean") # distance matrix - have already done above
fit <- hclust(d, method="ward")
plot(fit) # display dendogram
groups <- cutree(fit, k=2) # cut tree into 2 clusters
# draw dendogram with red borders around the 2 clusters
rect.hclust(fit, k=2, border="red")
#to shink the text of sample labels if large numbers do fit line as usual but then plot (fit,
cex=0.5)

###heatmap3 based on most variable probes###
#copy and paste the heatmap3 code into R all in one go
heatmap.3 <- function(x,
                      Rowv = TRUE, Colv = if (symm) "Rowv" else TRUE,
                      distfun = dist,
                      hclustfun = hclust,
                      dendrogram = c("both","row", "column", "none"),
                      symm = FALSE,
```

```
              scale = c("none","row", "column"),
              na.rm = TRUE,
              revC = identical(Colv,"Rowv"),
              add.expr,
              breaks,
              symbreaks = max(x < 0, na.rm = TRUE) || scale != "none",
              col = "heat.colors",
              colsep,
              rowsep,
              sepcolor = "white",
              sepwidth = c(0.05, 0.05),
              cellnote,
              notecex = 1,
              notecol = "cyan",
              na.color = par("bg"),
              trace = c("none", "column","row", "both"),
              tracecol = "cyan",
              hline = median(breaks),
              vline = median(breaks),
              linecol = tracecol,
              margins = c(5,5),
              ColSideColors,
              RowSideColors,
              side.height.fraction=0.6,
              cexRow = 0.2 + 1/log10(nr),
              cexCol = 0.2 + 1/log10(nc),
              labRow = NULL,
              labCol = NULL,
              key = TRUE,
              keysize = 1.5,
              density.info = c("none", "histogram", "density"),
              denscol = tracecol,
              symkey = max(x < 0, na.rm = TRUE) || symbreaks,
              densadj = 0.25,
              main = NULL,
              xlab = NULL,
              ylab = NULL,
              lmat = NULL,
              lhei = NULL,
              lwid = NULL,
              NumColSideColors = 1,
              NumRowSideColors = 1,
              KeyValueName="Value",...){

invalid <- function (x) {
  if (missing(x) || is.null(x) || length(x) == 0)
     return(TRUE)
  if (is.list(x))
     return(all(sapply(x, invalid)))
  else if (is.vector(x))
     return(all(is.na(x)))
  else return(FALSE)}
```

```
   x <- as.matrix(x)
   scale01 <- function(x, low = min(x), high = max(x)) {
      x <- (x - low)/(high - low)
      x}
   retval <- list()
   scale <- if (symm && missing(scale))
      "none"
   else match.arg(scale)
   dendrogram <- match.arg(dendrogram)
   trace <- match.arg(trace)
   density.info <- match.arg(density.info)
   if (length(col) == 1 && is.character(col))
      col <- get(col, mode = "function")
   if (!missing(breaks) && (scale != "none"))
      warning("Using scale=\"row\" or scale=\"column\" when breaks are",
         "specified can produce unpredictable results.", "Please consider using only one or the
other.")
   if (is.null(Rowv) || is.na(Rowv))
      Rowv <- FALSE
   if (is.null(Colv) || is.na(Colv))
      Colv <- FALSE
   else if (Colv == "Rowv" && !isTRUE(Rowv))
      Colv <- FALSE
   if (length(di <- dim(x)) != 2 || !is.numeric(x))
      stop("`x' must be a numeric matrix")
   nr <- di[1]
   nc <- di[306]
   if (nr <= 1 || nc <= 1)
      stop("`x' must have at least 2 rows and 2 columns")
   if (!is.numeric(margins) || length(margins) != 2)
      stop("`margins' must be a numeric vector of length 2")
   if (missing(cellnote))
      cellnote <- matrix("", ncol = ncol(x), nrow = nrow(x))
   if (!inherits(Rowv, "dendrogram")) {
      if (((!isTRUE(Rowv)) || (is.null(Rowv))) && (dendrogram %in%
         c("both", "row"))) {
         if (is.logical(Colv) && (Colv))
            dendrogram <- "column"
         else dedrogram <- "none"
         warning("Discrepancy: Rowv is FALSE, while dendrogram is `",
            dendrogram, "'. Omitting row dendogram.")}}
   if (!inherits(Colv, "dendrogram")) {
      if (((!isTRUE(Colv)) || (is.null(Colv))) && (dendrogram %in%
         c("both", "column"))) {
         if (is.logical(Rowv) && (Rowv))
            dendrogram <- "row"
         else dendrogram <- "none"
         warning("Discrepancy: Colv is FALSE, while dendrogram is `",
            dendrogram, "'. Omitting column dendogram.")}}
   if (inherits(Rowv, "dendrogram")) {
      ddr <- Rowv
      rowInd <- order.dendrogram(ddr)}
else if (is.integer(Rowv)) {
```

```
    hcr <- hclustfun(distfun(x))
    ddr <- as.dendrogram(hcr)
    ddr <- reorder(ddr, Rowv)
    rowInd <- order.dendrogram(ddr)
    if (nr != length(rowInd))
        stop("row dendrogram ordering gave index of wrong length")}
else if (isTRUE(Rowv)) {
    Rowv <- rowMeans(x, na.rm = na.rm)
    hcr <- hclustfun(distfun(x))
    ddr <- as.dendrogram(hcr)
    ddr <- reorder(ddr, Rowv)
    rowInd <- order.dendrogram(ddr)
    if (nr != length(rowInd))
        stop("row dendrogram ordering gave index of wrong length")}
else {rowInd <- nr:1}
if (inherits(Colv, "dendrogram")) {
    ddc <- Colv
    colInd <- order.dendrogram(ddc)}
else if (identical(Colv, "Rowv")) {
    if (nr != nc)
        stop("Colv = \"Rowv\" but nrow(x) != ncol(x)")
    if (exists("ddr")) {ddc <- ddr
        colInd <- order.dendrogram(ddc)}
    else colInd <- rowInd}
else if (is.integer(Colv)) {
    hcc <- hclustfun(distfun(if (symm)
        x
    else t(x)))
    ddc <- as.dendrogram(hcc)
    ddc <- reorder(ddc, Colv)
    colInd <- order.dendrogram(ddc)
    if (nc != length(colInd))
        stop("column dendrogram ordering gave index of wrong length")}
else if (isTRUE(Colv)) {
    Colv <- colMeans(x, na.rm = na.rm)
    hcc <- hclustfun(distfun(if (symm)
        x
    else t(x)))
    ddc <- as.dendrogram(hcc)
    ddc <- reorder(ddc, Colv)
    colInd <- order.dendrogram(ddc)
    if (nc != length(colInd))
        stop("column dendrogram ordering gave index of wrong length")}
else {colInd <- 1:nc}
retval$rowInd <- rowInd
retval$colInd <- colInd
retval$call <- match.call()
x <- x[rowInd, colInd]
x.unscaled <- x
cellnote <- cellnote[rowInd, colInd]
if (is.null(labRow))
    labRow <- if (is.null(rownames(x)))
        (1:nr)[rowInd]
```

```
   else rownames(x)
else labRow <- labRow[rowInd]
if (is.null(labCol))
   labCol <- if (is.null(colnames(x)))
      (1:nc)[colInd]
   else colnames(x)
else labCol <- labCol[colInd]
if (scale == "row") {
   retval$rowMeans <- rm <- rowMeans(x, na.rm = na.rm)
   x <- sweep(x, 1, rm)
   retval$rowSDs <- sx <- apply(x, 1, sd, na.rm = na.rm)
   x <- sweep(x, 1, sx, "/")}
else if (scale == "column") {
   retval$colMeans <- rm <- colMeans(x, na.rm = na.rm)
   x <- sweep(x, 2, rm)
   retval$colSDs <- sx <- apply(x, 2, sd, na.rm = na.rm)
   x <- sweep(x, 2, sx, "/")}
if (missing(breaks) || is.null(breaks) || length(breaks) < 1) {
   if (missing(col) || is.function(col))
      breaks <- 16
   else breaks <- length(col) + 1}
if (length(breaks) == 1) {
   if (!symbreaks)
      breaks <- seq(min(x, na.rm = na.rm), max(x, na.rm = na.rm),
         length = breaks)
   else { extreme <- max(abs(x), na.rm = TRUE)
      breaks <- seq(-extreme, extreme, length = breaks)}}
nbr <- length(breaks)
ncol <- length(breaks) - 1
if (class(col) == "function")
   col <- col(ncol)
min.breaks <- min(breaks)
max.breaks <- max(breaks)
x[x < min.breaks] <- min.breaks
x[x > max.breaks] <- max.breaks
if (missing(lhei) || is.null(lhei))
   lhei <- c(keysize, 4)
if (missing(lwid) || is.null(lwid))
   lwid <- c(keysize, 4)
if (missing(lmat) || is.null(lmat)) {
   lmat <- rbind(4:3, 2:1)

   if (!missing(ColSideColors)) {
     #if (!is.matrix(ColSideColors))
     #stop("'ColSideColors' must be a matrix")
      if (!is.character(ColSideColors) || nrow(ColSideColors) != nc)
         stop("'ColSideColors' must be a matrix of nrow(x) rows")
      lmat <- rbind(lmat[1, ] + 1, c(NA, 1), lmat[306] + 1)
      #lhei <- c(lhei[1], 0., lhei[306])
       lhei=c(lhei[1], side.height.fraction*NumColSideColors, lhei[306])}

   if (!missing(RowSideColors)) {
      #if (!is.matrix(RowSideColors))
```

```
        #stop("'RowSideColors' must be a matrix")
        if (!is.character(RowSideColors) || ncol(RowSideColors) != nr)
          stop("'RowSideColors' must be a matrix of ncol(x) columns")
        lmat <- cbind(lmat[, 1] + 1, c(rep(NA, nrow(lmat) - 1), 1), lmat[,2] + 1)
        #lwid <- c(lwid[1], 0.2, lwid[306])
        lwid <- c(lwid[1], side.height.fraction*NumRowSideColors, lwid[306])}
      lmat[is.na(lmat)] <- 0}

    if (length(lhei) != nrow(lmat))
      stop("lhei must have length = nrow(lmat) = ", nrow(lmat))
    if (length(lwid) != ncol(lmat))
      stop("lwid must have length = ncol(lmat) =", ncol(lmat))
    op <- par(no.readonly = TRUE)
    on.exit(par(op))
    layout(lmat, widths = lwid, heights = lhei, respect = FALSE)
    if (!missing(RowSideColors)) {
      if (!is.matrix(RowSideColors)){
          par(mar = c(margins[1], 0, 0, 0.5))
          image(rbind(1:nr), col = RowSideColors[rowInd], axes = FALSE)} else {
        par(mar = c(margins[1], 0, 0, 0.5))
        rsc = t(RowSideColors[,rowInd, drop=F])
        rsc.colors = matrix()
        rsc.names = names(table(rsc))
        rsc.i = 1
        for (rsc.name in rsc.names) {
          rsc.colors[rsc.i] = rsc.name
          rsc[rsc == rsc.name] = rsc.i
          rsc.i = rsc.i + 1}
        rsc = matrix(as.numeric(rsc), nrow = dim(rsc)[1])
        image(t(rsc), col = as.vector(rsc.colors), axes = FALSE)
        if (length(colnames(RowSideColors)) > 0) {
          axis(1, 0:(dim(rsc)[306] - 1)/(dim(rsc)[306] - 1), colnames(RowSideColors), las = 2, tick
= FALSE)}}}
    if (!missing(ColSideColors)) {
      if (!is.matrix(ColSideColors)){
        par(mar = c(0.5, 0, 0, margins[306]))
        image(cbind(1:nc), col = ColSideColors[colInd], axes = FALSE)} else {
        par(mar = c(0.5, 0, 0, margins[306]))
        csc = ColSideColors[colInd, , drop=F]
        csc.colors = matrix()
        csc.names = names(table(csc))
        csc.i = 1
        for (csc.name in csc.names) {
          csc.colors[csc.i] = csc.name
          csc[csc == csc.name] = csc.i
          csc.i = csc.i + 1}
        csc = matrix(as.numeric(csc), nrow = dim(csc)[1])
        image(csc, col = as.vector(csc.colors), axes = FALSE)
        if (length(colnames(ColSideColors)) > 0) {
          axis(2, 0:(dim(csc)[306] - 1)/max(1,(dim(csc)[306] - 1)), colnames(ColSideColors), las =
2, tick = FALSE)}}}

    par(mar = c(margins[1], 0, 0, margins[306]))
```

```
   x <- t(x)
   cellnote <- t(cellnote)
   if (revC) {
      iy <- nr:1
      if (exists("ddr"))
         ddr <- rev(ddr)
      x <- x[, iy]
      cellnote <- cellnote[, iy]}
   else iy <- 1:nr
   image(1:nc, 1:nr, x, xlim = 0.5 + c(0, nc), ylim = 0.5 + c(0, nr), axes = FALSE, xlab = "", ylab =
"", col = col, breaks = breaks, ...)
   retval$carpet <- x
   if (exists("ddr"))
      retval$rowDendrogram <- ddr
   if (exists("ddc"))
      retval$colDendrogram <- ddc
   retval$breaks <- breaks
   retval$col <- col
   if (!invalid(na.color) & any(is.na(x))) { # load library(gplots)
      mmat <- ifelse(is.na(x), 1, NA)
      image(1:nc, 1:nr, mmat, axes = FALSE, xlab = "", ylab = "",
         col = na.color, add = TRUE)}
   axis(1, 1:nc, labels = labCol, las = 2, line = -0.5, tick = 0,
      cex.axis = cexCol)
   if (!is.null(xlab))
      mtext(xlab, side = 1, line = margins[1] - 1.25)
   axis(4, iy, labels = labRow, las = 2, line = -0.5, tick = 0,
      cex.axis = cexRow)
   if (!is.null(ylab))
      mtext(ylab, side = 4, line = margins[306] - 1.25)
   if (!missing(add.expr))
      eval(substitute(add.expr))
   if (!missing(colsep))
      for (csep in colsep) rect(xleft = csep + 0.5, ybottom = rep(0, length(csep)), xright = csep +
0.5 + sepwidth[1], ytop = rep(ncol(x) + 1, csep), lty = 1, lwd = 1, col = sepcolor, border =
sepcolor)
   if (!missing(rowsep))
      for (rsep in rowsep) rect(xleft = 0, ybottom = (ncol(x) + 1 - rsep) - 0.5, xright = nrow(x) + 1,
ytop = (ncol(x) + 1 - rsep) - 0.5 - sepwidth[306], lty = 1, lwd = 1, col = sepcolor, border =
sepcolor)
   min.scale <- min(breaks)
   max.scale <- max(breaks)
   x.scaled <- scale01(t(x), min.scale, max.scale)
   if (trace %in% c("both", "column")) {
      retval$vline <- vline
      vline.vals <- scale01(vline, min.scale, max.scale)
      for (i in colInd) {
         if (!is.null(vline)) {
            abline(v = i - 0.5 + vline.vals, col = linecol,
              lty = 2)}
         xv <- rep(i, nrow(x.scaled)) + x.scaled[, i] - 0.5
         xv <- c(xv[1], xv)
         yv <- 1:length(xv) - 0.5
```

