Optimised ARID1A immunohistochemistry is an accurate predictor of *ARID1A* mutational status in gynaecological cancers

Saira Khalique[1,2], Kalnisha Naidoo[1], Ayoma D. Attygalle[3, 4], Divya Kriplani[1], Frances Daley[1], Anne Lowe[1], James Campbell[5], Thomas Jones[6], Michael Hubank[6], Kerry Fenwick[7], Nicholas Matthews[7], Alistair G. Rust[7], Christopher J. Lord[1,2,8], Susana Banerjee[3,9*] and Rachael Natrajan[1,2*]

[1] The Breast Cancer Now Toby Robins Research Centre, Division of Breast Cancer, The Institute of Cancer Research, London, UK

[2] Division of Molecular Pathology, The Institute of Cancer Research, London, UK

[3] Gynaecology Unit, The Royal Marsden NHS Foundation Trust, London, UK

[4] Department of Histopathology, The Royal Marsden NHS Foundation Trust, London, UK

[5] ICR Core Bioinformatics Facility, The Institute of Cancer Research, Sutton, UK

[6] Molecular Diagnostics Departments, The Centre for Molecular Pathology, The Royal Marsden NHS Foundation Trust, Sutton, UK

[7] Tumour Profiling Unit, The Institute of Cancer Research, London, UK

[8] The CRUK Gene Function Laboratory, The Institute of Cancer Research, London, UK

[9] Division of Clinical Studies , The Institute of Cancer Research, London, UK

Running title: ARID1A immunohistochemistry and mutational status in gyanecological cancers

*Corresponding authors:

Dr Rachael C. Natrajan

The Breast Cancer Now Toby Robins Research Centre, Mary-Jean Mitchell Green

Building, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB,

UK

Email: rachael.natrajan@icr.ac.uk

Dr Susana Banerjee

The Gynaecology Unit , The Royal Marsden NHS Foundation Trust , 203 Fulham

Road, London , SW3 6JJ, UK

Email: Susana.Banerjee@rmh.nhs.uk

Main text word count: 3335

Conflict of interest: The authors declare no conflict of interest

ABSTRACT

ARID1A is a tumour suppressor gene that is frequently mutated in clear cell and endometrioid carcinomas of the ovary and endometrium and is an important clinical biomarker for novel treatment approaches for patients with *ARID1A* defects. However, the accuracy of ARID1A immunohistochemistry (IHC) as a surrogate for mutation status has not fully been established for patient stratification in clinical trials. Here we tested whether ARID1A immunohistochemistry could reliably predict *ARID1A* mutations identified by next-generation sequencing. Three commercially available antibodies - EPR13501 (Abcam), D2A8U (Cell Signaling) and HPA005456 (Sigma) – were optimised for IHC using cell line models and human tissue, and screened across a cohort of 45 rare gynaecological tumours. IHC was scored independently by three pathologists using an immunoreactive score. *ARID1A* mutation status was assessed using 2 independent sequencing platforms and the concordance between *ARID1A* mutation and protein expresssion was evaluated using Receiver Operator Characteristics statistics. Overall, 21 *ARID1A* mutations in 14/43 assessable tumours (33%) were identified, the majority of which were predicted to be deleterious. Mutations were identified in 7/18 (39%) ovarian clear cell carcinomas, (4/7 (57%) of ovarian endometrioid carcinomas, 2/5 (40%) of endometrial carcinomas and 1/6 (17%) of carcinosarcomas. ROC analysis identified greater than 95% concordance between mutation status and IHC using a modified immunoreactive score for all three antibodies allowing a definitive cut-point for *ARID1A* mutant status to be calculated. Comprehensive assessment of concordance of ARID1A IHC and mutation status identified EPR13501, as an optimal antibody, with a 100% concordance between *ARID1A* mutation status and protein expression, across different gynaecological histological subtypes. It delivered the best inter-observer agreement between all pathologists, as well as a a clear cost-benefit advantage. This could allow patients to be accurately stratified based on their ARID1A IHC status into early phase clinical trials.

