

## Molecular stratification of endometrioid ovarian carcinoma predicts clinical outcome

Robert L Hollis<sup>1,a</sup>, John P Thomson<sup>1,a</sup>, Barbara Stanley<sup>1,a</sup>, Michael Churchman<sup>1</sup>, Alison M Meynert<sup>2</sup>, Tzyvia Rye<sup>1</sup>, Clare Bartos<sup>1</sup>, Yasushi Iida<sup>1,4</sup>, Ian Croy<sup>1</sup>, Melanie Mackean<sup>3</sup>, Fiona Nussey<sup>3</sup>, Aikou Okamoto<sup>4</sup>, Colin A Semple<sup>2</sup>, Charlie Gourley<sup>1,b</sup> and C. Simon Herrington<sup>1,b,\*</sup>

<sup>1</sup>Nicola Murray Centre for Ovarian Cancer Research, Cancer Research UK Edinburgh Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, UK

<sup>2</sup>MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, UK

<sup>3</sup>Edinburgh Cancer Centre, Western General Hospital, Edinburgh, UK.

<sup>4</sup>The Jikei University School of Medicine, Tokyo, Japan

<sup>a, b</sup> These authors contributed equally

**\*Corresponding Author:** Professor C Simon Herrington - Nicola Murray Centre for Ovarian Cancer Research, Cancer Research UK Edinburgh Centre, MRC IGMM, University of Edinburgh. Crewe Road South, Edinburgh, EH4 2XU, UK. Fax: 0044 131 651 8711; Tel: 0044 131 651 8608; email: simon.herrington@ed.ac.uk.

**Funding:** RLH was supported by an MRC-funded Research Fellowship. BS was supported by the Oncology Endowment Fund (University of Edinburgh) and Edinburgh Lothian Health Fund. We extend our thanks to The Nicola Murray Foundation for their generous support of our laboratory. CAS and AMM are supported by core MRC funding to the MRC Human Genetics Unit.

**Conflicts of interest:** MM: honoraria from Tesaro, BristolMyersSquibb and Roche. FN: non-personal interests in AstraZeneca and Tesaro. CG discloses: research funding from AstraZeneca, Aprea, Nucana, Tesaro and Novartis; honoraria/consultancy fees from Roche, AstraZeneca, MSD, Tesaro, Nucana, Clovis, Foundation One, Cor2Ed and Sierra Oncology; named on issued/pending patents related to predicting treatment response in ovarian cancer outside the scope of the work described here. All other authors declare no conflicts of interest.

**Running title:** Molecular stratification of endometrioid ovarian carcinoma

**Keywords:** ovary; cancer; carcinoma; endometrioid; molecular; genomic; stratification; algorithm; survival

## Abstract

Endometrioid ovarian carcinoma (EnOC) is an under-investigated type of ovarian carcinoma. Here, we report the largest genomic study of EnOCs to date, performing whole exome sequencing of 112 cases following rigorous pathological assessment. High frequencies of mutation were detected in *CTNNB1*(43%), *PIK3CA*(43%), *ARID1A*(36%), *PTEN*(29%), *TP53*(26%) and *SOX8*(19%), a novel target of recurrent mutation in EnOC. *POLE* and mismatch repair protein-encoding genes were mutated at lower frequency (6%, 18%) with significant co-occurrence. A molecular taxonomy was constructed using a novel algorithm (PRISTINE), identifying clinically distinct EnOC subtypes: *TP53m* cases demonstrated greater genomic complexity, were frequently FIGO stage III/IV at diagnosis (48%) and incompletely debulked (44%), and demonstrated inferior survival; conversely, *CTNNB1m* cases demonstrated low complexity and excellent clinical outcome, were predominantly stage I/II at diagnosis (89%) and completely resected (87%). Tumour complexity provides further resolution within the *TP53wt/CTNNB1wt* group. Moreover, we identify the WNT, MAPK/RAS and PI3K pathways as good candidate targets for molecular therapeutics in EnOC.

## Introduction

Ovarian carcinomas (OC) are a heterogeneous group of malignancies comprising 5 core histological types, each with distinct pathological characteristics, molecular landscapes and clinical behaviour<sup>1, 2</sup>. Endometrioid OC (EnOC) accounts for approximately 10% of all OCs, with the majority of cases diagnosed as low grade, early stage disease with excellent clinical outcome<sup>3, 4</sup>.

Currently, the management of EnOC follows the historic ‘one size fits all’ approach of aggressive cytoreductive surgery with adjuvant platinum-taxane chemotherapy for patients with disease that has progressed beyond the ovary/fallopian tube. By contrast, routine molecular stratification of care is emerging in other OC types, most notably with the advent of poly(adenosine diphosphate-ribose) polymerase (PARP) inhibitor therapy in *BRCA1/2*-mutant high grade serous (HGS) OC cases<sup>5, 6</sup>.

Targeted sequencing approaches have identified *PTEN*, *ARID1A*, *PIK3CA*, *KRAS*, *CTNNB1* and genes encoding mismatch repair (MMR) proteins as frequently mutated in relatively small cohorts of EnOC<sup>7, 8, 9</sup>, reminiscent of endometrioid endometrial carcinomas (EnECs)<sup>10</sup>, with a *TP53* mutation rate markedly lower than their high grade serous (HGS) OC counterparts<sup>11</sup>. Recent whole genome sequencing of a small EnOC case series has recapitulated these findings and identified a small proportion of EnOC with extensive copy number aberrations more akin to the genomic instability demonstrated by HGSOCs<sup>12</sup>.

The majority of EnOC are believed to arise from endometriosis<sup>1</sup>, and most grade 1 and 2 (low grade) EnOC display a classical immunohistochemical (IHC) profile comprising WT1 negativity, wild-type p53 expression, and oestrogen receptor (ER) positivity<sup>4</sup>. These classical low grade EnOC bear close histological resemblance to EnECs<sup>13</sup>.

Grade 3 (high grade) EnOC can be challenging to differentiate from HGSOC on the basis of morphology alone<sup>4, 13</sup>. In particular, HGSOCs demonstrating the solid, pseudo-endometrioid and/or transitional-cell-like (SET) histological pattern, which may be associated with *BRCA1* mutations<sup>14</sup>, represent a population easily misclassified as EnOC. Indeed, it is now recognised that many historically diagnosed high grade EnOCs in fact represent HGSOCs, supported by transcriptomic studies demonstrating that a proportion of high grade EnOCs cluster with HGSOCs<sup>15, 16, 17, 18</sup>. As such, true high grade EnOC are increasingly rare, representing only around 5-19 % of EnOC cases<sup>4, 8, 19, 20</sup>; these cases reportedly experience poor clinical outcome, in contrast to their low grade counterparts<sup>3, 20</sup>.

To date, the overwhelming body of clinical and molecular EnOC characterisation has been confounded by the inclusion of historically misclassified HGSOCs. Mutational analyses performed by more recent studies have either been applied to low grade EnOC alone<sup>21</sup>, or lack information on grade or diagnostic

criteria used<sup>7, 22</sup> and have ubiquitously analysed small patient cohorts with insufficient power to confidently correlate molecular events with patient outcome. As a result, the molecular landscape of EnOC, in particular high grade EnOC, is poorly defined.

WT1 immunohistochemistry (IHC) is a useful tool to discriminate high grade EnOC (WT1 negative) from HGSO (WT1 positive), reducing inter-observer variation<sup>15, 19, 23, 24, 25, 26</sup>. Here we perform molecular characterisation of contemporarily defined EnOC, with the use of IHC for WT1. We perform whole exome sequencing (WES) to define the genomic landscape of EnOC, including high grade EnOC, in a sizeable cohort of otherwise unselected patients and correlate specific molecular events with differential patient outcome.

## Results

### **Clinical characteristics.**

Of 289 historically diagnosed EnOC cases identified with available tumour material, 112 WT1 negative cases were characterised by WES following rigorous pathology review (18 non-evaluable, 8 non-ovarian primary, 120 WT1 positive, 14 clear cell/mucinous/carcinosarcoma/mixed histology, 2 concurrent metastatic malignancy, 15 DNA extraction/WES failure) (figure 1). The clinicopathological characteristics of these patients are shown in table 1. The majority (78.2%, 86 of 110 evaluable cases; 2 unknown stage) of patients presented with stage I or II disease. 19 cases (17.0%) had synchronous endometrial cancer diagnosis. The median follow-up time was 13.0 years. 5-year disease-specific survival (DSS) and progression-free survival (PFS) across the cohort was 72.8% (95% CI 64.8-81.8%) and 68.5% (95% CI 60.2-77.9%).