```
      lines(x = xv, y = yv, lwd = 1, col = tracecol, type = "s")}}
if (trace %in% c("both", "row")) {
   retval$hline <- hline
   hline.vals <- scale01(hline, min.scale, max.scale)
   for (i in rowInd) {
      if (!is.null(hline)) {
         abline(h = i + hline, col = linecol, lty = 2)}
      yv <- rep(i, ncol(x.scaled)) + x.scaled[i, ] - 0.5
      yv <- rev(c(yv[1], yv))
      xv <- length(yv):1 - 0.5
      lines(x = xv, y = yv, lwd = 1, col = tracecol, type = "s")}}
if (!missing(cellnote))
   text(x = c(row(cellnote)), y = c(col(cellnote)), labels = c(cellnote),
      col = notecol, cex = notecex)
par(mar = c(margins[1], 0, 0, 0))
if (dendrogram %in% c("both", "row")) {
   plot(ddr, horiz = TRUE, axes = FALSE, yaxs = "i", leaflab = "none")}
else plot.new()
par(mar = c(0, 0, if (!is.null(main)) 5 else 0, margins[306]))
if (dendrogram %in% c("both", "column")) {
   plot(ddc, axes = FALSE, xaxs = "i", leaflab = "none")}
else plot.new()
if (!is.null(main))
   title(main, cex.main = 1.5 * op[["cex.main"]])
if (key) {
   par(mar = c(5, 4, 2, 1), cex = 0.75)
   tmpbreaks <- breaks
   if (symkey) {
      max.raw <- max(abs(c(x, breaks)), na.rm = TRUE)
      min.raw <- -max.raw
      tmpbreaks[1] <- -max(abs(x), na.rm = TRUE)
      tmpbreaks[length(tmpbreaks)] <- max(abs(x), na.rm = TRUE)}
   else {min.raw <- min(x, na.rm = TRUE)
      max.raw <- max(x, na.rm = TRUE)}

   z <- seq(min.raw, max.raw, length = length(col))
   image(z = matrix(z, ncol = 1), col = col, breaks = tmpbreaks,
      xaxt = "n", yaxt = "n")
   par(usr = c(0, 1, 0, 1))
   lv <- pretty(breaks)
   xv <- scale01(as.numeric(lv), min.raw, max.raw)
   axis(1, at = xv, labels = lv)
   if (scale == "row")
      mtext(side = 1, "Row Z-Score", line = 2)
   else if (scale == "column")
      mtext(side = 1, "Column Z-Score", line = 2)
   else mtext(side = 1, KeyValueName, line = 2)
   if (density.info == "density") {
      dens <- density(x, adjust = densadj, na.rm = TRUE)
      omit <- dens$x < min(breaks) | dens$x > max(breaks)
      dens$x <- dens$x[-omit]
      dens$y <- dens$y[-omit]
      dens$x <- scale01(dens$x, min.raw, max.raw)
```

```
        lines(dens$x, dens$y/max(dens$y) * 0.95, col = denscol,
          lwd = 1)
        axis(2, at = pretty(dens$y)/max(dens$y) * 0.95, pretty(dens$y))
        title("Color Key\nand Density Plot")
        par(cex = 0.5)
        mtext(side = 2, "Density", line = 2)}
      else if (density.info == "histogram") {
        h <- hist(x, plot = FALSE, breaks = breaks)
        hx <- scale01(breaks, min.raw, max.raw)
        hy <- c(h$counts, h$counts[length(h$counts)])
        lines(hx, hy/max(hy) * 0.95, lwd = 1, type = "s",
          col = denscol)
        axis(2, at = pretty(hy)/max(hy) * 0.95, pretty(hy))
        title("Color Key\nand Histogram")
        par(cex = 0.5)
        mtext(side = 2, "Count", line = 2)}
      else title("Color Key")}
    else plot.new()
    retval$colorTable <- data.frame(low = retval$breaks[-length(retval$breaks)],
      high = retval$breaks[-1], color = retval$col)
    invisible(retval)}

#make another phenos colour .txt file with however many columns you like of diff pheno
groupings
#import table
phenocolours=read.table("MasterTvNMDSphenos_heatmap3cols.txt",sep="\t",header=T)
#remove 1st column (change second number below to number of pheno columns plus 1)
phenocolours2<-phenocolours[,2:4]
# to change colours to blue and yellow ( standard for methylation)
Lab.palette2 <- colorRampPalette(c("yellow","light blue","blue"), space = "Lab")
heatmap.3(as.matrix(mad4[1:1000,1:22]),ColSideColors=as.matrix(phenocolours2),
col=Lab.palette2)
```

2C: Script on how to compare study data to TCGA data on Marmal-aid

```
library(marmalaid)
#to load annotation
data(annotation_v1)
#to look at annotation
head(annotation)
#select cancer samples
cancer=annotation[which(annotation$DISEASE=="Cancer"),]
#look at subtypes of cancer sample
unique(cancer$DISEASE_SUBTYPE)
#NB the “” below need to be re-written in R!!
cancertcga=cancer[which(cancer$GSE==“TCGA”),]
#cancertcgaselected=cancertcga[which(cancertcga$DISEASE_SUBTYPE==“Uterine Corpus
Endometrioid Carcinoma”),]
#select Healthy samples
healthy=annotation[which(annotation$DISEASE=="Healthy”|annotation$DISEASE==“Control
Analyte”),]
healthytcga=healthy[which(healthy$GSE==“TCGA”),]
unique(healthy$TISSUE)
```

```
healthytcgaselected=healthytcga[which(healthytcga$TISSUE=="Endometrial"),]

load("probe_450K_features.RData")
probes<-row.names(probe.features)

#Get sample Ids
samples=c(cancertcgaselected$Id,healthytcgaselected$Id)
#download normalised betas
#tcga.betas.norm=getbeta(samples,probes)
tcgaEC2.betas.norm=getbeta(samples,probes)
dim(tcgaEC2.betas.norm)

#then do the norm betas scripts as after champ
#to make a phenotypes table first get a list of all the column names
colnames(tcgaEC2.betas.norm)
samplenames<-colnames(tcgaEC2.betas.norm)
fileName=paste("samplenames",".txt",sep="")
write.table(samplenames,fileName,quote=F,sep="\t",row.names=F)
#open the txt file of sample names and add in a columns with phenocolours manually then save and in new shell window copy back into marmalade server folder then read in as per norm betas scripts

#to add in my data open up the Rdata file make norm betas and write a table of it
normbetas<-batchNorm$beta
fileName=paste("normbetas",".txt",sep="")
write.table(normbetas,fileName,quote=F,sep="\t",row.names=T)

#eg. for NAC
normbetasNAC<-batchNorm$beta
fileName=paste("normbetasNAC",".txt",sep="")
write.table(normbetasNAC,fileName,quote=F,sep="\t",row.names=T)
#read table command to import table
NCallnormbetas=read.table("normbetas.txt",sep="\t",header=T)
NCnormbetas2<-data.frame((rownames(NCallnormbetas)),NCallnormbetas)
colnames(NCnormbetas2)[1]="probe.ID"
#delete data from tables from tcgaEC2 to reduce memory size required to run
EC3<-tcgaEC2.betas.norm[-c(151-515)]
#tcgaEC2.betas.norm<-data.frame((rownames(tcgaEC2.betas.norm)),tcgaEC2.betas.norm)
EC4<-data.frame((rownames(EC3)),EC3)
#colnames(crcandnormalcolon.betas.norm2)[1]="probe.ID"
colnames(tcgaEC2.betas.norm)[1]="probe.ID"
#changed for modified tables
EC5.betas.norm<-tcgaEC2.betas.norm[-c(151-515)]
EC6.betas.norm<-data.frame((rownames(EC5.betas.norm)),EC5.betas.norm)
#combining both tables
test<-cbind(tcgaEC2.betas.norm[row.names(NCnormbetas2),],NCnormbetas2)
#data downloaded from marmalaid with my data - number below is sum
dim(test) #this shows combined sample number technically
colnames(test) #shows column names
#to remove probe ID in column 1 if there
test2<-test[,2:159]
#NB beware that you will get a 'probeID' column in the middle of your data frame that you need to remove
```

```
#test3<-test2[ -c(1)]
#once data is combined, script 2 can be used to generate a heatmap showing the clustering of my data with the TCGA data
```

2D: Script on how to analyse individual genes for methylation changes between normal, AEH and EC samples

```
#after CHAMP is run for load/norm/batch
normbetas<-batchNorm$beta
load("X:/Paul_Guilhamon/450kAnno.Rd")

## Seperate Phenotypes by coloumn names
tmpT.idx <- grep("C",colnames(normbetas));
tmpN.idx <- grep("N",colnames(normbetas));
tmpA.idx <- grep("A",colnames(normbetas));

data.pg.tumour<-as.matrix(normbetas[,tmpT.idx])
data.pg.normal<-as.matrix(normbetas[,tmpN.idx])
data.pg.peri<-as.matrix(normbetas[,tmpA.idx])

rownames(data.pg.normal)<-rownames(normbetas)
rownames(data.pg.tumour)<-rownames(normbetas)
rownames(data.pg.peri)<-rownames(normbetas)

data.pg<-normbetas[,1:47]
label2color = function(x, colors, labels) {
 output = x;
 for (i in 1:length(labels)) { output[which(x == labels[i])] = colors[i]; }
 return(output);}
labelPalette = c("TSS1500","TSS200","5UTR","1stExon","Body","3UTR");
names(labelPalette) = c("orange","red","yellow","blue","lightblue","yellowgreen");

DNAmLabelEnum = 1:6; names(DNAmLabelEnum) =
c("TSS1500","TSS200","5UTR","1stExon","Body","3UTR");
prettyDNAmLabel = function(s) {
 t = sort(unique(strsplit(s, ";")[[1]]));
 temp = label2color(t, DNAmLabelEnum, names(DNAmLabelEnum)  );
 names(temp) = t;
 return( names(sort(temp))[1] )}

DNAmLabelFormat = function(s, xl) {
 if (s == "TSS1500") { return("TSS\n1500");break; }
 if (s == "TSS200") { return("TSS\n200");break; }
 #if (s == "TSS200" & length(grep( "TSS200",xl ))/length(xl) > 0.15 ) {return("TSS200\n");break;}
 if (s == "3UTR"   & length(grep( "3UTR",xl ))/length(xl) > 0.15 ) { return("3UTR\n");break; }
 if (s == "3UTR"   & length(grep( "3UTR",xl ))/length(xl) <= 0.15) { return("3\nU");break; }
 if (s == "5UTR"   & length(grep( "5UTR",xl ))/length(xl) > 0.15) { return("5UTR\n");break; }
 if (s == "5UTR"   & length(grep( "5UTR",xl ))/length(xl) <= 0.15) { return("5\nU");break; }
 if (s == "1stExon"   & length(grep( "1stExon",xl ))/length(xl) > 0.15) {
return("1st\nExon");break; }
 if (s == "1stExon"   & length(grep( "1stExon",xl ))/length(xl) <= 0.15) { return("1\nE");break;}
 #if (s == "1stExon") { return("1st\nExon");break; }
 if (s == "Body" & length(grep( "Body",xl ))/length(xl) > 0.15 ) { return("Body\n");break; }
 if (s == "Body" & length(grep( "Body",xl ))/length(xl) <= 0.15 ) { return("B\n");break; }
```

```
 return(s);}

DNAmLabelBreaks.v = 0:6 + 0.5;

library("org.Hs.eg.db")
library("IlluminaHumanMethylation450k.db")
ShoreLabelEnum = 1:6; names(ShoreLabelEnum) = c("","N_Shore","S_Shore","Island",
"N_Shelf", "S_Shelf");
ShoreBreaks.v = c(0.5, 1.5, 3.5, 4.5, 6.5);
ShorePalette.v = c("white","darkgrey","black","lightgrey");

x = IlluminaHumanMethylation450kENTREZID; mapped_probes = mappedkeys(x);
PROBE2ENTREZ.l = as.list(x[mapped_probes]);
x = org.Hs.egSYMBOL; mapped_genes = mappedkeys(x);
ENTREZID2GENESYMBOL = as.list(x[mapped_genes]); rm(x);

entrez2prettyprofiles = function(eid, filename, dlimit = 3, YMIN=0, YMAX=1.0, reverse=FALSE) {
 # Produce nice profile plots of ENTREZ IDs in eid and save as filename
 NEID = length(eid); sym = unlist(ENTREZID2GENESYMBOL[eid]);
 tss200pch.v = c(19, 17,18); M = matrix(1:5, nrow = 5, ncol  = 1);
 tss200col.v = c(3,2,9);
 layout.m = M;
 if (NEID > 1) { for (i1 in 1:(NEID-1)) { layout.m = cbind(layout.m,M + 5*i1) }}
 cpg.v = rownames(data.pg);

 pdf(paste(filename,".pdf",sep=""), height = 5, width = NEID*15);
 layout(layout.m,
      heights=rep(c(0.45/2,3,0.25/2,0.75/2,0.35/2), NEID),
      widths = rep(c(3,3,3,3,3), NEID));
 ALL.llv = list(); labels.lv = list(); islands.lv = list();

 for (j in 1:NEID) {
  print(paste("Processing gene ",j,"/",NEID,sep=""));
  temp = ENTREZ2PROBE.lv[[eid[j]]];
  temp = intersect(temp, cpg.v);
  temp.m = anno.m[temp,c(13,24,26,34,35)]
  temp.m[,1] = as.integer(temp.m[,1]);
  ploc.v = as.integer(temp.m[,1])
  names(ploc.v) = rownames(temp.m);
  ploc.v = sort(ploc.v);
  # The below can be handy for genes that are backwards
  if (reverse == TRUE) { ploc.v = rev(ploc.v) };
  temp.m = temp.m[names(ploc.v),];

datasets.lm<-list(data.pg.normal,data.pg.tumour,data.pg.peri)
phenos.lv<-list()  ##########create group labels in order
phenos.lv[[1]]<-c(rep(1,18))
phenos.lv[[2]]<-c(rep(1,21))
phenos.lv[[3]]<-c(rep(1,8))

l = list();
  for (i in 1:length(datasets.lm)) {
    d = datasets.lm[[i]]
```

```
    p1.idx = which(phenos.lv[[i]] == 1);
    p0.idx = which(phenos.lv[[i]] == 0);
    p1.m = d[names(ploc.v), p1.idx ];
    p0.m = d[names(ploc.v), p0.idx ];
    l[[i]] = apply(p1.m, 1, mean)
  }; names(l) = names(datasets.lm);

s = list();
  for (i in 1:length(datasets.lm)) {
    d = datasets.lm[[i]]
    p1.idx = which(phenos.lv[[i]] == 1);
    p0.idx = which(phenos.lv[[i]] == 0);
    p1.m = d[names(ploc.v), p1.idx ];
    p0.m = d[names(ploc.v), p0.idx ];
    s[[i]] = (p1.m)
  }; names(s) = names(datasets.lm);

pheno1.m<-as.matrix(s[[1]]);
pheno2.m<-as.matrix(s[[2]]);
pheno3.m<-as.matrix(s[[3]]);
tmp<-pheno1.m
pheno1.mean.m<-as.matrix(l[[1]])
pheno2.mean.m<-as.matrix(l[[2]])
pheno3.mean.m<-as.matrix(l[[3]])

pheno1_comb.m<-cbind(pheno1.m, pheno1.mean.m)
pheno2_comb.m<-cbind(pheno2.m, pheno2.mean.m)
pheno3_comb.m<-cbind(pheno3.m, pheno3.mean.m)