INTRODUCTION

Recent studies have highlighted that somatic loss of function mutations in the tumour suppressor gene *ARID1A* underpin 57% of ovarian clear cell carcinomas (OCCC)[1, 2], 40% of uterine endometrioid carcinomas[3] and between 20-36% of uterine carcinosarcomas[4-6], however are rare in high-grade serous ovarian cancer (HGSOC)[1, 2]. HGSOC are underpinned by *TP53* mutations, where p53 IHC is routinely used in clinical practice to aid diagnosis[7]. However, OCCC rarely harbour *TP53* mutations; these differences in driver mutational profiles highlight the distinct aetiology of the two diseases. Moreover, the presence of *ARID1A* missense mutations may not alter ARID1A protein expression unlike *TP53* missense mutations. Loss or significant reduction of ARID1A protein expression is associated with heterozygous *ARID1A* mutations[1, 8, 9], suggesting a dominant-negative tumour suppressor role (for a comprehensive review on this see[10]). ARID1A forms a key DNA binding subunit in the ATP dependent SWI/SNF chromatin-remodelling complex that modulates the winding of DNA around histone cores allowing access to the DNA to enable transcription, DNA repair and replication[11, 12], leading to aberrant cell cycle and loss of proliferation control[13]. Two genetically engineered mouse models (GEMM) have been created with *ARID1A* alterations which led to tumour formation with loss of ARID1A protein expression observed by IHC. The first was an ovarian endometrioid tumour with co-existent *PTEN* loss[14] and a more recently a GEMM harbouring both *ARID1A* loss and a *PIK3CA* (H1047R) mutation promoted ovarian clear-cell tumorigenesis[15]. *ARID1A* mutations are also detected at high frequencies in other solid tumours including 23% of pancreatic ductal adenocarcinoma (PDAC), 23% of advanced urothelial carcinomas, 18% hepatocellular carcinomas and 6% of metastatic castrate resistant prostate cancers[16-20].

Novel ways of treating patients with *ARID1A* mutations have focused largely on using synthetic-lethal approaches. Bitler et al highlighted the potential of targeting the antagonistic activity between SWI/SNF and EZH2 methyltransferase, with the EZH2 small molecule inhibitors GSK126, that triggered apoptosis in *ARID1A* mutated cells[21]. This was mediated via upregulation of *PIK31P1*, a direct target of EZH2, and selectivity was further enhanced upon inhibition of PI3K-AKT signalling[21]. Subsequent work by Bitler et al has shown that *ARID1A*-mutated ovarian cancers are selectively dependent on HDAC6 activity, due to HDAC6 upregulation in *ARID1A* mutant cells that mechanistically inactivates the apoptosis-promoting function of TP53 due to deacetylation of lysine 120[22]. The work showed that treating *ARID1A*-mutated tumours with the small molecule HDAC6 inhibitor, ACY1215, showed a survival benefit *in vivo*. Additional approaches identified reliance on the *ARID1A* paralog, *ARID1B* in mutant cells[23]. Our group has found *ARID1A* defective OCCC tumours can be targeted with the multi-kinase inhibitor dasatinib, mediated through addiction to the dasatinib target YES1[9] and more recently a profound sensitivity to inhibition of the DNA repair kinase, ATR, leading to premature mitotic entry, genomic instability and apoptosis[24]. Shen et al have additionally shown that loss of ARID1A leads to an impaired G2-M DNA damage checkpoint activation and repair of DNA double strand breaks (DSB), causing sensitivity to DSB inducing treatments such as the PARP inhibitor talozaparib, both *in vitro* and *in vivo*[25]. Together, these observations open up the possibility of assessing these therapeutic approaches in clinical trials, in particular, for tumour types with high frequencies of *ARID1A* mutations, such as OCCC.

Currently there are no clinical trials recruiting patients that prospectively assess *ARID1A* mutational status. However, there are three early phase trials investigating *ARID1A* mutational status and response to therapy; the first of which allocates treatment to patients whose biopsies are sequenced as part of on-going clinical

sequencing programmes outside of the remit of the clinical trial[26]. Patients with *PIK3CA*, *AKT* or *ARID1A* mutations will receive olaparib with the AKT inhibitor AZD5363[26]. A second is a randomised phase II study of nintedanib (an oral tyrosine kinase inhibitor targeting VEGF receptor 1-3, FGFR 1-3 and PDGFR α and β) compared to chemotherapy in patients with clear cell carcinoma of the ovary or endometrium, which will assess *ARID1A* mutational status retrospectively and correlate with outcome[27]. A third trial which is ongoing, but not currently recruiting, will assess dasatinib in patients with recurrent or persistent ovarian, fallopian tube, endometrial or peritoneal cancer and will retrospectively compare ARID1A mutational and IHC status[28]. These trials highlight that prospectively assessing *ARID1A* mutational status is potentially cost-prohibitive and the turn-around time can make it difficult for trial recruitment. Therefore, a surrogate biomarker of mutational status such as IHC is needed.