### **Molecular landscape of WT1 negative EnOC from whole exome sequencing**

The most commonly mutated target genes included *CTNNB1* (48 cases, 42.9%), *PIK3CA* (48 cases, 42.9%), *ARID1A* (40 cases, 35.7%), *PTEN* (33 cases, 29.5%), *KRAS* (29 cases, 25.9%) and *TP53* (29 cases, 25.9%) (figure S1 and S2). Unsupervised hierarchical clustering across the 50 most commonly mutated genes across the cohort revealed mutation of *TP53* and *CTNNB1* (*TP53m* and *CTNNB1m*) as the most prominent stratifying events (figure 2). *TP53m* and *CTNNB1m* were largely mutually exclusive, with significant depletion of *CTNNB1m* in the *TP53m* group ( $P<0.001$ ; co-occurrence in 1 case, 0.9%).

Mutation in *SOX8*, a novel target of recurrent mutational disruption in EnOC, was also identified at high frequency (21 cases, 18.8%), alongside other novel targets of mutational events (figure 3A, supplemental figure S3A). There was significant enrichment of *SOX8m* in the *TP53m* group (10/29, 34.5% *SOX8m* in the *TP53m* group vs 11/83, 13.3% *SOX8m* in the *TP53* wild-type (*TP53wt*) group;  $P=0.025$ ). Events in other genes previously reported as mutated in EnOC or endometrial cancers were identified at lower frequency, including *FBXW7m* (14 cases, 12.5%), *KMT2D/MLL3m* (12 cases, 11%), *APC* (6 cases, 5.4%), *BRCA1/2m* (14 cases, 12.5%), *MTORm* (7 cases, 6.5%), *PIK3R1m* (10 cases, 8.9%) and *PPP2R1Am* (4 cases, 3.6%) (figure 3A, S1 and S2). Combined analysis of frequently mutated pathways across the total EnOC tumour set identified a large number of mutations across four major oncogenic pathways: PI3K-AKT, WNT, RAS & NOTCH (supplemental figure S4).

### **Tumour mutational burden and microsatellite instability in EnOC**

A median of 78 variants were detected per sample (range 42-2,894) (supplemental figure S5 & s6A). 10 tumours (8.9%) were considered ‘hypermutated’ (>250 mutations per sample) and 3 (2.7%) were considered ‘ultramutated’ (>1000 mutations) (Figure 3A). Overall analysis of tumour mutational burden (TMB) against TCGA derived datasets places EnOC alongside HGSOC (median mutation count = 72), colonic adenocarcinoma (median = 76) and EnEC (median = 78) (supplemental figure S6B).

Mutations in one or more genes encoding MMR proteins were identified in 20 (17.9%) cases (figure 3C), most commonly in *MSH3*, *MSH6*, *MLH3* or *MSH4*. The majority of MMR-mutant (MMRm) tumours were *TP53*wt (18/20 cases, 90.0%). Frameshift or nonsense MMRm were associated with significantly higher microsatellite instability (MSI) scores compared to those containing missense MMRm (median 236 vs 142.5, P<0.001) and MMRwt samples (median 236 vs 162, P<0.001) (figure 3A & supplemental figure S7). There was no significant difference in MSI score between MMRwt cases and those with missense MMRm (P=0.181).

*POLE*m were detected in a minority of tumours (7 cases, 6.3%) with 3 (42.8%) found to occur over a hotspot within the exonuclease domain (Figure 3A & supplemental figure S3B). There was a high frequency of concurrent *POLE*m and MMRm (5 of 7 *POLE*m cases, 71.4%), with significant enrichment for *POLE*m in the MMRm group versus MMR wild-type cases (5/20, 25.0% vs 2/92, 2.2%, P=0.002). Together, the *POLE*m and nonsense/frameshift MMRm cases accounted for majority of high TMB cases (figure 3D). Cases with concurrent *POLE*m and MMRm accounted for all three ultramutated tumours; 8 of the 10 (80.0%) hypermutated tumours contained either *POLE*m or frameshift/nonsense MMRm.

### **Mutational spectrum in endometrioid ovarian carcinoma**

We observed a bias towards C>T and C>A transversion and transition molecular signatures across the 112 tumours (figure 3A). A shift in signatures was observed in samples harbouring *POLE*m, with a greater proportion of T>G changes, and depletion of C>G and T>A substitutions in this population (supplemental figure S8).

### **Tumour complexity and copy number variation**

Distribution of per-sample global variant allele frequency (VAF) density and calculation of mutant-allele tumour heterogeneity (MATH) score across the 112 EnOC samples was used to infer tumour complexity (figure 4). *TP53*wt tumours were predominantly low complexity, demonstrating lower

MATH scores (median 27.5 vs 54.7, P<0.001) (figure 4B & supplemental figure S9) and fewer discrete VAF peaks (P<0.001) (supplemental table S1) compared to *TP53m* cases.

Analysis of copy number variation (CNV) across the samples revealed differential CNV burden across the molecular subgroups (supplemental figure S10A & s10B), with *TP53m* cases harbouring greater CNV burden compared to *TP53wt* tumours (P<0.0001). The most frequent CNVs across the cohort were CN gain of *ZNF43* (30 cases, 26.8%), *RABA1C* (24 cases, 21.4%) and *AMY1C* (23 cases, 20.5%), and CN loss of *PKNOX1* (42 cases, 37.5%), *CEP68* (29 cases, 25.9%) and *GLB1L* (20 cases, 17.9%) (supplemental figure S10C) with CNV events also occurring over some of the genes identified as containing frequent SNVs (supplemental figure 10D).

### ***Molecular events in EnOC define clinically distinct subgroups of disease***

*TP53m* cases demonstrated significantly inferior survival upon univariable analysis (HR for DSS=4.43, 95% CI 2.27-8.64, Bonferroni-adjusted P<0.001) (figure 5Ai, and table S2), were more likely to be diagnosed at advanced stage (14 of 29 evaluable cases, 48.3% stage III/IV vs 10 of 81, 12.4%; P<0.001) (table S3), less likely to be successfully resected to zero macroscopic residual disease (44.4%, 12 of 27 evaluable cases with macroscopic RD after surgical debulking vs 14.1%, 11 of 78; P=0.003), and demonstrated a trend for greater age at diagnosis which did not meet statistical significance (median 61 vs 57 years, P=0.063). Multivariable analysis accounting for patient age, stage at diagnosis and extent of residual disease following primary cytoreduction identified *TP53* status as independently associated with survival (P=0.031 for DSS; P=0.011 for PFS) (table S4 and S5, figure S11). The *TP53m* group demonstrated significant depletion of cases with concurrent endometrial cancer diagnosis (3.4%, 1 of 29 *TP53m* vs 21.7%, 18 of 83 *TP53wt*, P=0.022).

By contrast, cases with *CTNNB1m* were almost ubiquitously stage I/II at diagnosis (89.1%, 41 of 46 evaluable cases) and debulked to zero macroscopic RD (87.0%, 42 of 46 evaluable cases), with markedly favourable outcome (HR for DSS=0.23, 95% CI 0.10-0.56, Bonferroni-adjusted P=0.010) (table S2) which was significant upon multivariable analysis (P=0.017 for DSS; P=0.006 for PFS) (figure 5Aii, supplemental s6 and s7). *CTNNB1m* was significantly associated with favourable outcome specifically in the context of *TP53wt* cases (HR for DSS 0.30, 95% CI 0.11-0.87), and *TP53* wild-type/*CTNNB1m* cases were less genetically complex versus their *TP53wt/CTNNB1wt* counterparts (median MATH score 36.3 vs 31.9, P=0.035) (figure 4C).