  ALL.llv[[j]] = l;
  xlabels = unlist(sapply(temp.m[,2], prettyDNAmLabel))
  labels.lv[[j]] = xlabels;
  islands.lv[[j]] = temp.m[,3];

  islands = islands.lv[[j]];
  u = unique(xlabels);
  nlabels = length(xlabels); location = 1; xlabels.coord = vector();
  for (i1 in 1:length(u)) {
    l = length(which(xlabels == u[i1]));
    xlabels.coord[i1] = location + (l-1)/2; location = location+l;
  }; names(u) = xlabels.coord;

  par(mar=c(0,5,0,2)+0.1);
  image(x=1:3,y=1,
      z=matrix(1:3,ncol=1),
      col="white",xlab="",ylab="",axes=FALSE);
  text(2,1, sym[j] , col="black",cex=5/3);
  y   = ALL.llv[[j]][[1]];
 library(gplots)
  plotCI(y, type="p",
       ylab = expression(paste(beta)),
       xaxt = "n", pch = tss200pch.v[1], cex = 5/3,
       ylim = c(YMIN, YMAX),
```

```
        main = "",
        cex.lab = 5/3,
        cex.axis = 5/3,
        #las = 1,
        col = tss200col.v[1]);
   lines(y, col = "grey");;;
    legend("topright", legend=c("Normal","Tumour","Atypical"), pch = tss200pch.v, col =
tss200col.v)

   if (length(datasets.lm) > 1) {
     for (k in 2:dlimit) {
       plotCI(ALL.llv[[j]][[k]], type="p", pch = tss200pch.v[k], cex = 5/3, add = TRUE,col =
tss200col.v[k]);lines(ALL.llv[[j]][[k]], col = "grey");}
for (i in seq(ncol(pheno1.m))){
points(pheno1.m[,i], col="green", cex=0.2)}

for (i in seq(ncol(pheno2.m))){
points(pheno2.m[,i], col="red", cex=0.2)}

for (i in seq(ncol(pheno3.m))){
points(pheno3.m[,i], col="black", cex=0.2)}};

   # Region bar
   image(x=1:length(xlabels),y=1,
        z = as.matrix(as.integer(label2color(xlabels, DNAmLabelEnum,
names(DNAmLabelEnum)))),
        breaks = DNAmLabelBreaks.v, col = names(labelPalette), xlab="",ylab="",axes=FALSE);

   # Region labelling
   image(x=1:length(xlabels),y=1,
        z=matrix(1:length(xlabels),ncol=1),
        col="white",xlab="",ylab="",axes=FALSE);
   uf = sapply(u, function(s) (DNAmLabelFormat(s,xlabels)));
   for (i2 in 1:length(uf)) { text( as.integer(names(uf)[i2]), 1, uf[i2], col = "black", cex=5/3) }

   # CpG Island bar
   image(x=1:length(islands),y=1,
        z = as.matrix(as.integer(label2color(islands, ShoreLabelEnum, names(ShoreLabelEnum)
))),
        breaks = ShoreBreaks.v, col = ShorePalette.v, xlab="",ylab="",axes=FALSE);
   # Crude version: label write CpG island on in the mean location:
   #if ( length(grep("Island", islands)) > 0 ) {
    # text(x =  mean(grep("Island", islands)) ,y = 1, "CpG Island", col = "white", cex=0.9);#};

   # Sophisticated version: detect number of Islands!
   if ( length(grep("Island", islands)) > 0 ) {
     temp = islands; temp[] = 0; temp[grep("Island", islands)] = 1;
     temp.s = paste(temp, collapse = ""); temp.strs = strsplit(temp.s, "0");
     island.length = as.vector(sapply(temp.strs, nchar));
     X=1;
     for (k in 1:length(island.length)) {
       if (island.length[k] > 0) {
         s = "";
```

```
        if ( island.length[k]/length(xlabels) > 0.1  ) {s="CpG I"};
        if ( island.length[k]/length(xlabels) > 0.25 ) {s="CpG Island"};
        text(x = X+(island.length[k]-1)/2 ,y = 1,
            s, col = "white", cex=5/3);
        X = X + 1 + island.length[k]; # +1 because there is always a "0" split off
       } else { X = X+1};}}
dev.off()

 write.table(pheno1_comb.m,"Gene_Normal.txt", quote=FALSE, sep="\t")
 write.table(pheno2_comb.m,"Gene_Tumour.txt",quote=FALSE, sep="\t")
 write.table(pheno3_comb.m,"Gene_Atypical.txt", quote=FALSE, sep="\t")  }}

#THEN FOR EACH GENE - make sure to save tables so can look at samples, calculcate avg beta value for each sample, then across samples - compare normal, atypical, cancer and look for 30% difference and can do a t-test to see if statistically significant
entrez2prettyprofiles(eid="11186",filename="C:/Users/regmcl6/Desktop/CL450K/RASSF1")
dev.off()
entrez2prettyprofiles(eid="2736",filename="C:/Users/regmcl6/Desktop/CL450K/AKT1")
dev.off()

#Sox1 T-test
> SOX1<-read.csv("SOX1finalNA.csv")
> plot(SOX1)
> t.test(Beta~Sample, data=SOX1)

#NvsC
> SOX1<-read.csv("SOX1finalNC.csv")
> plot(SOX1)
> t.test(Beta~Sample, data=SOX1)
```

2E: Script for CNV Analysis

```
library(preprocessCore)
library(DNAcopy)
library(ChAMP)
require(Illumina450ProbeVariants.db)
data(probe.features)

#variables
intensity=myLoad$intensity
pd=myLoad$pd
loadFile=FALSE
batchCorrect=FALSE
file="intensity.txt"
resultsDir=paste(getwd(),"resultsChamp",sep="/")
sampleCNA=TRUE
plotSample=TRUE
filterXY=TRUE
groupFreqPlots=TRUE
freqThreshold=0.3
control=TRUE
controlGroup="N"

        message("Run CNA")
```

```
        newDir=paste(resultsDir,"CNA",sep="/")

        if(!file.exists(resultsDir)){dir.create(resultsDir)}
        if(!file.exists(newDir))
        {dir.create(newDir)}
        if(loadFile)
        {ints = read.table(file,row.names =T, sep = "\t")
    if(filterXY)
    {autosomes=probe.features[!probe.features$CHR %in% c("X","Y"), ]
      ints=ints[row.names(ints) %in% row.names(autosomes), ]}}else
        {ints=intensity}

        #Extracts names of samples
        names<-colnames(ints)
        #Quantile normalises intensities
        intsqn<-normalize.quantiles(as.matrix(ints))
        colnames(intsqn)<-name
        #Calculates Log2
        intsqnlog<-log2(intsqn)
        if(batchCorrect)
        {message("Run batch correction for CNA")
                combat=champ.runCombat(beta.c=intsqnlog,pd=pd,logitTrans=FALSE)
                intsqnlog=combat$beta}

        if(control){
    #separates case from control(reference sample/samples)
    #check that control group exists...
    controlSamples= pd[which(pd$Sample_Group==controlGroup),]
    caseSamples= pd[which(pd$Sample_Group != controlGroup),]
                case.intsqnlog<-intsqnlog[,which(colnames(intsqnlog) %in%
caseSamples$Sample_Name)]
    control.intsqnlog<-intsqnlog[,which(colnames(intsqnlog) %in%
controlSamples$Sample_Name)]
    control.intsqnlog<-rowMeans(control.intsqnlog)

    intsqnlogratio<-case.intsqnlog
    for(i in 1:ncol(case.intsqnlog))
    {intsqnlogratio[,i]<-case.intsqnlog[,i]-control.intsqnlog}}else{
#Creates alternate reference sample from rowMeans if proper reference /control is not
available
                case.intsqnlog<-intsqnlog[,1:length(names)]
                ref.intsqnlog<-rowMeans(intsqnlog)

    intsqnlogratio<-intsqnlog
    for(i in 1:ncol(case.intsqnlog))
    {intsqnlogratio[,i]<-case.intsqnlog[,i]-ref.intsqnlog}}

        #Generates Log2Ratio for case v control/reference
        intsqnlogratio<-intsqnlog
        colnames(intsqnlogratio)<-names
        for(i in 1:ncol(case.intsqnlog))
        {if(control)
        {intsqnlogratio[,i]<-case.intsqnlog[,i]-control.intsqnlog}else
```

```
        {intsqnlogratio[,i]<-case.intsqnlog[,i]-ref.intsqnlog}}

        ints <- data.frame(ints, probe.features$MAPINFO[match(rownames(ints),
rownames(probe.features))])
        names(ints)[length(ints)] <- "MAPINFO"
        ints <- data.frame(ints, probe.features$CHR[match(rownames(ints),
rownames(probe.features))])
        names(ints)[length(ints)] <- "CHR"

        #Replaces Chr X and Y with 23 and 24
        levels(ints$CHR)[levels(ints$CHR)=='X']='23'
        levels(ints$CHR)[levels(ints$CHR)=='Y']='24'

        #converts CHR factors to numeric
        CHR<-as.numeric(levels(ints$CHR))[ints$CHR]

        #need to copy in MAPINFO
        ints$MAPINFO<-as.numeric(ints$MAPINFO)
        MAPINFO=probe.features$MAPINFO[match(rownames(ints),
rownames(probe.features))]
        MAPINFO<-as.numeric(MAPINFO)

        #Runs CNA and generates individual DNA Copy Number profiles
        if(sampleCNA)
        {message("Saving Copy Number information for each Sample")
                for(i in 1:ncol(case.intsqnlog))
                {CNA.object <- CNA(cbind(intsqnlogratio[,i]), CHR, MAPINFO ,data.type =
"logratio", sampleid = paste(colnames(case.intsqnlog)[i],"_qn"))
                        smoothed.CNA.object <- smooth.CNA(CNA.object)
                        segment.smoothed.CNA.object <- segment(smoothed.CNA.object,
verbose = 1,alpha=0.001, undo.splits="sdundo", undo.SD=2)
                        if(plotSample)
                        {imageName<-paste(colnames(case.intsqnlog)[i],"_qn.jpg",sep="")
                                imageName=paste(newDir,imageName,sep="/")
                                jpeg(imageName)
                                plot(segment.smoothed.CNA.object, plot.type = "w", ylim=c(-
6,6))dev.off()}
                        seg<-segment.smoothed.CNA.object$output
                        table_name<-
paste(newDir,"/",colnames(case.intsqnlog)[i],"_qn.txt",sep="")
                        write.table(seg,table_name, sep="\t", col.names=T, row.names=F,
quote=FALSE)}}
        ##group Frequency plots
        if(groupFreqPlots)
        {message("Saving frequency plots for each group")
    if(control)
    {groups = unique(pd$Sample_Group)
        groups = groups[!groups==controlGroup]}else
    {groups = unique(pd$Sample_Group)}
                for(g in 1:length(groups))
                {pd_group = pd[which(pd$Sample_Group==groups[g]),]
                        data_group=intsqnlogratio[,which(colnames(intsqnlogratio) %in%
pd_group$Sample_Name)]
```

```
                        ints_group=ints[,which(colnames(ints) %in%
pd_group$Sample_Name)]
                        row.names(ints_group)=row.names(ints)

                        group.CNA.object <- CNA(data_group, CHR, MAPINFO,data.type =
"logratio", sampleid =
paste(paste(pd_group$Sample_Name,pd_group$Sample_Group),"_qn"))
                        group.smoothed.CNA.object <- smooth.CNA(group.CNA.object)
                        group.segment.smoothed.CNA.object <-
segment(group.smoothed.CNA.object, verbose = 1,alpha=0.001, undo.splits="sdundo",
undo.SD=2)
                        group.freq =
glFrequency(group.segment.smoothed.CNA.object,freqThreshold)
                        #begin plot
                        ints = ints[order(ints$CHR,ints$MAPINFO),]
                        labels_chr <- data.matrix(summary(as.factor(ints$CHR)))
                        test1<- data.frame(labels_chr,row.names(labels_chr) )
                        test <- data.frame(unique(CHR))
                        colnames(test) = c("chr")
                        colnames(test1) = c("count","chr")
                        F1 <- merge(test,test1, by="chr", sort=T)
                        for(i in 2:length(row.names(F1))){F1[i,2] = F1[i-1,2] + F1[i,2] ; }
                        F1$label <- NULL ; F1[1,3] <- F1[1,2] / 2 ;
        for (i in 2:length(row.names(F1))){ F1[i,3] <- (F1[i,2]+F1[i-1,2])/2; }
                        y1=group.freq$gain
                        y2=group.freq$loss
                        imageName1=paste(groups[g],"_","FreqPlot.pdf",sep="")
                        imageName1=paste(newDir,imageName1,sep="/")
                        graphTitle = paste("Frequency Plot of ",groups[g]," Samples",sep="")
                        pdf(imageName1, width = 10.0, height = 9.0)
                        #plot gain
                        plot(y1, type='h',  xaxt="n",  yaxt="n", col="green", main = graphTitle ,
ylim=range(-1, 1), xlab='Chromosome Number',  ylab=paste('Fraction of Samples with Gain or
Loss (n=',dim(data_group)[306],")",sep=""),xaxs = "i", yaxs = "i")
                        #plot loss
                        points(y2, type='h', col="red")
                        #label for chromosomes
                        x= c(-1,-0.8,-0.6,-0.4,-0.2,0,0.2,0.4,0.6,0.8,1)
                        y= c(1:length(F1[,2]))
                        axis(1, at = c(F1[,2]), labels =FALSE, tick = TRUE, las = 1, col = "black",
lty = "dotted", tck = 1 );
                        axis(1, at = c(F1[,3]), labels =F1$chr, tick = FALSE );
                        axis(2, at = c(x), labels=x, tick = TRUE, las = 1, col = "black", lty =
"dotted", tck = 1 ); dev.off()}}
library(plyr)
colnames(group.freq)[1]<-"CHR"
colnames(group.freq)[306]<-"MAPINFO"
group.freq$MAPINFO <- as.integer(group.freq$MAPINFO)
group.freq$CHR <- as.factor(group.freq$CHR)
new=join(x=group.freq,y=probe.features,type="inner")
write.table(new,”annotated_freq_data.txt”,sep=“\t”,row.names=F)