Despite the clinical importance of *ARID1A* mutations, rapid sequencing is not widely available hence immunohistochemistry (IHC) would be the commonest method to infer mutational status, highlighted by the fact that a number of open clinical open clinical trials are performing retrospective *ARID1A* mutational assessment. Studies that have so far assessed the correlation between *ARID1A* mutational status and IHC have demonstrated a good concordance[1, 3, 29] however the accuracy of IHC as a predictor of *ARID1A* mutation in ovarian carcinoma has not been precisely defined as there is no uniform scoring system or specific antibody that is recommended for IHC clinical use. In the largest concordance study to date, Wiegand et al assessed ARID1A in 182 gynaecological tumours, defining positive staining as definitive nuclear staining and negative as no immunoreactivity[1]. There was a statistically significant correlation between the loss of ARID1A protein expression and *ARID1A* mutational status in both OCCC and endometrioid carcinomas[1]. In total, 73% of OCCC (27/37 cases) with a known *ARID1A* mutation

showed loss of ARID1A expression, however 11% of OCCC (4/36 cases) with no *ARID1A* mutation also showed loss of ARID1A protein expression[1]. Of note this study used the mouse clone 3H2 (Abgent, CA), which targets a region of 111 amino acids (aa 1216 to 1326) but is no longer commercially available. Patients with recurrent or metastatic disease who will be entering the growing number of clinical trials where a robust assessment of ARID1A protein expression in the tumour could be informative will require a test that is accurate, reproducible with a fast turn-around time.

Here our aims were to develop ARID1A IHC as a surrogate predictive biomarker for diagnostic assessment of *ARID1A* mutational status in gynaecological tumours. In particular, we sought to compare the concordance of a number of commercially available antibodies using a standardised scoring system and identify the most optimal assay for clinical assessment of mutation status.

MATERIALS AND METHODS

Cell lines

Cell lines ES2 and TOV21G were obtained from the American Type Tissue Collection (ATCC). HCT116 isogenic ARID1A (Q456*/Q456*) and parental lines were purchased from Horizon Discovery (Cambridge, UK). These are developed by knock-in of premature stop codon (Q456*). Cell lines were cultured in a humidified 37°C incubator with 5% CO_2. Cell lines were tested to confirm no mycoplasma infection using Mycoalert™ ®Mycoplasma Detection Kit as per manufacturer's instructions (Lonza, Slough UK). Cell line identity was confirmed with short tandem repeat typing using the Promega GenePrint®10 system (Promega, Southampton, UK). Cell pellets were formalin fixed and paraffin-wax embedded (FFPE) for antibody optimisation.

Clinical Samples

All patients gave written consent for the use of material for research purposes and tissue samples were obtained with appropriate ethical approval under the Royal Marsden Hospital (RMH) NHS Foundation Trust study: CCR3705 "Analysis of tumour specimens for biomarkers in gynaecological cancers" (Table 1, Supplementary Table S1). All patient samples were reviewed at RMH and appropriate FFPE tissue blocks were selected from their histology reports. Haematoxylin and Eosin (H&E) sections were reviewed by a pathologist (DK) to confirm appropriate tumour tissue and content. Five thick (eight µm) sections were cut for DNA extraction, with an additional H&E slide and three unstained sections for ARID1A immunohistochemistry (IHC). Whole serial sections were cut from the same diagnostic block to minimise heterogeneity between analysis for the ARID1A IHC and next-generation sequencing. If no germline blood sample was available, then non-malignant FFPE blocks were obtained and sections cut for DNA extraction.

DNA extraction and library preparation

DNA extraction and NGS sequencing took place in GCLP accredited laboratories at The Centre for Molecular Pathology, The Royal Marsden NHS Foundation Trust, Sutton. Genomic DNA from FFPE tissue sections was extracted using QIAamp FFPE Tissue Kit (Qiagen, Manchester, UK) according to the manufacturer's instructions for both tumour and non-malignant content. Genomic DNA from blood was extracted using the QIAamp Blood mini kit (manual) or QIAsymphony DNA Midi Kit (automated) (Qiagen, Manchester, UK) according to the manufacturer's instructions. DNA quality was assessed on the Agilent 2200 Tapestation (Agilent, Stockport, UK) and the Qubit Fluorometer (Fisher Scientific, Loughborough, UK).