Greater tumour complexity was associated with inferior survival when defined by the number of VAF peaks (P=0.020 for DSS) or continuous MATH score (P<0.001 for DSS) across the cohort (supplemental

table S7). Moreover, MATH score at a threshold of the median score in low complexity tumours (with a single VAF peak, figure 4Aiii), provided additional meaningful prognostic information specifically in the absence of *TP53m* or *CTNNB1m* (figure 5B, figure S11). Together, these data support a clinically meaningful classification system driven firstly by *TP53m* status, then *CTNNB1m* status, with genomic complexity providing further discrimination of outcome within the *TP53wt/CTNNB1wt* population; we define this step-wise molecular classification system as the **PRISTINE algorithm :PRognostic genoMlc StratificaTion of EnOC** (figure 5C).

## Discussion

The molecular landscape of EnOC is poorly defined, particularly in high grade cases, due to under-investigation and historic misclassification of HGSOC as high grade EnOC in older studies. WT1 negativity has emerged as an important discriminator of high grade EnOC from HGSOC, which displays morphological similarities<sup>4, 23, 24, 25, 26</sup>. To our knowledge, this is the largest report of genetically characterised EnOC to date, utilising routine WT1 IHC to exclude pseudo-endometrioid HGSOCs carcinomas that have contaminated previous studies of this OC type.

In line with previous sequencing studies in small cohorts of EnOC, we identified a high mutation rate of genes known to be perturbed in EnEC, the most frequent of which were *CTNNB1*, *PIK3CA*, *PTEN*, *ARID1A*, *KRAS* and *TP53*<sup>7, 10, 21, 27, 28</sup>. Our EnOC cohort demonstrated a similar rate of *TP53*m to the TCGA study of EnEC<sup>10</sup>; by contrast, the mutation rate of *PTEN* was markedly lower.

We demonstrate that EnOC tumours contain a moderate tumour mutational burden with respect to other cancer types, with a small proportion hyper- or ultra- mutated in nature (11.6%, 13 of 112). MMR deficiency due to mutations or methylation in the MMR genes results in microsatellite instability, and loss of MMR protein expression has previously been demonstrated in approximately 7-14% of EnOC<sup>29, 30, 31, 32, 33</sup>. As in EnEC, we identify a subgroup of EnOC harbouring MMR mutations, with a similar overall MMR mutation rate as reported previously in EnOC<sup>29, 30, 31, 32, 33</sup>. We demonstrate that cases with truncating MMRm (nonsense or frameshifting) demonstrate MSI and that these cases account for many of the samples demonstrating high TMB. Moreover, mutation of *POLE* was rare in this cohort (6.3% of cases), there was significant enrichment for this event in the MMRm population, and concurrent mutation of *POLE* and MMR genes was present in all three ultramutated tumours. Moreover, *POLEM* cases demonstrated a distinct mutational profile. However, unlike in EnEC<sup>10</sup>, *POLEM* was not associated with a significant survival benefit, though the power of these analyses was severely limited. This low *POLEM* rate is consistent with previous reports in EnOC<sup>34, 35</sup>.

Collectively, the data presented here identify *TP53*m EnOC as a distinct clinical and biological subtype of EnOC. *TP53*m cases demonstrated higher levels of copy number aberrations, greater tumour genomic complexity, higher rate of advanced stage at diagnosis, inferior rate of complete macroscopic tumour resection, and overall poor clinical outcome. This is consistent with the poor outcome reported in EnEC harbouring *TP53*m<sup>10</sup>, is reminiscent of HGSOC<sup>11</sup>, and is in line with several studies of EnOC<sup>27, 36</sup>. In particular, the study performed by Parra-Herran et al<sup>27</sup> applied the PROMISE algorithm, a surrogate of the EnEC molecular classifier, to a cohort of WT1 negative EnOC and found the group with aberrant p53 expression to have the worst survival. In our cohort, *TP53*m cases also represented those least likely to demonstrate concurrent endometrial cancer.

While the copy number and survival profile of our *TP53m* EnOC group provides one rationale for reclassification of these tumours as HGSOC, the high frequency (48.3%, 14 of 29) of MMRm or classic EnOC mutations (*CTNNB1*, *PTEN*, *ARID1A*, *KRAS* or *PIK3CA*), lack of WT1 expression (in all cases) and high rate of early stage diagnosis in this cohort (51.7%, 15 stage I/II) form a compelling argument that these represent true EnOC. Only 9 cases (8.0% of our cohort) represented feasible candidates as possible true WT1-negative HGSOCs (MMR, *PTEN*, *CTNNB1*, *PIK3CA*, *ARID1A*, *KRAS* wild-type tumours with *TP53m* and diagnosed at stage III/IV), which are recognised as a rare phenomenon (<5% HGSOC cases)<sup>23</sup>.

Conversely, *CTNNB1m* – which appears mutually exclusive with *TP53m* – is associated with low genomic complexity, early-stage disease that is easily debulked to zero macroscopic RD and demonstrates excellent clinical outcome. This is in contrast to findings in low grade early stage EnEC associating *CTNNB1m* with a greater chance of recurrence<sup>37</sup>. Within our EnOC cohort, *CTNNB1m* status was also associated with superior outcome specifically in the context of *TP53wt* tumours, suggesting clinical impact independent of its anti-correlation with *TP53m*. These data support the notion of a tiered classification by *TP53m* and *CTNNB1m* status. The outcome of the remaining *TP53wt/CTNNB1wt* patients, who demonstrate an overall intermediate survival profile and moderate genomic complexity, can be further resolved by tumour complexity score. *TP53* status, *CTNNB1* status and tumour complexity can therefore be used in succession within the PRISTINE algorithm for defining molecular subtypes with markedly differential clinical outcome and clinicopathological features.

The high rate of genomic disruption in *CTNNB1*, *KRAS*, *PTEN* and *PIK3CA* suggests that inhibitors of the WNT, MAPK and PI3K pathways represent agents with potential clinical utility in EnOC treatment. Efforts to identify novel therapeutic strategies should focus on cases with greatest unmet clinical need, namely *CTNNB1wt* cases. In particular, *TP53m* cases represent those where further treatment options are urgently required to improve outcome, and we identified potentially clinically actionable mutations in a large proportion of these cases (13.8% with *KRASm*, 27.6% with *PTENm/PIK3CAm*, 27.6% with *BRCA1/2m*).

Finally, we identify *SOX8* as a potential novel target of mutational disruption in EnOC. SRY-related high mobility group (HMG) box (*SOX*) genes encode a family of transcription factors which act as critical regulators of cellular programming and are frequently altered in many cancers<sup>38</sup>. Interestingly, analysis of TCGA data reveals that *SOX8m* occur at low frequencies in uterine cancers (2.64% in uterine corpus endometrial carcinoma, 1% in uterine carcinosarcoma) as well as in colonic adenocarcinoma (2.25%) (data from the TCGA data portal<sup>39</sup>). As recent studies suggest that *SOX8* in part regulates the activity of genes associated with the WNT/β-catenin pathway, a commonly mutated pathway in EnOC,

mutation of this gene may impact on classically defined EnOC pathways through this route<sup>40</sup>. Genomic disruption of *SOX8* therefore represents a novel candidate mechanism by which EnOCs frequently activate the WNT/β-catenin pathway; given the frequent co-occurrence of *SOX8m* and *TP53m* in our cohort (34.5% *SOX8m* in the *TP53m* group), selection bias against true high grade EnOCs in previous studies - leading to depletion of *TP53m* cases in those cohorts - may well explain why *SOX8m* has not previously been identified as a common genomic event in EnOC.

In summary, we have demonstrated that EnOC is a molecularly heterogeneous disease, comprising molecular subtypes of disease which we define here using the PRISTINE algorithm. These subtypes demonstrate differential clinical outcome and clinicopathological features. In particular, our study highlights *CTNNB1m* and *TP53m* as markers of biologically distinct subtypes of EnOC with contrasting clinical behaviour. These markers have the potential to inform future prognostication and molecular stratification within EnOC. Gene sequencing of *TP53* and *CTNNB1*, or IHC directed at their respective gene products, represent mechanisms by which these findings could readily be translated into routine risk-stratification of newly diagnosed cases. EnOCs demonstrating absence of *CTNNB1m* and/or presence of *TP53m* represent patients with greatest unmet clinical need, many of whom harbour activating mutations in pathways that represent opportunities for therapeutic intervention with molecular agents. Investigating the clinical efficacy of inhibitors targeting the MAPK/RAS, WNT and PI3K pathways in this context has the potential to identify agents that will improve EnOC patient survival.