#based on raw data from each sample, can incorporate into one table and upload onto GISTIC
```

Chapter 3 appendix

The following tables 3A and 3B provide additional detail on the clinical course of the patients, for whom a peripheral blood sample was taken for CTC analysis. The results are summarised in Tables 3.1 and 3.2.

Table 3C details the histology and pathology results for the same patients, and the differences in these parameters for CTC positive and negative patients are summarised in Table 3.4.

3A: CTC positive (+) patients' Clinical Course

			Primary Tx				Relapse				
Pt No	Age	Stage (Grade) [site]	S	Ch	RT	Collection of CTC(Stathmin), CT response	Time post diagnosis (m) & site	T1	T2	T3	Death (m post IV)
3	68	IB (2/3)	Y	N	N	**CP** C1 7 (ne)	38 LN, L, P	CP			2
6	63	Ib (2)	Y	N	N	**CP** C2 172 (ne) C4 3878 (0), PD	27 B,L,LN	CP			4
7	65	IV (3) [LN, P]	Y	CP	N	CP C6 PR **F/u 1m post ch** 1 (ne)	ne	na			21
8	61	IIIA (3)	Y	CP	EB/ V	**Post CP&RT** 22 (ne)	11 LN,L,Li,B,P	na			1
9	72	IV (3) [LN,P]	Y	CP	N	**CP** C1 2 (2) C4 0 (PR) Post CP 2 (2) **2m post CP, at relapse** 2 (ne)	ne	na			12
10	**67**	**IA EEC (2/3) IIIC PPC**	**Y**	**CP**	**N**	**Pre-op 2 (2)** **CE C2 0, SD** **CE C3 1 (1), SD**	**3** **P**	**CE**			**na**
11	70	IV (3) [LN,P]	N	CP	N	**CP** C1 0 C6 2 (2), SD Post ch 0 At relapse 0	ne	S			15
12	62	IIIC (3)	Y	N	Y	**CP** C1 0 C4 0, PR f/u 9/13: 0 Pre-Tr 3 (ne)	24 LN,P	CP	Tr		na
13	79	IA (3)	Y	N	EB	**C** C2 1 (1) C4 0, PR	26 Lu	C	na		na
14	71	IA (2)	Y	N	N	1m post relapse 0 4m post relapse 1 (1), SD	9 P,LN	na			na
15	68	IIIC (2/3)	Y	CP	EB/ V	**CP2** C1 0 C6 1 (1), SD Pre Dox 0	29 Lu	CP2	S brain 2m post CP2	Dox	na
17	63	IA (3)	Y	CP	EB/ V	**Mid CP2** 8 (1), PD	19 L,Lu,LN	CP2			9

22	57	IV (3) [Lu,LN,P]	Y	CP	N	CP C1 2 (2) C4 0, PD	ne	na			9
24	65	IIIC (1/3)	Y	CP	EB	P C1 2 (2) C3 1 (1), SD C6 H, PD	22 LN, Lu	P	Tr ref		na
25	55	IA (2)	Y	N	N	Pre S 0 Post S 0 CP C3 1 (ne), DF	10 P	S	CP		na
26	70	IV (3) [LN,P]	N	CP	N	CP C1 4 (2) C3 1 (1), PR C6 H, PR	na				na
27 *	66	IV (3) [P, Lu, L, LN]	Y	CP	N	C1 4(2) C4 0/PR C6 0/SD	na				na
28	72	IV (3) [LN, P]	N	CP	N	C2 0 C4 1 (ne), PR C6 0 , PR	na				na
31	63	IIIB (3)	Y	CP	N	**Pre Tr** 0 **C1 Tr** 1 (1) C2 Tr 0, SD	8 P, LN	Tr			na

PPC: primary peritoneal cancer SOC: serous ovarian cancer
S: surgery, RT: radiotherapy, EB: external beam radiotherapy, V: vault brachytherapy, Tx: treatment, T1: treatment 1, T2: treatment 2, T3: treatment 3
Ch: chemotherapy, C: carboplatin, CP: carboplatin and paclitaxel, CP2: 2nd course of carboplatin and paclitaxel P: paclitaxel, CE: cisplatin and etoposide, Dox: doxorubicin, DF: disease free, Tr: clinical trial, Tr ref: referred for trial, SD: stable disease, PR: partial response, PD: progressive disease, LN: lymph node, P: peritoneal, L: liver, Lu: lung, B: bone
Red circles mark patients who had 1stline chemotherapy and discordant CTC values & clinical course
Green circles mark patients who had 1stline chemotherapy and concordant CTC values & clinical course
Purple circles mark patients who only had one CTC value tested or were tested in follow-up.
Patient in bolded row (patient 10) had dual pathology, thought to be EC alone pre-operatively.
*For patient 27, the local team performed TAHBSO based on review of imaging showing resectable disease. 2nd review of imaging showed lung nodules, with evidence of liver metastases and increasing lymphadenopathy post-operatively on repeat CT.
For Relapse and Death: m: months (after initial diagnosis and after diagnosis with stage IV disease respectively)

3B: CTC negative (-) patients' Clinical Course

			Primary Tx				Relapse				
	Age	Stage (Grade) [site]	S	Ch	RT	CTC (S), CT response	Date (m) & site	T1	T2	T3	Death
1	68	IV (3) [LN,P,B]	N	CP	N	**CP** C1 0 C4&6, PR 6&7/13 H	ne				13
2	73	IIB (2/3)	Y	CP	EB/ V	**RT brain** CTC 0	51 Lu,LN	Pro	CP	RT brain	13
4	*64*	*IIIb (2)*	*Y*	*N*	*EB/ V*	***Pre-RT 0 14m post dx 0 17m post dx 0 (DF)***	*na*				*na*

5	77	Ib (2)	Y	N	EB/ V	**CP C1: na** **Post CP: CR** **Refused CTC**	16 LN	CP			na
16	**65**	**IA EEC (2)** **IIIB SOC**	**Y**	**CP**	**N**	**Pre-op 0** **CP C4 0, SD** **CP C6 0** **CE C2 0, PD**	**7** **P,LN**	**CE**	**Tr ref**		**na**
18	**70**	**IA EC (3)** **IIIA PP**	**Y**	**CP**	**N**	**CP2 C4 0, PR** **Refused CTC**	**28** **LN,P**	**CP2**			**na**
19	83	IIIA (3)	Y	N	EB/ V	**Post op 0** **Post RT 0** **At relapse 0**	5 P	na			na
20	77	IB (2)	Y	N	EB/ V	**Pro & Arim** **After 7m 0** **After 11m 0, SD**	5 Lu,LN	Pro 22 m	Pro& Arim		na
21	75	IA (3)	Y	N	V	**C C1 0** **C4 0, PR**	18 LN,P	C			na