## Methods

### *Ethical approval*

Ethical approval for the use of human tissue specimens for research was obtained from South East Scotland Scottish Academic Health Sciences Collaboration (SAHSC) BioResource (reference: 15/ES/0094-SR494). Correlation of molecular data to clinical outcome and clinicopathological variables in ovarian cancer was approved by NHS Lothian Research and Development (reference 2007/W/ON/29).

### *Pathology review*

505 patients diagnosed with OC between August 1968 and May 2014, and whose pathology reports contained the term ‘endometrioid’, were identified through the Edinburgh Ovarian Cancer Database (figure 1); tumour material was available for 289 cases<sup>41</sup>. Chemotherapy naïve tumour from the primary site was selected where available. Pathology review was conducted as per WHO 2014 classification, including immunohistochemistry (IHC) for WT1 in every case (Supplementary Methods

Section 1A), by an expert gynaecological pathologist (CSH). The presence of endometriosis was recorded from the reviewed slides or pathology report.

Cases with non-interpretable morphology, non-evaluable tumour and cases representing metastases from primary endometrial cancer, as defined by WHO criteria, were excluded. Ovarian metastases, WT1 positive tumours, carcinosarcomas and carcinomas of clear cell, mucinous or mixed histology were also excluded (figure 1). IHC for CK7 (Leica, Clone RN7, 1:100 dilution) and CK20 (Leica, clone KS20.8, 1:50 dilution) was performed to exclude colorectal adenocarcinoma metastases (Supplementary Materials Section 1B).

### ***Clinical data***

Baseline characteristics and outcome data were extracted from the Edinburgh Ovarian Cancer Database, wherein the diagnostic, treatment and follow-up data for every ovarian cancer patient treated at the Edinburgh Cancer Centre is prospectively entered as part of routine care. Disease-specific survival (DSS) was calculated from the date of pathologically confirmed OC diagnosis. Progression-free survival (PFS) was recorded as the duration between the date of diagnosis to the date of first radiological progression or recurrence, or death from EnOC.

### ***DNA extraction***

H&E-stained slides were marked by an expert gynaecological pathologist (CSH) to identify tumour areas suitable for macrodissection in order to enrich for tumour cellularity. DNA extraction was performed using the QIAamp DNA FFPE Tissue Kit (Qiagen) and Deparaffinization Solution according to the manufacturer's instructions.

### ***Whole exome sequencing***

Exome capture was performed using the Illumina TruSeq Exome Library Prep kit (supplementary materials section 2A) and WES was performed on the Illumina NextSeq 550 (Illumina, Inc., San Diego, CA, USA). The median per-sample on-target coverage in the successfully sequenced samples was 89.5X (range 36X-289X). Data were aligned to the GRCh38 human reference genome using bwa-0.7.17<sup>42</sup>, duplicates marked and base quality scores recalibrated with the GenomeAnalysisToolkit (GATK) v4<sup>43</sup> in the bcbio 1.0.6 pipeline (Supplementary Materials Section 2B).

### ***Variant calling and classification***

Variant calling was performed using a majority vote system from three variant caller algorithms: VarDict<sup>44</sup>, Mutect2<sup>45</sup> and Freebayes<sup>46</sup>. Filtering for C>T (FFPE artifacts) and G>T (oxidation artifacts) was applied using GATK (CollectSequencingArtifactMetrics and FilterByOrientationBias). Variants associated with low sequence depth (<20X) or low variant allele frequency (<10%) were removed.

Common variants were excluded using the 1000 genomes and ExAC reference datasets; known pathogenic and benign variants were flagged using ClinVar<sup>47</sup>, and remaining variants were filtered to remove likely non-functional variation using the Polymorphism Phenotyping (PolyPhen)<sup>48</sup> and Sorting Intolerant from Tolerant (SIFT)<sup>49</sup> functional prediction tools (Supplementary Materials Section 2C).

Microsatellite instability (MSI) score was assessed as the number of short insertions or deletions (InDels) detected in a given sample. TMB was defined as the number of mutations present in a given tumour following the aforementioned filtering steps. TMB across other cancer datasets in The Cancer Genome Atlas (TCGA) were contrasted against those in our EnOC datasets<sup>50</sup>. Transitions and transversions were calculated using the titv function in maftools<sup>51</sup>.

Unsupervised analysis was performed across the top 50 differentially mutated genes within the total tumour dataset represented as a binary matrix (0, wild-type; 1, mutant). Samples were clustered by Euclidian distance and Ward's linkage based on the overall Pearson correlation score of these binary signatures. Heat maps were drawn in R using the ggplot package. Supervised mutational analysis was performed using the most commonly mutated genes across sequenced samples. Genomic events in these genes and overall tumour mutational burden analysis were visualised using the R package maftools<sup>51</sup>. Pathway analysis was carried out using the OncogenicPathways function<sup>52</sup>.

### **Tumour complexity scoring**

Tumour genomic complexity was assessed by VAF density using the inferHeterogeneity function in the R package maftools<sup>51, 53</sup> (supplementary materials section 3). Resulting MATH scores represent the width of the VAF distribution; specimens of low complexity with a single driver event and associated outgrowth demonstrate fewer VAF peaks with a lower MATH score. Conversely, highly complex tumours with multiple driver events, branched evolution and multiple subclonal populations demonstrate multiple VAF peaks and higher MATH score. With the PRISTINE algorithm, the TP53wt/CTNNB1wt group were split into higher and lower complexity using MATH score, at a cutoff equal to the median MATH score of low complexity tumours (with a single VAF peak, figure 4Aiii).

### **Copy number variation (CNV) detection**

Copy number analysis was performed using GeneCN pipelines using Bio-DB-HTS version 2.10 (see Supplementary Materials Section 4) to identify regions of significant copy number gain or loss (CN score >5 standard deviations from reference, P<0.05) using the pooled TP53wt samples as a reference population.

### **Statistical analysis**

Statistical analyses were performed using R version 3.5.1. Comparisons of continuous data were made with the Mann Whitney-U test or T-test, as appropriate. Median follow-up time was calculated using the reverse Kaplan-Meier method. Survival analysis was performed using the Cox proportional hazards regression analysis in the Survival package<sup>54</sup>. Multivariable analyses accounted for FIGO stage, patient age at diagnosis, decade of patient diagnosis and extent of residual disease following surgical cytoreduction. Comparisons of frequency were performed using the Chi-squared test or Fisher's exact test, as appropriate. Correction for multiplicity of testing was performed using the Bonferroni method where appropriate.

### Acknowledgements

We extend our thanks to the patients who contributed to this study and to the Edinburgh Ovarian Cancer Database from which the clinical data reported here were retrieved. We thank the Edinburgh Clinical Research Facility, Western General Hospital, Edinburgh, UK for their support with the high throughput sequencing described here. We are grateful to the NRS Lothian Human Annotated Bioresource, NHS Lothian Department of Pathology and Edinburgh Experimental Cancer Medicine Centre for their support.

### Data availability

The data used to generate the analyses presented here are currently being submitted to the European Genome-phenome Archive (EGA) and, once data upload is completed, will be made available upon request to our data access committee. For more information please see <https://ega-archive.org/access/data-access>

### Author contributions statement

RLH contributed conceptualisation, data analysis, visualisation, manuscript writing; JPT contributed data analysis, visualisation, manuscript writing; BS contributed conceptualisation, cohort collation, data collection, manuscript writing; MC contributed project management, tissue access, manuscript review and editing; AMM contributed data analysis, manuscript review and editing; TR, CB and YI contributed data collection, manuscript review and editing; IC contributed tissue processing, manuscript review and editing; MM and FN contributed access to patients/tissue, manuscript review and editing; AO contributed supervision, manuscript review and editing; CAS contributed supervision, manuscript review and editing; CG contributed: conceptualisation, access to patients/tissue, supervision, funding acquisition, manuscript review and editing; CSH contributed: conceptualisation, data collection, supervision, manuscript review and editing.