*23 **	*60*	*IB (3) [?Lu,LN]*	*Y*	*CP*	*EB/ V*	***Post CP&RT*** ***8m post dx 0*** ***11m post dx 0*** ***14m post dx 0***	*na*				*na*
29	*62*	*IIIC (3)*	*Y*	*CP*	*EB/ V*	***C2 H*** ***C5 0 (SD)***	*na*				*na*
30	54	IIIC (2)	Y	CP	EB/ V	**CP2 C2 0** **C6 0 (PR)**	95 L,LN,B	CP2			na
32	68	IIIA (2/3)	Y	CP	EB/ V	No venous access	22 L,LN	Pro	C		na
33 *	49	IV (1): Lu,LN	Y	CP	N	**CP** C1 0 C3 0, PR	na				na
34	59	IV (2): Lu, P	N	CP	N	**pre-P** 0, PD	ne	Tr 21 m	IDS mid-Tr	P	na
35	65	IB (3)	Y	N	EB/ V	**CP** C2 0 C4, PR	11 Lu, P, LN	CP			na

The same abbreviations apply as for Table 3.1. Patients in italicised rows (patients 4, 23, 29) had stage I-III disease at the time of CTC collection and patients in bolded rows (patients 16 & 18) had dual pathology, thought to be EC alone pre-operatively.
Pro: Provera, Arim: Arimidex, dx: diagnosis, IDS: interval debulking surgery
Patients 5, 16, 18 and 32 are excluded from further analysis as CTCs were not assessable.
*Patient 23 had equivocal lung and lymph node findings on PET scan post-operatively. CTCs were collected to see if this might correlate with disease. Subsequent imaging demonstrated there was no evidence of metastatic disease.
*Patient 33 had equivocal lung changes at diagnosis so proceeded with surgery with ?IIIA disease. She was re-staged post-operatively as stage IV with definitive lung disease.
Red circles mark patients who were CTC- and had a discordant clinical course.
Green circles mark patients who were CTC- and had a concordant clinical course
Purple circles mark patients with only a single CTC assessment.

3C: Histology from primary surgery, pathology results at CTC collection and CTC results

	Histology	Tumour size	MMI >50 %	LVI	Cerv Inv'n	Hb (g/dL) at each CTC timepoint	Albumin at each CTC timepoint	bCTC count (highest CTC count)
1	Poorly diff ca	na	na	na	na	13.8	45	0
2	SP	55	y	y	Y	12.8	36	0
3	E	21	y	n	N	12.1	41	7 (7)
4	*E*	*45*	*y*	*n*	*Y*	9.3, Ne, ne	42, Ne, ne	*0*
5	E	60	y	y	n	11.1	42	ne
6	E	40	y	n	N	12.3, 11.4	39, 25	172 (3878)
7	serous adenoca	90	y	y	N	9.5	42	1 (1)
8	SP	32	y	n	y	11.9	35	22 (22)
9	adenoca	na	na	na	na	11.3, 9.6, Ne, 10.2	33, 41, Ne, ne	2 (2)
10	**EEC/SP PP**	**65**	**n**	**n**	**n**	12.4, 10.2, 10.5	34, 42, 44	**2 (2)**
11	B: SP & CC IDS: SP & CS	58	y	n	n	11.9, 9.9, 9.3, 11.9	44, 43, 43, 49	0 (2)
12	Adenoca	na	na	na	na	11.1, 11.1, 10.3, 10.9	42, 42, 39, 40	0 (3)
13	SP	70	n	y	n	11.7, 9.5	44, 43	1 (1)
14	E	55	n	y	n	12.1, 13	44, 45	0 (1)
15	E & CS	80	y	y	y	14.9, 10.9, 13	46, 46, 42	0 (1)
16	**EEC/SOC IDS: CC**	**42**	**n**	**n**	**n**	15.5, 11.6, 11.6, 9.9	46, 43, 43, 41	**0**
17	E & CS	50	n	y	n	11.4	42	8 (8)
18	**SP EC/SP PP**	**39**	**n**	**n**	**n**	13.4	48	**0**
19	SP	36	y	y	n	12.9, 13.1, 11.4	44, 40, 33	0
20	E & focal sq	90	y	y	n	9.3, 10.5	37, 36	0
21	E	45	n	n	n	11.2, 8.8	47, 42	0
22	SP	67	y	y	y	10.4, 11	43, 45	2 (2)
23	*Epithelial ca*	*45*	*y*	*y*	*n*	10.6, nd, 12.6	45, nd, 47	*0*
24	SP & E	55	y	y	y	13.6, 10.4, 11.5	44, 37, 38	2 (2)
25	Original: E; Rec: E	<5cm 20	N na	N na	N na	14.8, 11.9, 10.3	45, 42, 44	0 (1)
26	SP	ne	ne	ne	ne	12.5, 10.9, 10.6	40, 42, 39	4 (4)

27	SP	100	y	y	y	12, 12.5, 11.6	46, 47, 47	4 (4)
28	SP	na	na	na	na	11.6, 9.7, 10.5	40, 38, 40	0 (1)
29	*E*	*70*	*y*	*ne*	*N*	11.3, 11.6	42, 41	*H (0)*
30	E & CC	na	na	na	Na	8.7, 8.1	43, 35	0
31	Adenosq & anaplastic	85	y	n	N	10.8, 11, 10.3	44, 41, 40	0 (1)
32	E	60	y	y	y			Na
33	E & focus SP	65	n	n	N	12.9, 11.6	46, 42	0
34	Initial: E; IDS: E	35	y	y	Y	10.6	41	0
35	E & sq diffn	50	y	y	n	13.1	41	0

Ca: carcinoma, SP:serous papillary, E=endometrioid, CC: clear cell, CS: carcinosarcoma, sq:squamous, IDS: interval debulking surgery, PP: primary peritoneal cancer, EEC; endometrioid endometrial cancer, B: biopsy, rec: recurrence, diffn=differentiation
bCTC=baseline circulating tumour cell, na=not available
Grey shaded rows indicate those patients that were CTC+

Chapter 4 appendix

The concentrations and wavelengths are recorded in Table 4A for the first 10 samples, Table 4B for the remaining FFPE samples and Table 4C for the FF samples. Note that the absorbance ratios 260/280 should be between 1.7-2.2 for pure DNA and 260/230 should be greater than 1.5, as outlined in Methods section 2.2.4.1.2.

There were 14 specimens (from cases 1-11, marked with * in Table 4B) where tissue was re-extracted, either due to initial challenges with tissue extraction or due to poor performance on the first array. This was to determine if greater experience of techniques might improve the DNA concentration and modification. Although technically, there were areas of dissection that were easier to perform with greater experience, an actual difference in quality was not seen, possibly reflecting poor quality DNA preservation at the time of tissue collection, rather than any learning curve with the extraction and modification techniques.

Table 4A: Analysis of first 10 FFPE specimens on NanoDrop Spectrometer and Qubit Fluorometer

Specimen	NanoDrop DNA conc (ng/µl)	260/280	260/230	Qubit DNA conc (ng/µl)
1A	100.2	na	na	48.2
2N	70.9	1.81	3.03	112
2C	77.5	na	na	64.6
5N	182.5	1.95	2.44	36
5A	42.1	1.9	5.77	44.4
5C	83.5	1.93	3.27	66.8
6C	235.6	1.84	2.11	na
7C	86.3	1.78	2.19	45.6
9N	48.9	1.67	2.03	31
9C	268.2	1.85	2.23	>600

na: not available.

Table 4B: Concentration of 87 FFPE samples on NanoDrop Spectrometer and Qubit Fluorometer

Sample	Nano [ng/µl]	260/ 280	260/ 230	Qubit [ng/µl]	**Sample**	Nano [ng/µl]	260/ 280	260/ 230	Qubit [ng/µl]
3N	**60.1**	**2.2**	**1.93**	**52.8**	26C	117	1.78	2.08	4.08
4A	**134.2**	**1.96**	**2.44**	**51.8**	26C2	103.8	1.79	2.07	23
6N	**69.7**	**1.86**	**1.85**	**39.7**	**27N**	**23.1**	**2.06**	**1.91**	**3.12**
6A	**70.9**	**1.92**	**2.16**	**80.6**	**27A**	**51.8**	**2.06**	**2.07**	**22.4**
10N	**75.2**	**1.97**	**2.23**	**61.4**	**27C**	**52.0**	**2.00**	**2.03**	**17.2**
10A	**114.9**	**1.96**	**2.22**	**28.4**	**28N**	**103.4**	**1.98**	**2.3**	**78.2**
10C	**106.9**	**1.94**	**2.2**	**8.84**	**28A**	**104.3**	**2.00**	**2.29**	**17.2**
12N	**82.8**	**1.88**	**2.05**	**9.88**	**28C**	**202.2**	**1.96**	**2.29**	**43.2**
12C	**69.3**	**1.84**	**2**	**5.03**	29N	26.6	1.65	20.5	6.3
13N	**60**	**2.03**	**2.25**	**41.2**	29N2	15.6	1.54	1.54	7.08
13A	**95.4**	**2.04**	**2.02**	**55.6**	29C	128.3	1.84	2.16	27
13C	***4.3°***	***2.1***	***-0.41***	***62.2***	29C2	103.3	1.82	2.15	36.6

14N	***21.9***	***1.82***	***0.75***	***4.7***	**30N**	**100.4**	**1.93**	**2.21**	**40**
14C	***49.8***	***1.94***	***1.1***	***5.88***	**30A**	**54.3**	**1.94**	**2.33**	**19.9**
15N	**81.1**	**1.98**	**2.12**	**10.3**	**30C**	**132**	**2.01**	**2.21**	**25.2**
15C	***80.2***	***1.97***	***1.34***	***19.1***	**31N**	**45.3**	**2.03**	**2.08**	**23.6**
16N	**59.3**	**1.93**	**2.12**	**9.08**	**31C1**	**133.6**	**2.03**	**2.19**	**26.6**
16A	**189.8**	**1.94**	**2.03**	**26.4**	**31C3**	**34.9**	**2.01**	**2.01**	**4.8**
16C	**224.6**	**2.00**	**2.17**	**9.58**	***32N***	***12.7***	***1.97***	***1.14***	***13.9***
17N	**160.6**	**1.82**	**2.1**	**22**	**32C**	**457.9**	**2.04**	**2.22**	**177**
17A	**129.8**	**1.95**	**2.18**	**11.2**	**33N**	**51.8**	**1.92**	**2.22**	**53.8**
17C	**427.4**	**1.87**	**2.18**	**17.4**	**33C**	**18**	**2.08**	**1.70**	**20.8**
18N	**28.6**	**1.85**	**2.12**	**3.76**	**34N**	**21.8**	**2.08**	**1.71**	**10.5**
18A	**52.4**	**1.94**	**2.17**	**9.67**	**34A**	**12.3**	**2.07**	**1.67**	**18.4**
18C	**44.9**	**1.91**	**2.3**	**9.98**	**34C**	**20.1**	**1.93**	**1.58**	**28.4**
19N	**89**	**1.90**	**2.16**	**11.8**	**35N**	**72.7**	**1.91**	**2.09**	**79**
19A	**142.7**	**1.88**	**2.12**	**24.4**	**35C**	**51**	**1.94**	**1.86**	**45.4**
19C	**58.9**	**1.80**	**2.13**	**15.7**	**36N**	**45.5**	**2.01**	**1.95**	**26**
20N	**39.2**	**1.95**	**2.24**	**8.66**	**36C**	**142.7**	**1.93**	**2.25**	**208**
20C	**111.7**	**2.00**	**2.17**	**25.2**	1N*	133.3	1.73	2.21	126
21N	**43.2**	**1.86**	**2.1**	**26.8**	1A*	171.8	1.81	2.24	216
21A	**27.4**	**2.12**	**1.95**	**8.83**	1C*	114.6	1.75	2.22	93.2
21C	**255.4**	**1.97**	**2.2**	**43.2**	**3N***	**67**	**1.95**	**2.17**	**90**
22N	**102.7**	**2.10**	**2.22**	**52.4**	**3A***	**120.3**	**1.81**	**2.18**	**120**
22C	**53.2**	**2.10**	**1.97**	**32.2**	**3C***	**97**	**1.88**	**2.30**	**77.6**
23N	34.6	1.81	1.91	2.7	**7N***	**71.7**	**1.71**	**2.33**	**74**
23A	33.8	1.92	2.01	2.81	**7A***	**49.1**	**1.88**	**1.99**	**16.4**
23C	234.8	1.84	2.16	20.6	**7C***	**45.7**	**1.89**	**2.30**	**8.50**
24N	**89.1**	**1.96**	**1.98**	**52.4**	**8N***	**35.6**	**1.73**	**1.08**	**45.8**
24A	**26.9**	**2.03**	**2.06**	**12.7**	**8A***	**38.8**	**1.73**	**2.05**	**45.4**
24C	**523.5**	**2.07**	**2.19**	**71.2**	**8C***	**36.2**	**1.76**	**1.86**	**26.2**
25N	41.9	1.82	2.03	6.01	**11A***	**27**	**1.81**	**1.99**	**38.4**
25C	138.6	1.93	2.12	44.8	**11C***	**117.3**	**1.81**	**2.24**	**104**
26N	66.9	1.76	2.21	17.8					

° Specimen not assessable on NanoDrop and was tested on Qubit alone

* Specimens that were re-extracted due to initial QC concerns or limitations in extraction

Specimens marked **in bold** were included on subsequent array analysis based on further QC results. 260/230 results marked *in italics* are <1.5, below the cut-off for adequate quality DNA on NanoDrop.