## References

1. Vaughan S, et al. Rethinking ovarian cancer: recommendations for improving outcomes. *Nat Rev Cancer* **11**, 719-725 (2011).
2. Hollis RL, Gourley C. Genetic and molecular changes in ovarian cancer. *Cancer Biol Med* **13**, 236-247 (2016).
3. Storey DJ, et al. Endometrioid epithelial ovarian cancer : 20 years of prospectively collected data from a single center. *Cancer* **112**, 2211-2220 (2008).
4. Lim D, Murali R, Murray MP, Veras E, Park KJ, Soslow RA. Morphological and Immunohistochemical Reevaluation of Tumors Initially Diagnosed as Ovarian Endometrioid Carcinoma With Emphasis on High-grade Tumors. *Am J Surg Pathol* **40**, 302-312 (2016).
5. Moore K, et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N Engl J Med* **379**, 2495-2505 (2018).
6. Mirza MR, et al. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* **375**, 2154-2164 (2016).
7. Huang HN, et al. Ovarian and endometrial endometrioid adenocarcinomas have distinct profiles of microsatellite instability, PTEN expression, and ARID1A expression. *Histopathology* **66**, 517-528 (2015).
8. Geyer JT, et al. Pathogenetic pathways in ovarian endometrioid adenocarcinoma: a molecular study of 29 cases. *Am J Surg Pathol* **33**, 1157-1163 (2009).
9. Wu R, et al. Mouse model of human ovarian endometrioid adenocarcinoma based on somatic defects in the Wnt/beta-catenin and PI3K/Pten signaling pathways. *Cancer Cell* **11**, 321-333 (2007).
10. Kandoth C, et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
11. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615 (2011).
12. Cybulski P, et al. Molecular profiling and molecular classification of endometrioid ovarian carcinomas. *Gynecol Oncol* **154**, 516-523 (2019).
13. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. *WHO Classification of Tumours of Female Reproductive Organs*. IARC press (2014).
14. Soslow RA, et al. Morphologic patterns associated with BRCA1 and BRCA2 genotype in ovarian carcinoma. *Mod Pathol* **25**, 625-636 (2012).
15. Madore J, et al. Characterization of the molecular differences between ovarian endometrioid carcinoma and ovarian serous carcinoma. *J Pathol* **220**, 392-400 (2010).
16. Schwartz DR, et al. Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Res* **62**, 4722-4729 (2002).
17. Winterhoff B, et al. Molecular classification of high grade endometrioid and clear cell ovarian cancer using TCGA gene expression signatures. *Gynecol Oncol* **141**, 95-100 (2016).
18. Tothill RW, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* **14**, 5198-5208 (2008).

19. Assem H, et al. High-grade Endometrioid Carcinoma of the Ovary: A Clinicopathologic Study of 30 Cases. *Am J Surg Pathol* **42**, 534-544 (2018).
20. Soyama H, et al. A Pathological Study Using 2014 WHO Criteria Reveals Poor Prognosis of Grade 3 Ovarian Endometrioid Carcinomas. *In vivo* **32**, 597-602 (2018).
21. McConechy MK, et al. Ovarian and endometrial endometrioid carcinomas have distinct CTNNB1 and PTEN mutation profiles. *Mod Pathol* **27**, 128-134 (2014).
22. Wang YK, et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat Genet* **49**, 856-865 (2017).
23. Kobel M, et al. An Immunohistochemical Algorithm for Ovarian Carcinoma Typing. *Int J Gynecol Pathol* **35**, 430-441 (2016).
24. Acs G, Pasha T, Zhang PJ. WT1 is differentially expressed in serous, endometrioid, clear cell, and mucinous carcinomas of the peritoneum, fallopian tube, ovary, and endometrium. *Int J Gynecol Pathol* **23**, 110-118 (2004).
25. Cathro HP, Stoler MH. The utility of calretinin, inhibin, and WT1 immunohistochemical staining in the differential diagnosis of ovarian tumors. *Hum Pathol* **36**, 195-201 (2005).
26. Al-Hussaini M, Stockman A, Foster H, McCluggage WG. WT-1 assists in distinguishing ovarian from uterine serous carcinoma and in distinguishing between serous and endometrioid ovarian carcinoma. *Histopathology* **44**, 109-115 (2004).
27. Parra-Herran C, et al. Molecular-based classification algorithm for endometrial carcinoma categorizes ovarian endometrioid carcinoma into prognostically significant groups. *Mod Pathol* **30**, 1748-1759 (2017).
28. Stewart CJ, Leung Y, Walsh MD, Walters RJ, Young JP, Buchanan DD. KRAS mutations in ovarian low-grade endometrioid adenocarcinoma: association with concurrent endometriosis. *Hum Pathol* **43**, 1177-1183 (2012).
29. Lu FI, et al. Prevalence of loss of expression of DNA mismatch repair proteins in primary epithelial ovarian tumors. *Int J Gynecol Pathol* **31**, 524-531 (2012).
30. Aysal A, Karnezis A, Medhi I, Grenert JP, Zaloudek CJ, Rabban JT. Ovarian endometrioid adenocarcinoma: incidence and clinical significance of the morphologic and immunohistochemical markers of mismatch repair protein defects and tumor microsatellite instability. *Am J Surg Pathol* **36**, 163-172 (2012).
31. Liu J, et al. Microsatellite instability and expression of hMLH1 and hMSH2 proteins in ovarian endometrioid cancer. *Mod Pathol* **17**, 75-80 (2004).
32. Bennett JA, Pesci A, Morales-Oyarvide V, Da Silva A, Nardi V, Oliva E. Incidence of Mismatch Repair Protein Deficiency and Associated Clinicopathologic Features in a Cohort of 104 Ovarian Endometrioid Carcinomas. *Am J Surg Pathol* **43**, 235-243 (2019).
33. Rambau PF, et al. Significant frequency of MSH2/MSH6 abnormality in ovarian endometrioid carcinoma supports histotype-specific Lynch syndrome screening in ovarian carcinomas. *Histopathology* **69**, 288-297 (2016).
34. Hoang LN, et al. Polymerase Epsilon Exonuclease Domain Mutations in Ovarian Endometrioid Carcinoma. *Int J Gynecol Pathol* **25**, 1187-1193 (2015).
35. Zou Y, et al. Frequent POLE1 p.S297F mutation in Chinese patients with ovarian endometrioid carcinoma. *Mutat Res* **761**, 49-52 (2014).

36. Okuda T, et al. p53 mutations and overexpression affect prognosis of ovarian endometrioid cancer but not clear cell cancer. *Gynecol Oncol* **88**, 318-325 (2003).
37. Kurnit KC, et al. CTNNB1 (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence. *Mod Pathol* **30**, 1032-1041 (2017).
38. Thu KL, Becker-Santos DD, Radulovich N, Pikor LA, Lam WL, Tsao MS. SOX15 and other SOX family members are important mediators of tumorigenesis in multiple cancer types. *Oncoscience* **1**, 326-335 (2014).
39. Grossman RL, et al. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**, 1109-1112 (2016).
40. Xie SL, et al. SOX8 regulates cancer stem-like properties and cisplatin-induced EMT in tongue squamous cell carcinoma by acting on the Wnt/beta-catenin pathway. *Int J Cancer* **142**, 1252-1265 (2018).
41. Hollis RL, et al. Hormone receptor expression patterns define clinically meaningful subgroups of endometrioid ovarian carcinoma. *Gynecol Oncol* **155**, 318-323 (2019).
42. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
43. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
44. Lai Z, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* **44**, e108 (2016).
45. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).
46. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]*, (2012).
47. Landrum MJ, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985 (2014).
48. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249 (2010).
49. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814 (2003).
50. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
51. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747-1756 (2018).
52. Sanchez-Vega F, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e310 (2018).
53. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* **49**, 211-215 (2013).
54. Therneau T. A Package for Survival Analysis in S. (2015).

**Table 1.** Clinical characteristics of the 112 cases characterised by whole exome sequencing

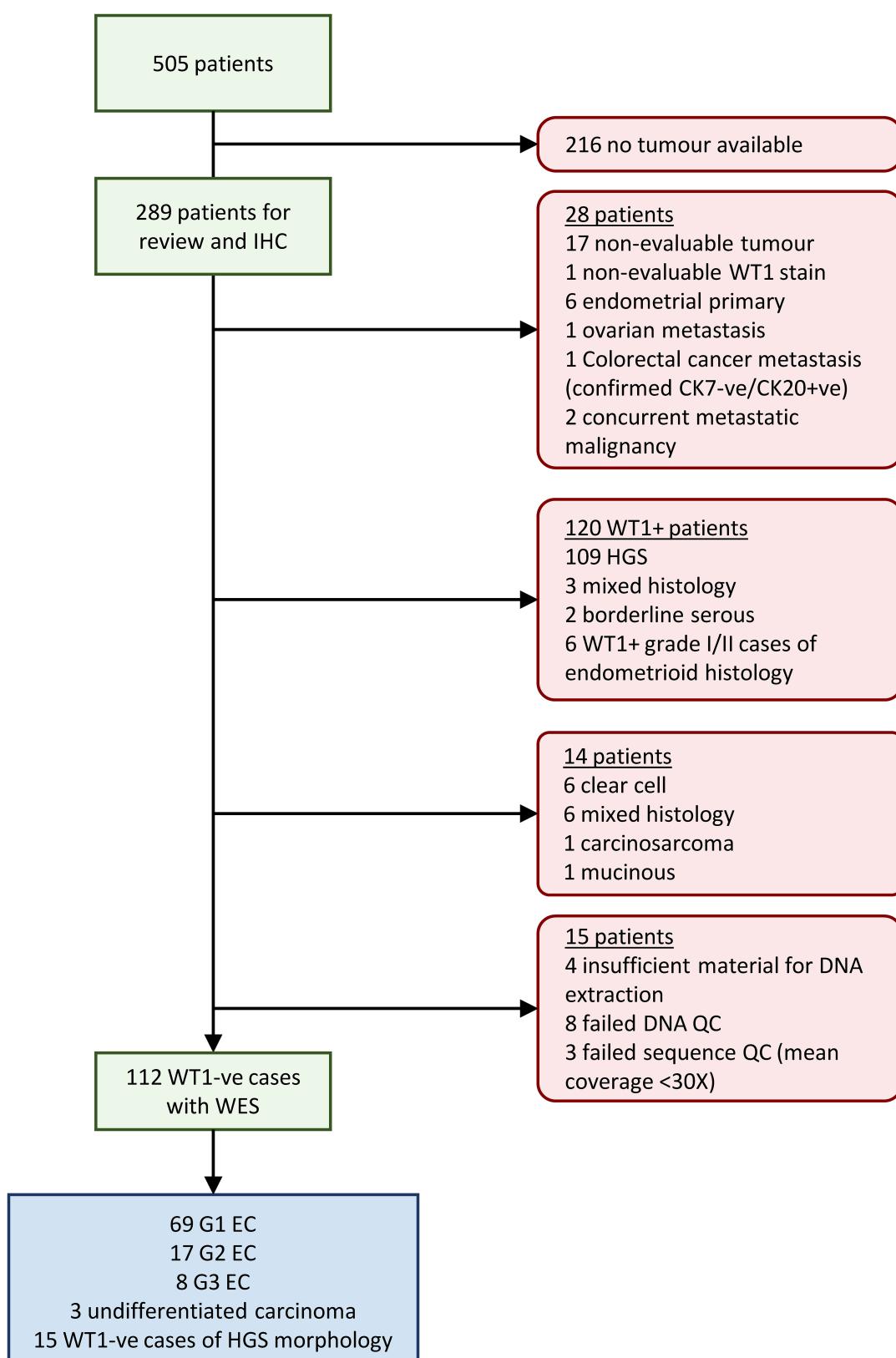
	N / median	% / range
<b>Cases</b>	112	
<b>Age (years)</b>	58.5	28-88
<b>BMI</b>	25.4	18.0-44.0
<b>Concurrent endometrial cancer<sup>a</sup></b>	19	17.0
<b>Endometriosis<sup>b</sup></b>	39	34.8
<b>Morphology</b>		
G1 EnOC	69	61.6
G2 EnOC	17	15.2
G3 EnOC	8	7.1
WT1 negative tumour of serous or undifferentiated morphology	18	16.1
<b>Vital status at last follow-up</b>		
Deceased, ovarian cancer	35	31.2
Deceased, other causes	19	17.0
Alive	58	51.8
<b>Period of Diagnosis</b>		
1980s	17	15.2
1990s	44	39.3
2000s	32	28.6
2010s	19	17.0
<b>FIGO stage at diagnosis</b>		
I	47	42.7
II	39	35.5
III	15	13.6
IV	9	8.2
NA	2	
<b>Primary debulking status</b>		
Zero macroscopic RD	82	78.1
Macroscopic RD	23	21.9
NA	7	
<b>Adjuvant chemotherapy stage I/II (n=86)</b>		
Single-agent platinum	31	36.5
Platinum-taxane combination	17	20.0
Other platinum combination	3	3.5
Other chemotherapy regime	7	8.2
No adjuvant chemotherapy	27	31.8
NA	1	
<b>Adjuvant therapy stage III/IV (n=24)</b>		
Single-agent platinum	14	58.3

Platinum-taxane combination	2	8.3
Other platinum combination	2	8.3
No adjuvant therapy	6	25.0

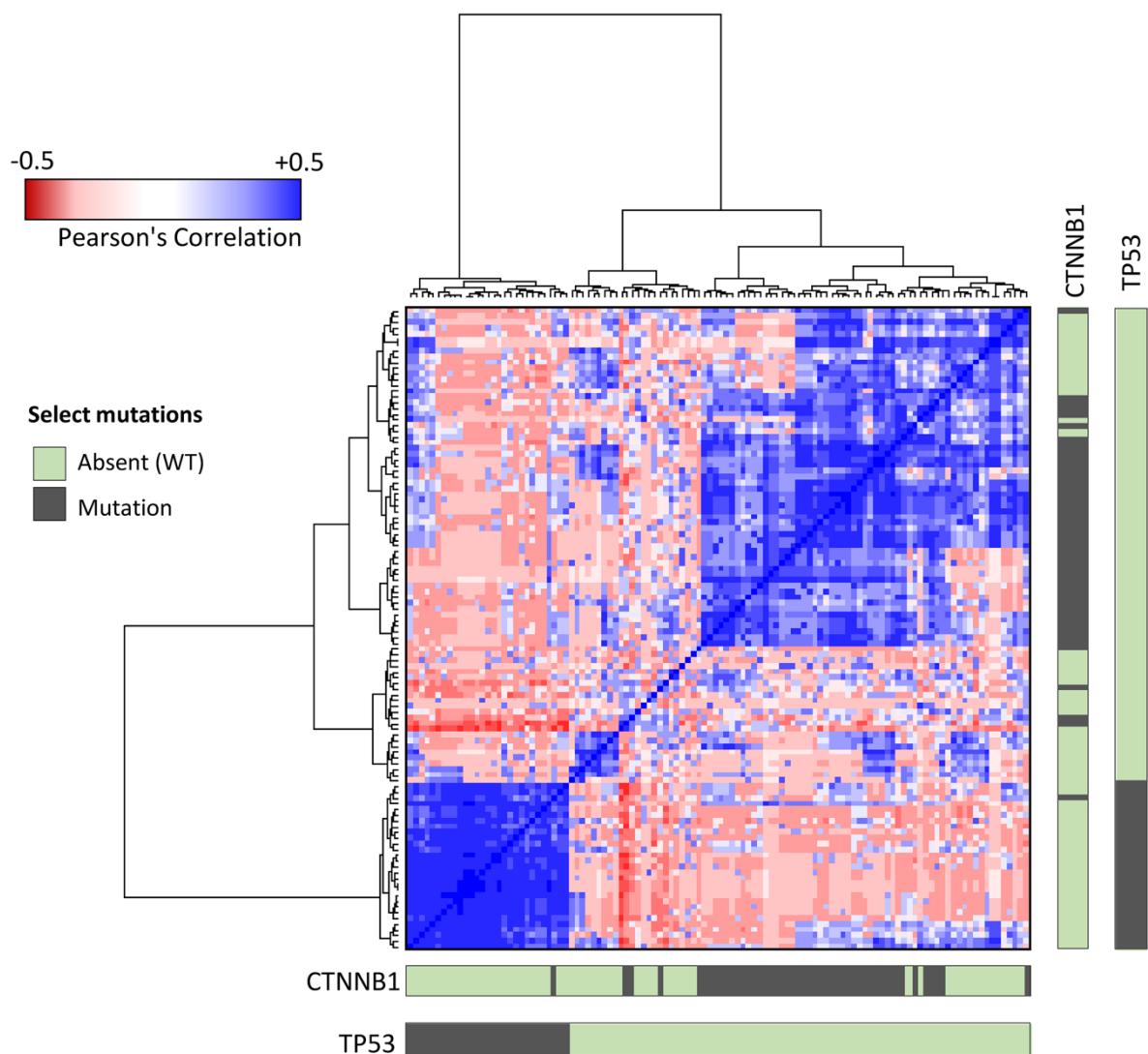
BMI, body mass index; G1, grade I; G2, grade II; G3, grade III; EnOC, endometrioid ovarian carcinoma; NA, not available; RD, residual disease