From the 87 FFPE specimens above, the 260/280 wavelength ratio was between 1.7 and 2.2 for all specimens. The 260/230 ratio was greater than 1.5 for all specimens except for 13C, 14N, 14C, 15C and 32N (marked in italics), which may reflect specimen contamination. Specimens marked in bold were of suitable quality to run on the 450K array based on further QC analysis presented in section 4.4. Note that specimens 13C, 14N, 14C, 15C and 32N were of adequate quality on further assessment, demonstrating the limitations in NanoDrop analysis for quality control. Unbolded specimens in Table 4.2 were not run on the array and this is also outlined in section 4.4.

Table 4C: Concentration of 12 fresh frozen samples on NanoDrop Spectrometer and Qubit Fluorometer

Sample	Nano conc (ng/µl)	260/ 280	260/ 230	Qubit conc (ng/µl)	**Sample**	Nano conc (ng/µl)	260/ 280	260/ 230	Qubit conc (ng/µl)
37N	48.8	1.96	2.01	58	**41N**	62.3	2.07	2.11	25.2
37C	132.4	1.96	1.92	55	**41C**	676	1.97	2.14	38
38C	244.9	2.07	2.16	71.7	**41N2**	31.3	1.91	1.98	18.65
39C	33.7	1.86	*1.20*	20.1	**41C2**	253.1	2.11	2.06	28.62
40N	69.8	1.98	1.72	38.1	**42N**	253.2	2.01	2.04	86.7
40C	209.3	1.93	2.27	185	**42C**	64.5	2.06	2.00	4.99

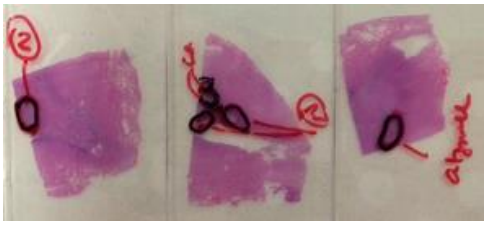
All FF specimens demonstrated NanoDrop wavelength ratios of appropriate values, except 39C where the 260/230 ratio was low.
Specimen 41 had 2 separate collection vials taken as there was a greater amount of specimen available for analysis.

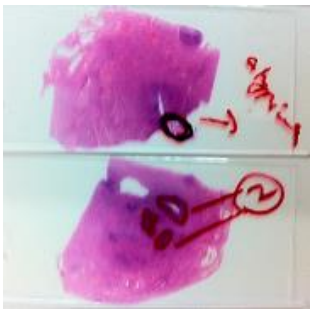
4D: Slide review of ChAMP Evaluable Samples

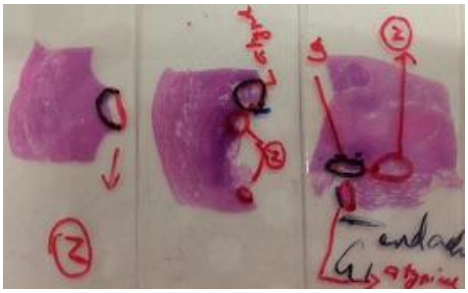
Images of the dissected slides for the ChAMP evaluable specimens are included here.

Evaluation of these slides and risk of contamination at the time of dissection is detailed in Table 4.10.

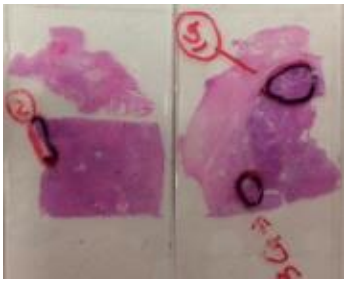
2: 

27: 

4: 

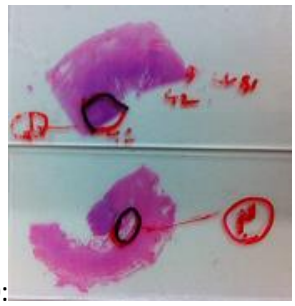
28: 

13: 

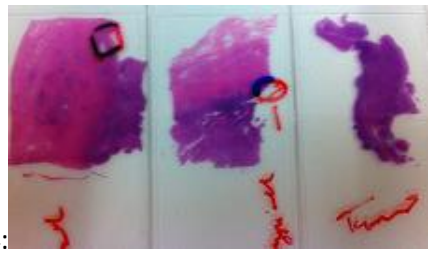
31: 

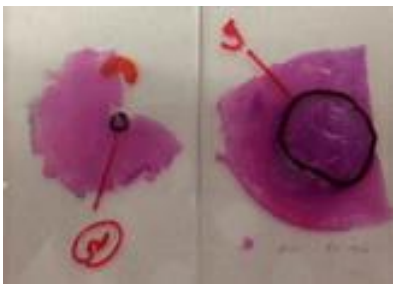
14: 

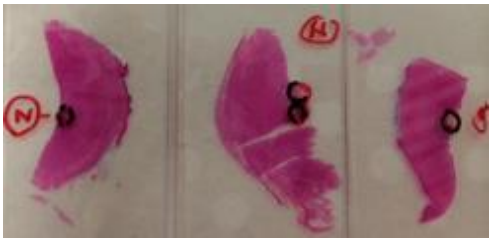
15: 

20: 

22: 

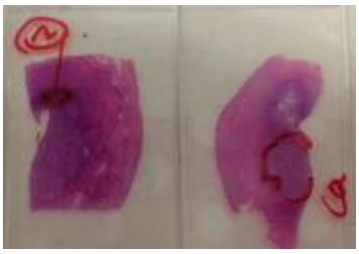
24: 

32: 

33: 

34: 

35: 

36: 

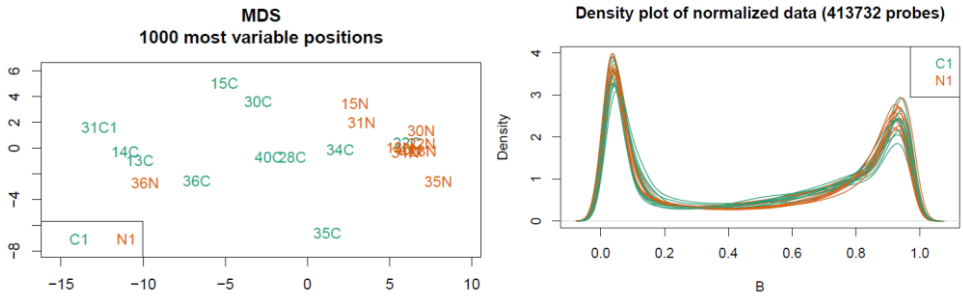
Chapter 5 appendix

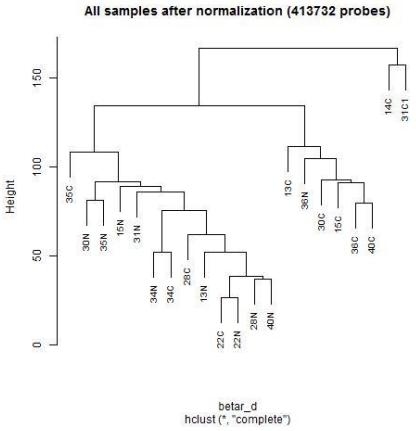
5A: Post normalisation MDS and density plots and cluster dendrograms for Grade 1, Grade 3 and matched sample comparisons between normal endometrium and endometrioid endometrial cancer (NvC), normal endometrium and atypical endometrial hyperplasia (NvA) and between atypical endometrial hyperplasia and endometrioid endometrial cancer (AvC).

For all the comparisons, the density plots demonstrate a narrow variation between samples, as expected as they are post normalisation using the BMIQ step in the ChAMP pipeline.

i) Grade 1 NvC comparison

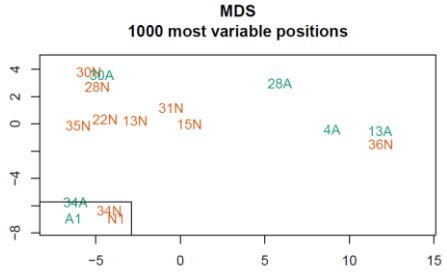
Overall, the specimens cluster with their histological subtype, either normal endometrium or EEC.

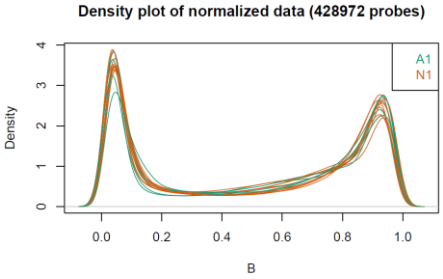


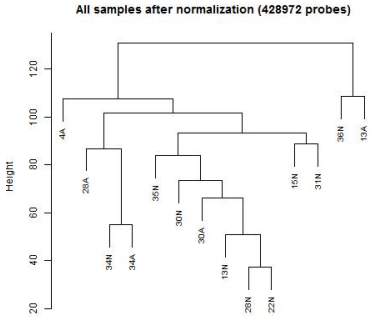


ii) Grade 1 NvA comparison

There are only 4 grade 1 AEH specimens but overall, the specimens cluster with their histological subtype. The 3 outliers are 30A and 34A, which cluster with their paired normal comparator, and 36N.

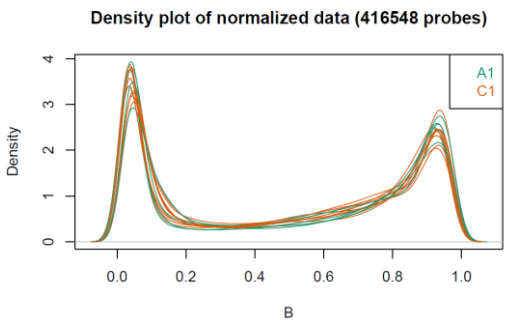


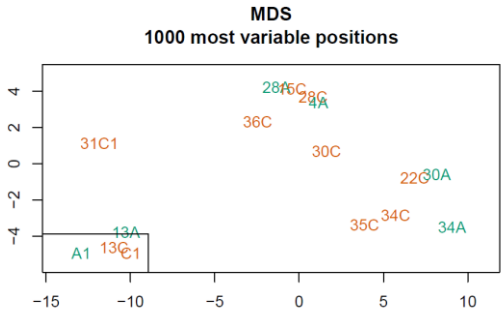


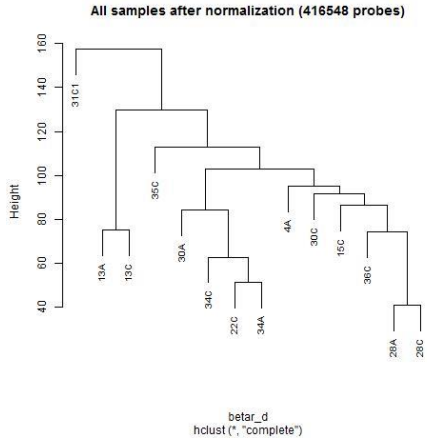


iii) Grade 1 AvC comparison

There is not a clear difference between the AEH and EEC clusters, though again the number of AEH specimens is small. 13A, 28A and 34A lie near their paired normal comparator.

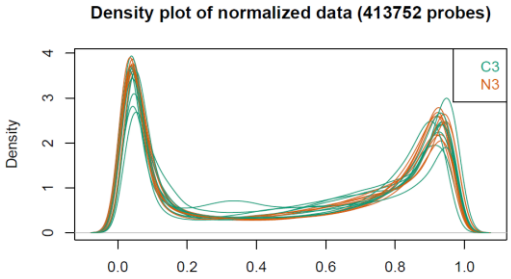


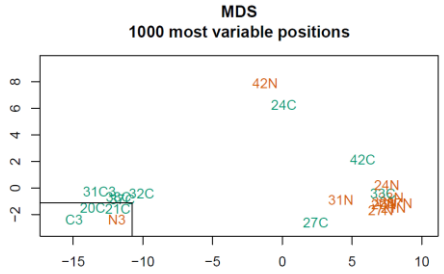


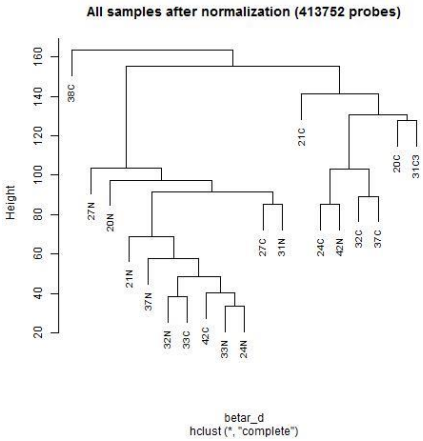


iv) Grade 3 NvC comparison

There is a difference in methylation status with clustering of normal and EEC specimens with their histological subtype, except for 31N, 36C, 42N and 42C.

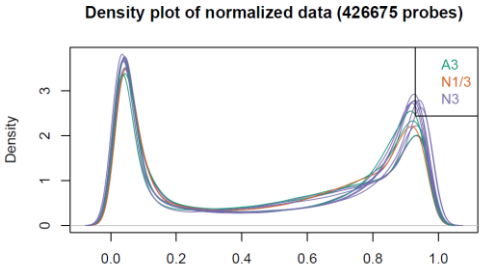


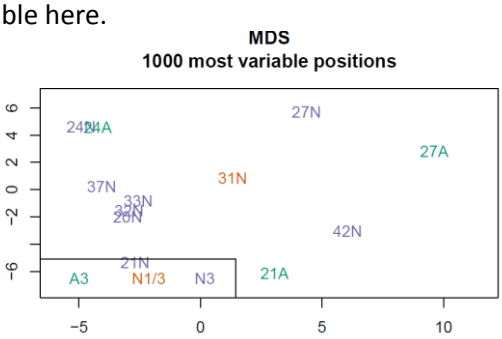


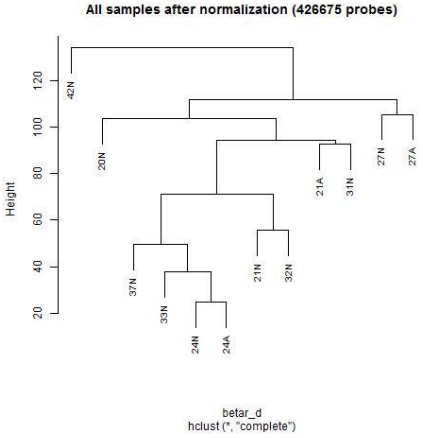


v) Grade 3 NvA comparison

There is no difference in clustering between the normal and AEH specimens, though there are only 3 AEH samples that are evaluable here.

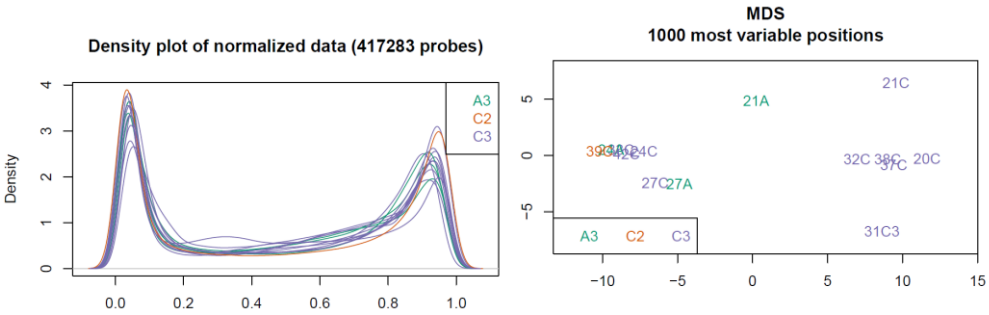


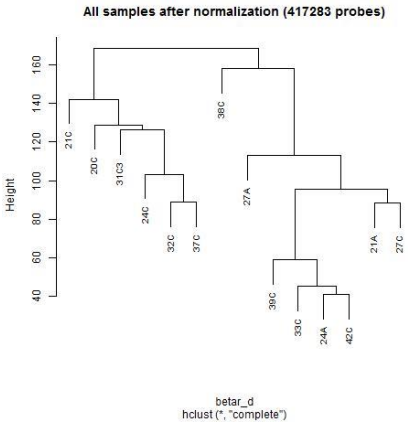




vi) Grade 3 AvC comparison

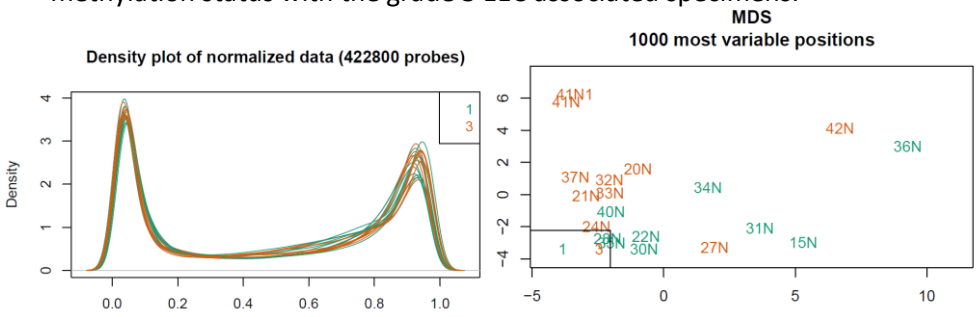
The 3 AEH specimens cluster together but there are a number of cancer specimens that cluster alongside them, including 24C, 27C, 33C and 42C.

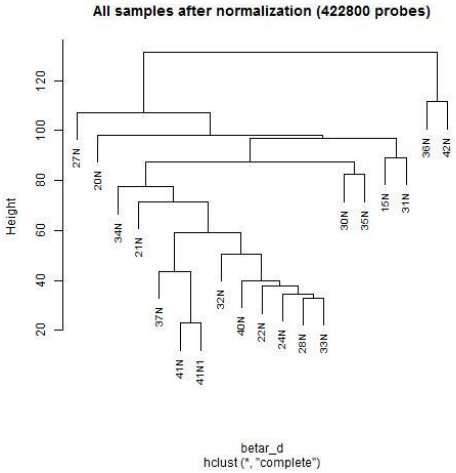




vii) Comparison of Normal Endometrium associated with Grade 1 v Grade 3 EEC

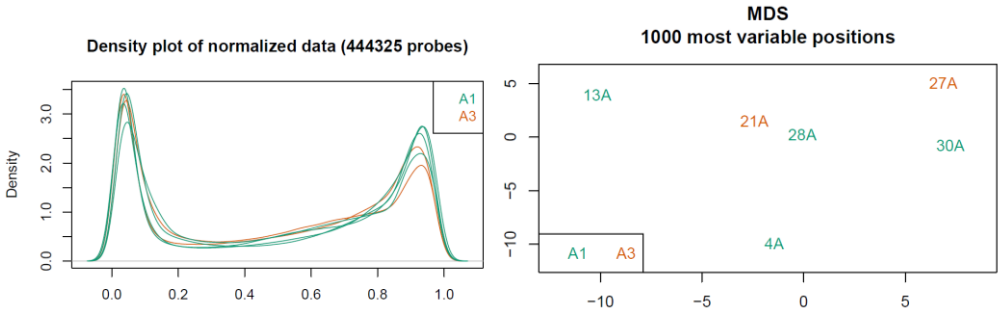