<sup>a</sup>as documented on the Edinburgh Ovarian Cancer Database

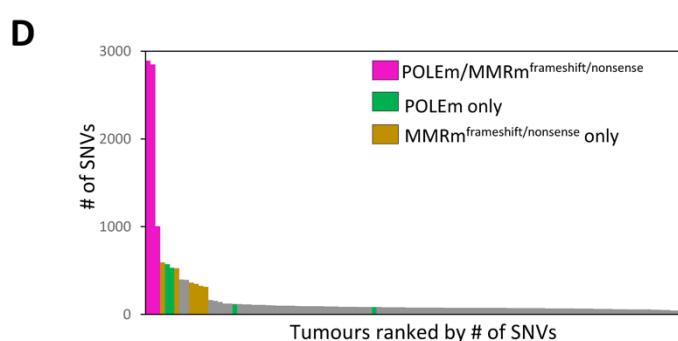
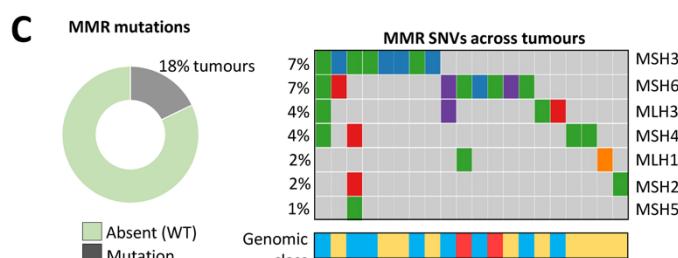
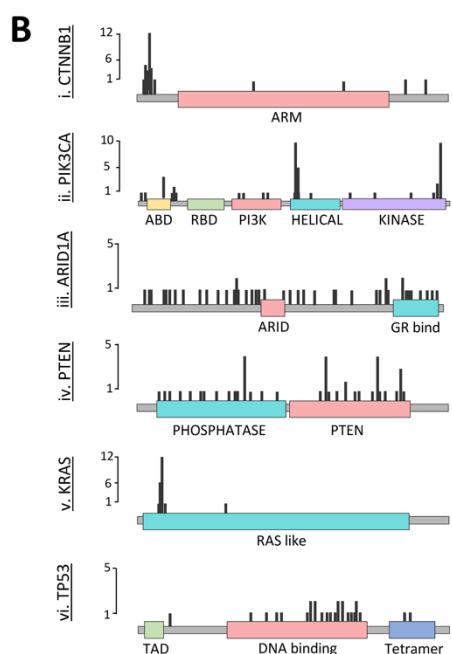
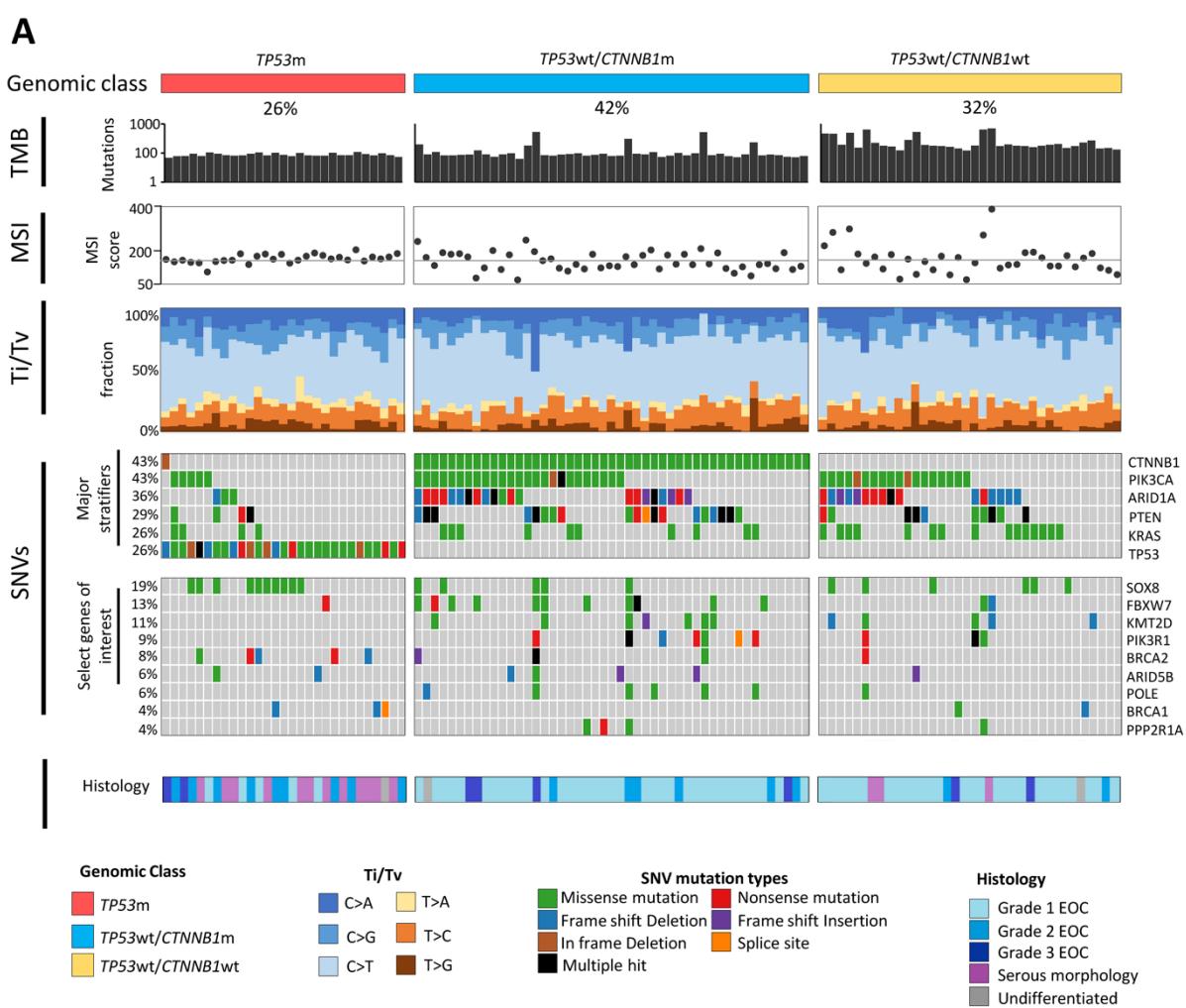
<sup>b</sup>as documented on diagnostic pathology report or identified from reviewed archival tissue