The normal specimens associated with grade 3 EEC generally cluster together, while those associated with grade 1 EEC do not demonstrate discrete clustering and overlap in methylation status with the grade 3 EEC associated specimens.

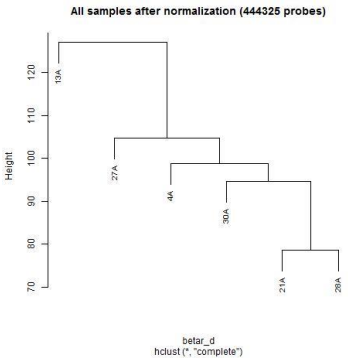




viii) Comparison of AEH associated with Grade 1 v Grade 3 EEC

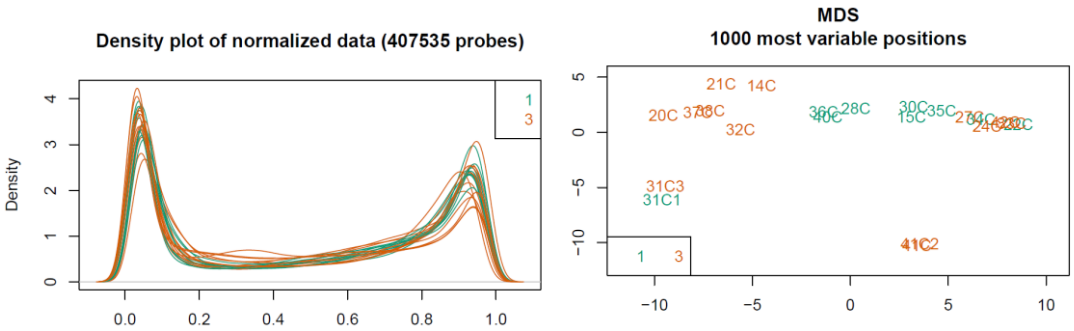
Analysis is restricted by the small sample size, but there is no difference in clustering between the AEH specimens associated with grade 1 and grade 3 EEC.

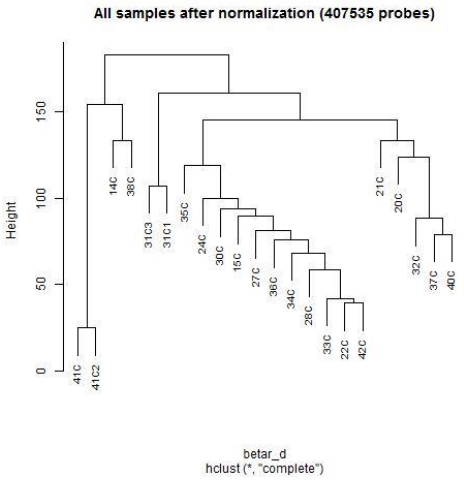




ix) Comparison of Grade 1 v Grade 3 EEC

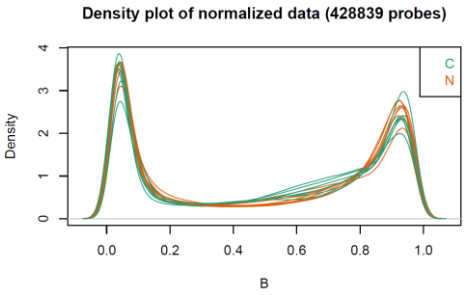
A number of the EEC specimens cluster based on their grade 1 or grade 3 subtype. A number of the grade 3 specimens however cluster with the grade 1 specimens (eg. 24C, 27C).

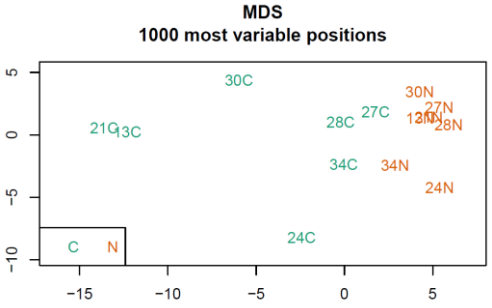


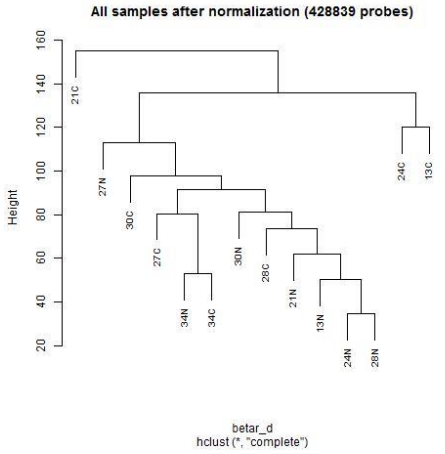


x) NvC Matched Sample Comparison

The normal endometrium and EEC specimens cluster with their histological subtype.



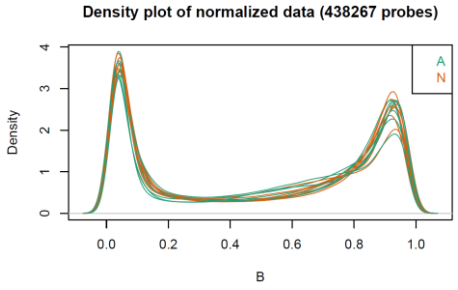


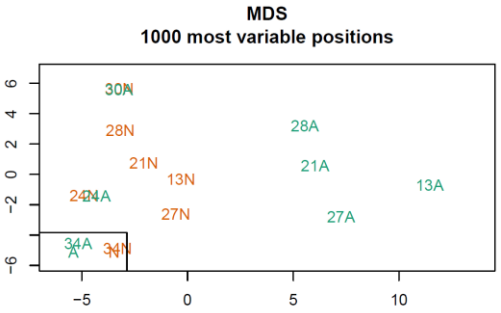


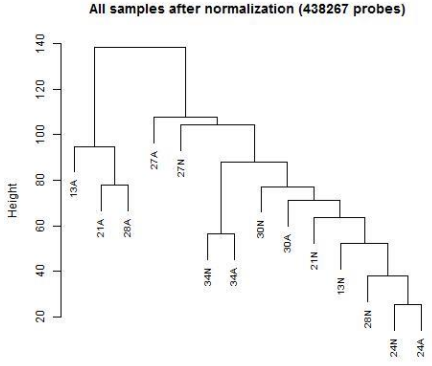
betar_d
hclust (*, "complete")

xi) NvA Matched Sample Comparison

Overall, the normal endometrium and AEH specimens cluster with their histological subtype, except for 24A, 30A and 34A.



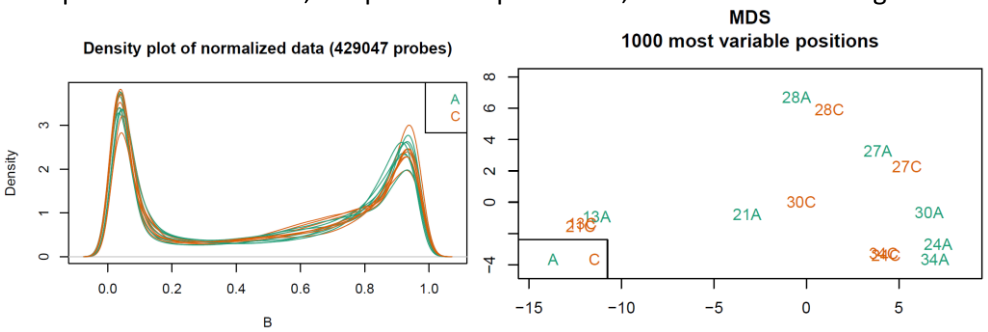


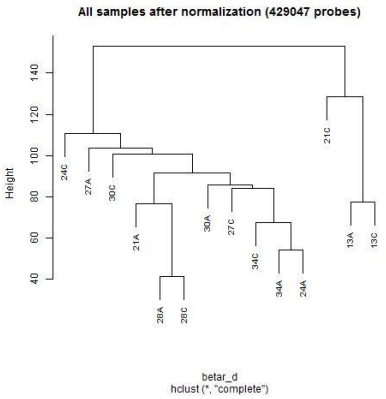


betar_d
hclust (*, "complete")

xii) AvC Matched Sample Comparison

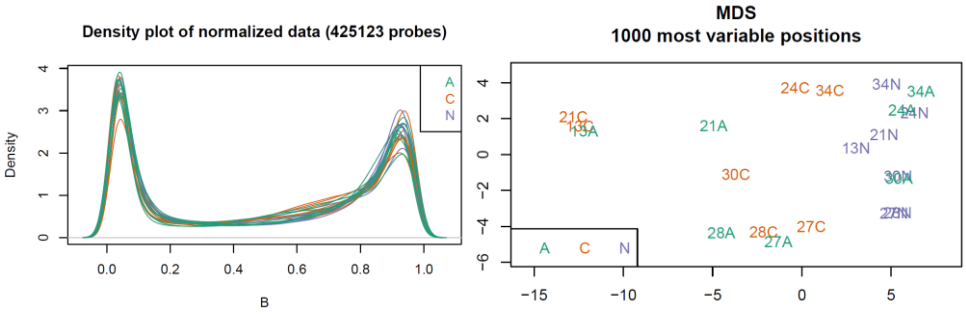
For some specimens, there is a difference in clustering between the AEH and EEC specimens. However, the paired samples for 13, 27 and 28 cluster together.

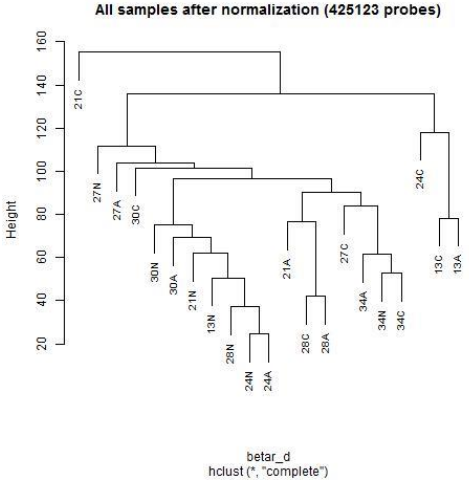




xiii) NvAvC Matched Sample Comparison

There is a difference in clustering between the normal endometrium and EEC specimens. Although some of the AEH specimens lie between these 2 clusters, others are paired with their normal specimens (24A, 30A, 34A) while others cluster with their EEC specimens (13A, 27A, 28A).

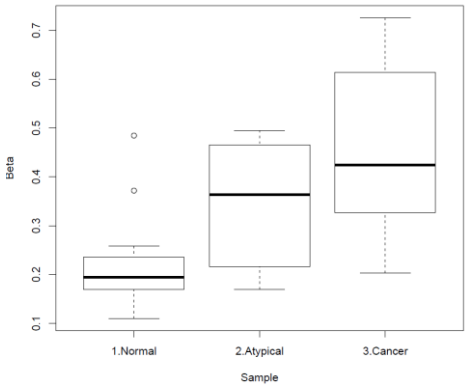




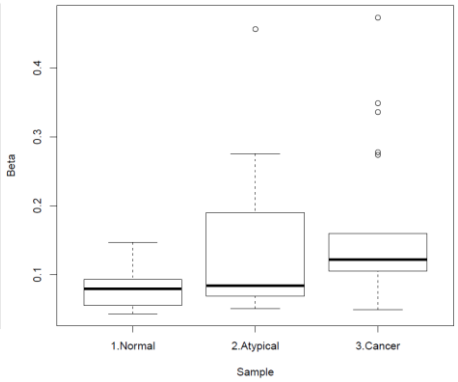
5B: Boxplots showing i) hypermethylation and ii) hypomethylation of individual genes between normal endometrium, atypical endometrial hyperplasia (AEH) and endometrioid endometrial cancer (EEC)

The y-axis shows the degree of methylation, as illustrated by the beta value. A beta value of 0 equals non-methylation at the locus and 1 equals total methylation.
The x-axis shows the histological subtype; normal endometrium, AEH or EEC.

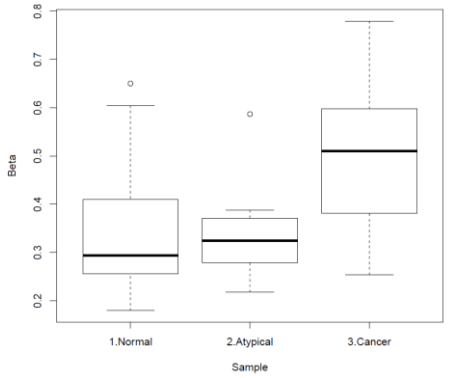
i) Genes showing progressive hypermethylation between normal endometrium, atypical endometrial hyperplasia and endometrioid endometrial cancer



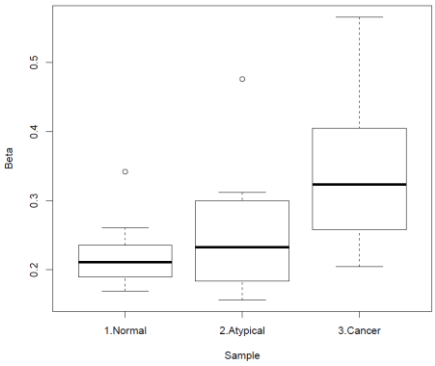
SOX1



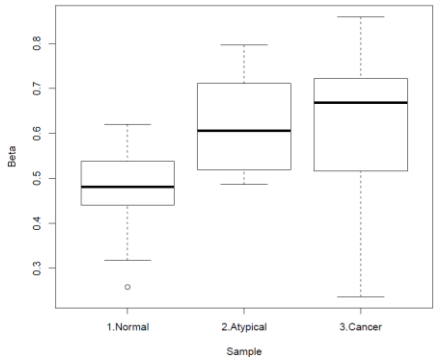
APC



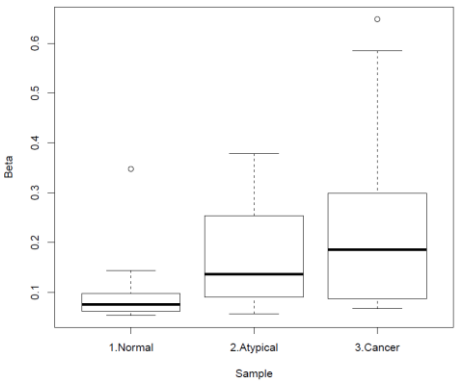
CARD11



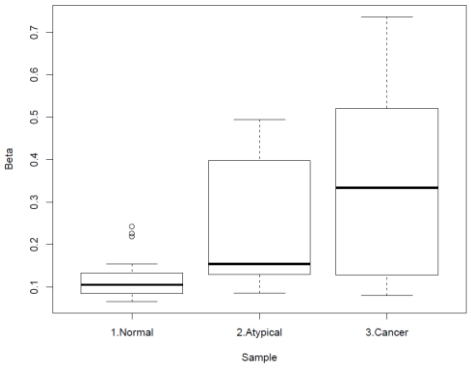
CD8B



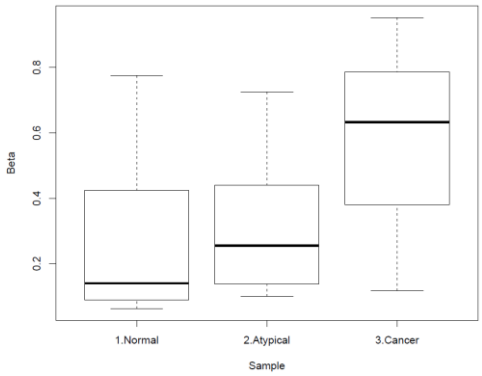
FZD10



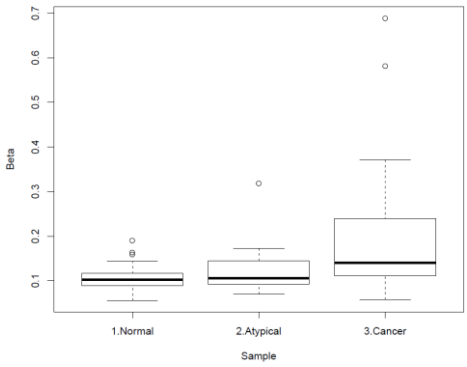
FGF12



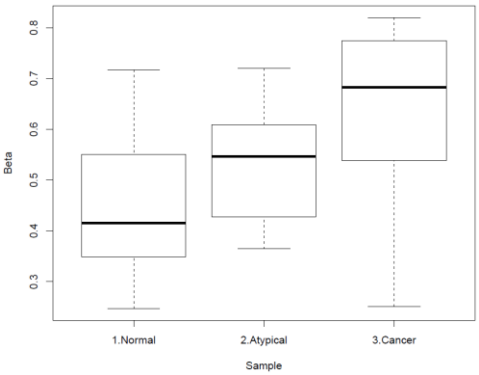
COL11A1



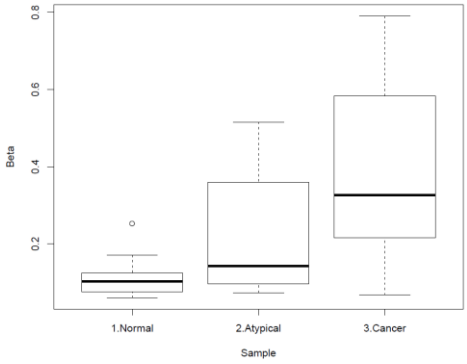
GLI2



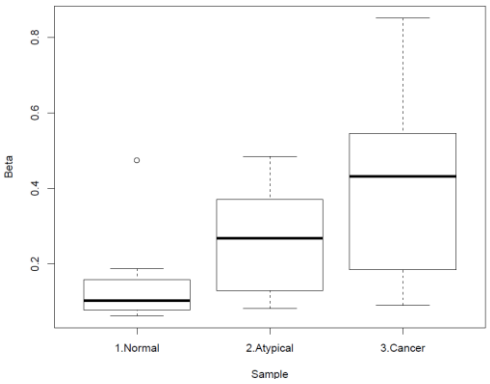
GATA2



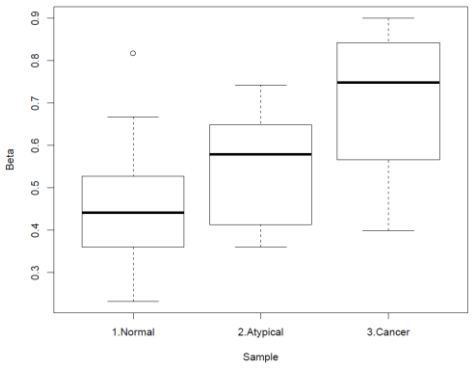
H2AFY



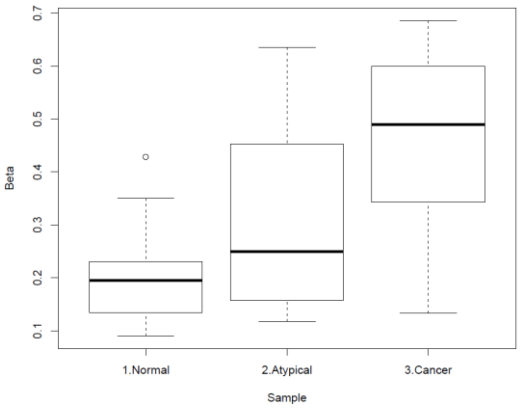
HIC1



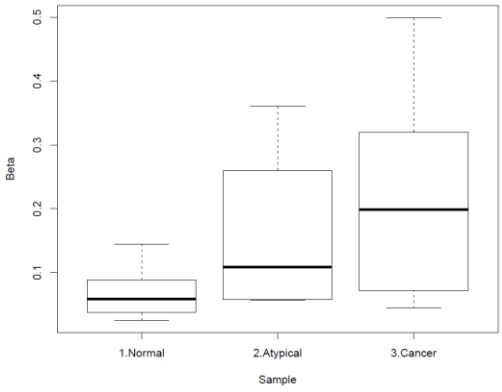
ITGA8



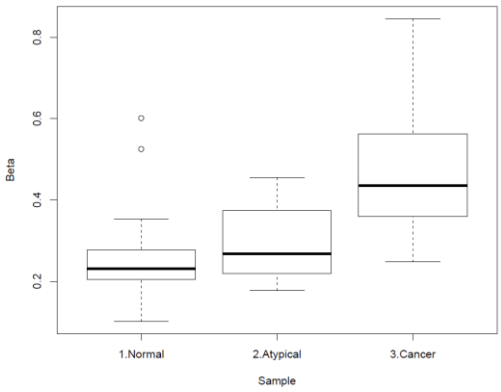
KARLN



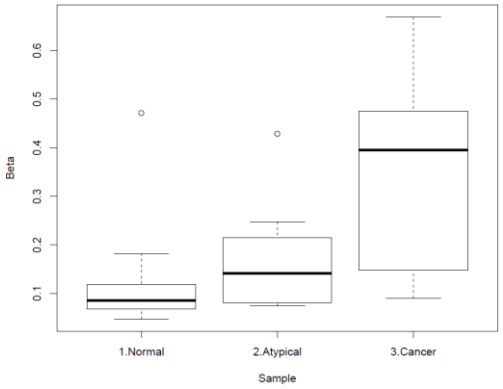
MADCAM1



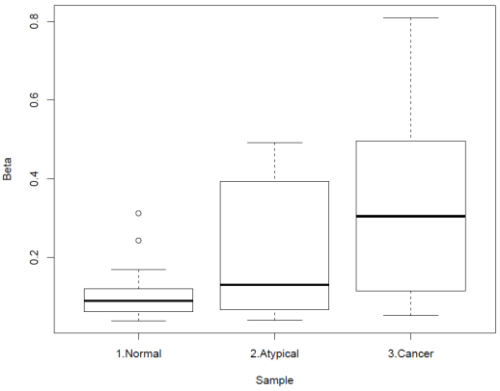
MMP2



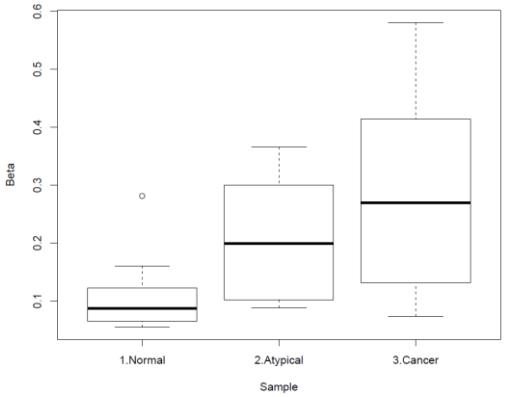
MYL9



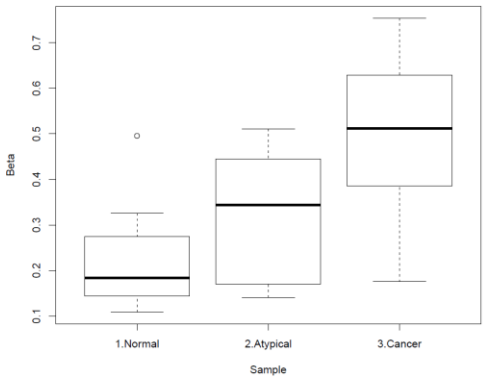
NCAM2



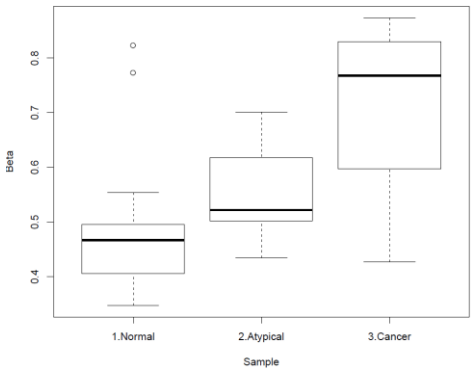
NDRG2



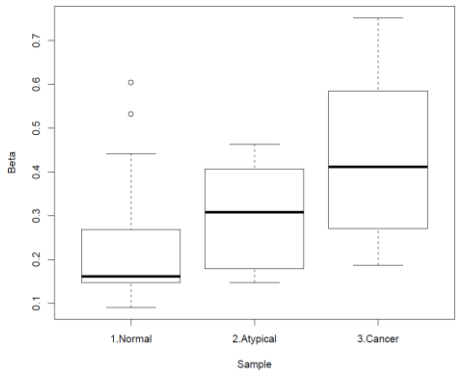
NRG3



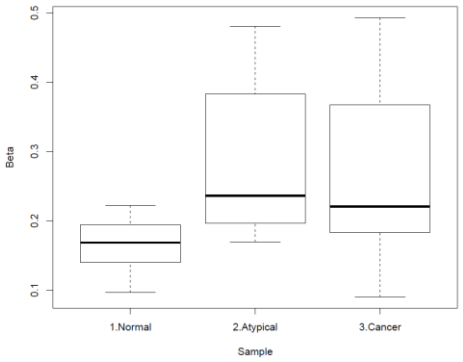
PENK



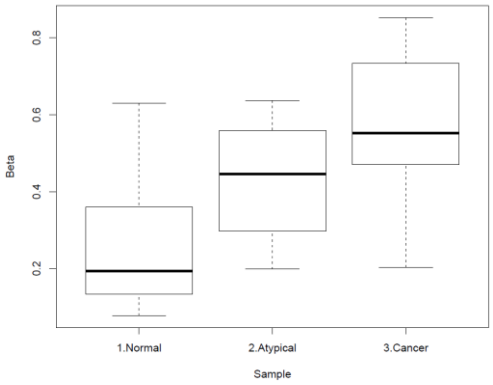
PTPN6



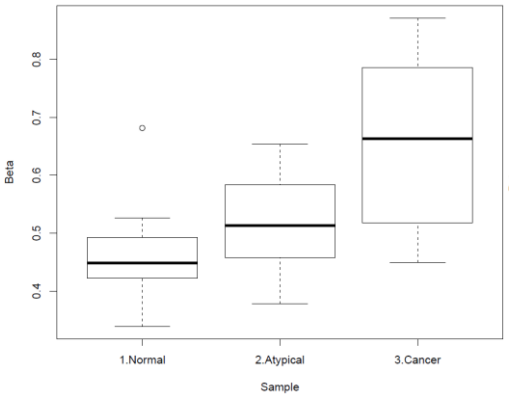
RIN1



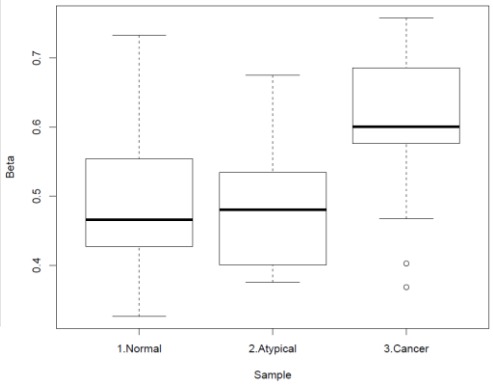
SFRP2



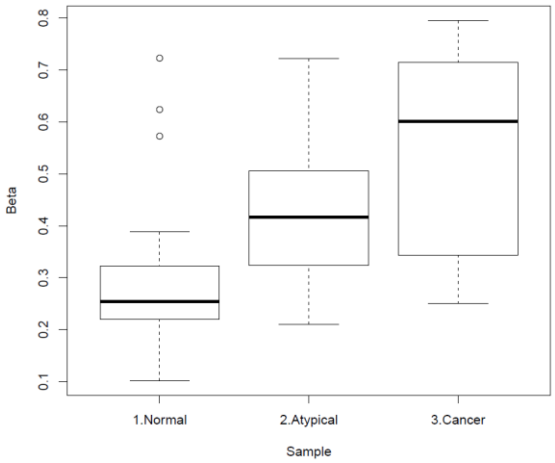
SHC1



TGIF1

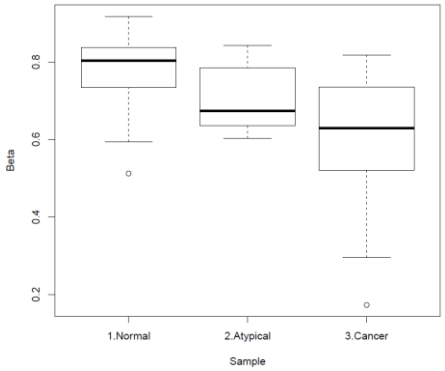


THY1

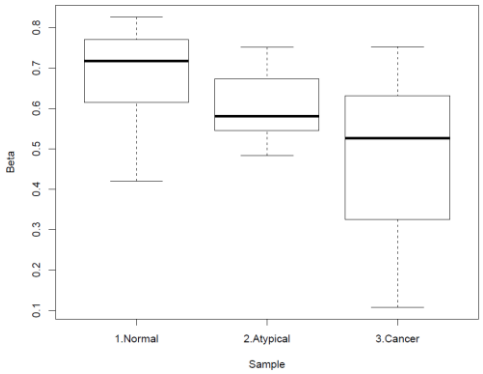


TGFB3

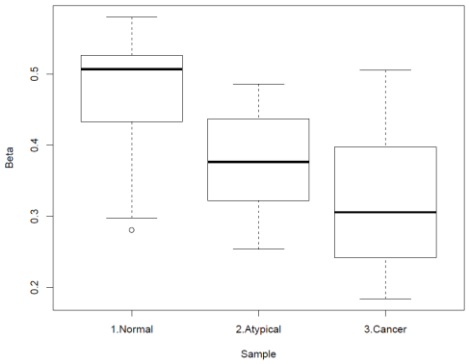
ii) Genes showing progressive hypomethylation between normal endometrium, atypical endometrial hyperplasia and endometrioid endometrial cancer



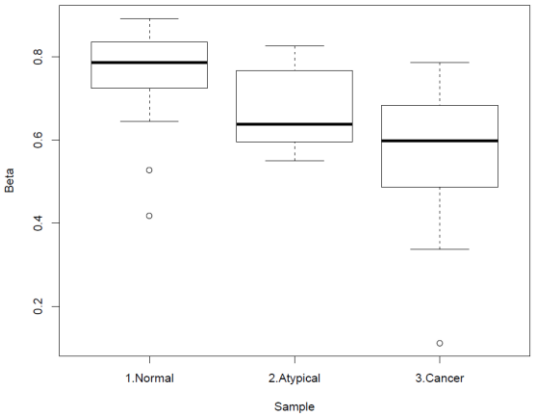
AQP5



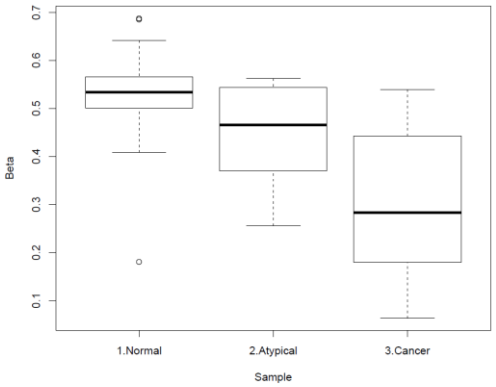
CLDN4



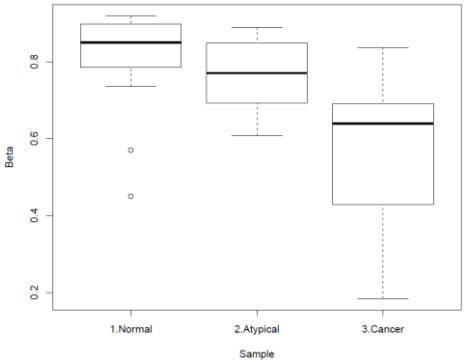
CFLAR



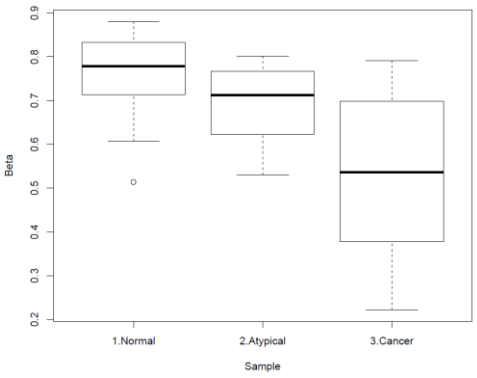
LAMA3



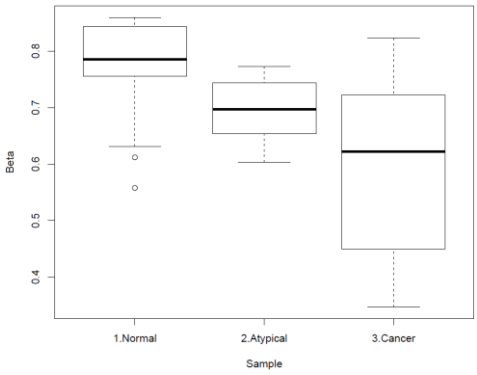
RARA



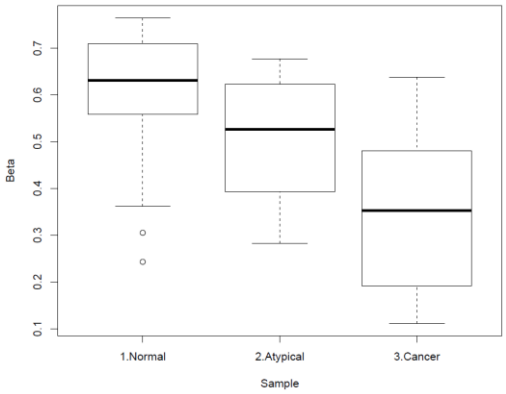
SERPINB5



SFN



SPP1



TJP2