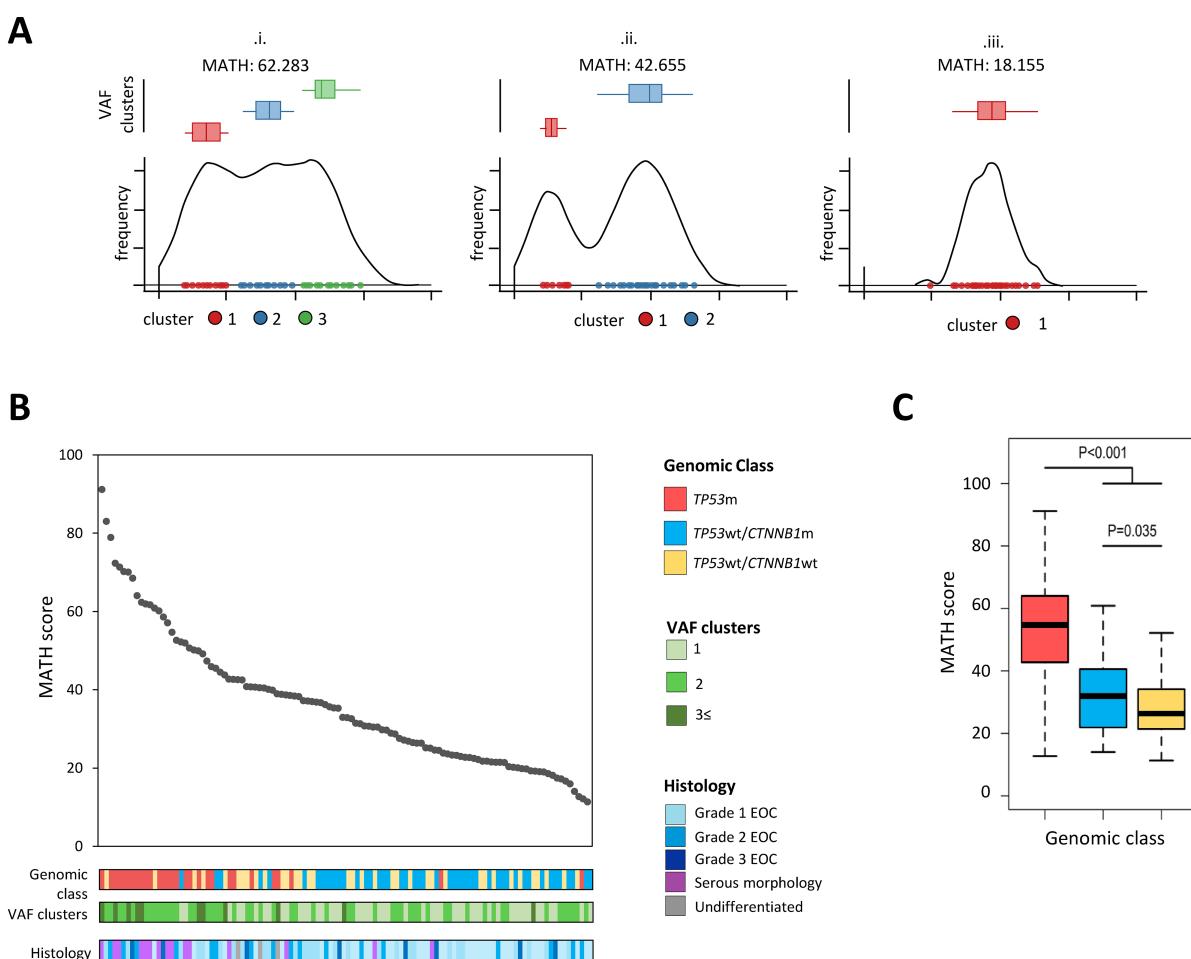
**Figure 1:** CONSORT diagram demonstrating identification of EnOC cases for whole exome sequencing. IHC, immunohistochemistry; WES, whole exome sequencing; G1, grade I; G2, grade II; G3, grade III; ca, carcinoma



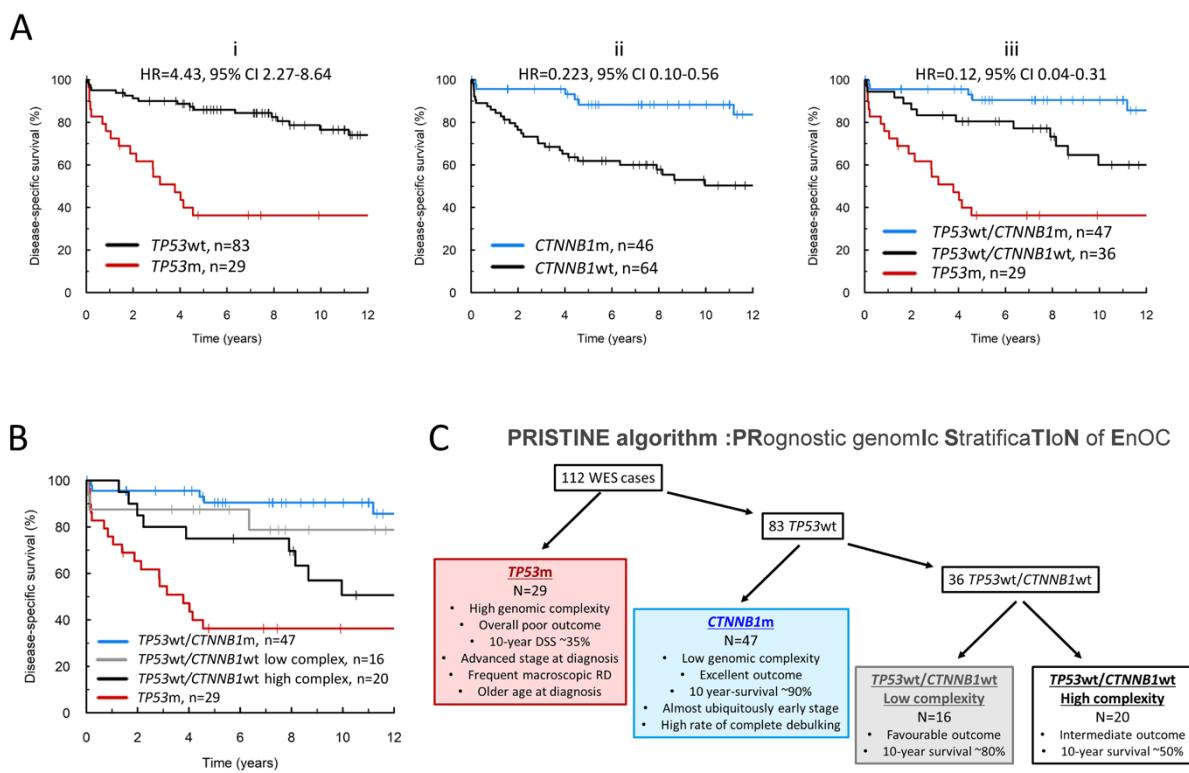
**Figure 2:** Unsupervised hierarchical clustering of endometrioid ovarian carcinomas using Pearson's correlation scores based on the mutational status of the most frequently mutated genes. Bars denote major stratifier genes, *CTNNB1* and *TP53*.



**Figure 3:** Genomic characterisation of endometrioid ovarian carcinoma. **A.** Whole exome sequencing identifies three major genomic subtypes of endometrioid ovarian carcinoma based on SNV status. TMB is plotted as number of SNVs per sample. Microsatellite instability (MSI) scores are plotted based on INDEL counts and colour coded based on mutations across mismatch repair genes. Molecular signatures for each tumour are plotted as the fraction of transversions and transitions (Ti/Tv). SNVs are displayed as an oncoplot. Grey denotes no mutation. Percentages on the left indicate % of samples with a given mutation. Upper plot shows the six most frequently mutated genes which act as major stratifiers, lower plot shows select genes of interest. **B.** Lollipop plots of five common gene targets of mutation. **C.** MMR mutations across cases. **D.** Relationship between TMB and mutations in *POLE* and/or MMR genes.



**Figure 4.** Genomic complexity of endometrioid ovarian carcinomas. **A.** Example density plots for (i) complex, (ii) intermediate and (iii) low complexity tumours. Clusters of VAF scores are shown and summarized within box plots above the density plot. **B.** plot of MATH scores across all 112 tumours, ranked high to low. Bars below show genomic class, number of discrete VAF peaks and histologically defined tumour grade. **C.** Box plot of MATH scores across the three genomic classes.



**Figure 5.** Disease-specific survival of molecularly-defined endometrioid ovarian carcinoma (EnOC) subtypes. (A) Disease-specific survival based on *TP53m* (i) , *CTNNB1m* (ii) or combined mutational status (iii). (B) Disease-specific survival based on the PRISTINE algorithm for EnOC classification. (C) Summary of the PRISTINE algorithm for molecular subtyping in EnOC.