Targeted *ARID1A* sequencing

ARID1A mutations were identified using a targeted capture panel (Nimblegen, Roche, Welwyn Garden City, UK), designed to target 59 genes for the FOrMAT

clinical trial (Feasibility of Molecular Characterisation Approach to Treatment, CCR3994, Royal Marsden NHS Hospital, Foundation Trust). The panel typically covers *ARID1A* at 99% >250X (Supplementary Table S2). Individual sample library preparation was performed using the KAPA Biosystem HyperPlus kit using 50-200ng DNA. Pooling and capture steps were performed using the Nimblegen SeqCap Capture Protocol (Roche). KAPA Library Quant Kit Universal qPCR Mix (Illumina), was used to quantify the libraries before combining in equimolar concentrations prior to sequencing on the Illumina MiSeq. A high confidence call covered >95% of a region; medium confidence >85% and a failure if the coverage was <85% coverage. MiSeq Reporter software (v2.5.1) was used to align sequences to version Hg19 of the human genome using aligner: BWA v0.61 and Somatic Caller v3.5.2.1 to call variants alongside in-house web server 1.0 and 2.0 software. Variants were called at a frequency of 5% and present in at least 5 reads, with a minimum read depth of 10 reads. Mutations were manually visualised in the Integrated Genomics Viewer (IGV). All mutations were validated using a custom AmpliSeq panel (Thermo Fisher, Wilmington, USA) and run on the Ion Proton panel (Thermo Fisher, Wilmington, USA) with 10 ng input DNA. Sequencing data was analysed using the Ion Torrent software suite (version 5.2.2). Sequences were aligned to version Hg19 of the human genome and mutations were called using the variantCaller (version v5.2.0.34). For pairs of tumour-normal samples, mutation calls were intersected using vcf-isec (VCFtools version 0.1.10) and filtered to retain somatic mutations. Oncotator and SnpEff (version 3.3h) were used to annotate somatic variants reporting the most deleterious effect. Further annotations were added to each mutation using ANNOVAR and the following databases: ClinVar, COSMIC, dbnsfp33a, clinvar_20170130, ljb23_ma, exac03 and exac03nontcga. Only those mutations that were confirmed with both sequencing platforms were included in the final concordance analysis. Raw targeted sequencing data have been deposited into

the NCBI Sequence Read Archive under the accession PRJNA432413 and PRJNA432343.

ARID1A Immunohistochemistry

IHC was performed on 3-4μm thick whole tissue sections. The slides were incubated with antibodies to: Anti-ARID1A, rabbit monoclonal 1:1000, EPR13501 (Abcam, Cambridge, UK), ARID1A/BAF250A, rabbit monoclonal 1:250, D2A8U (Cell Signaling Technology, Europe, Leiden, The Netherlands) and anti-ARID1A, rabbit polyclonal 1:400, HPA005456 (Sigma-Aldrich, Gillingham, Dorset, UK), using the Dako-Autostainer Link 48 with the EnVision FLEX kit as per manufacturer's instructions (Agilent Technologies, Cheadle, Cheshire, UK). Human breast, prostate and kidney tissue were used as positive controls and xenograft models were obtained as previously described[24].

Cases were independently scored using an immunoreactive scoring system[30] by three pathologists DK (Pathologist 1) KN (Pathologist 2) and AA (Pathologist 3) who were blinded to the sequencing results. Sections were evaluated for both intensity (0=negative, 1= weak staining; 2= moderate; 3= strong) and proportion of cells staining positively expressed as a percentage (0= 0%; 1+ $\leq$10%; 2+= 11-50%; 3+= 51-80%; 4+>80%). The intensity and proportion of cells staining was multiplied to get the final score between 0-12[30]. Stromal cells were used as an internal positive control. The pathologists' scores were averaged to give a combined score used in further analyses. An IHC score "cut-off" for loss of expression was defined in conjunction with the genomic deleterious mutation results for each antibody using ROC statistical analysis. Fleiss kappa statistics were used to assess inter-rater variability[31].

11

RESULTS

ARID1A IHC optimisation shows nuclear immunoreactivity in *ARID1A* wild-type cases and absence in *ARID1A* mutant cases

We first evaluated three ARID1A antibodies based upon their current availability and recent use in the literature[24, 32, 33]; Abcam EPR13501 monoclonal antibody, Cell Signaling D2A8U monoclonal antibody and Sigma HPA005456 polyclonal antibody, hereafter abbreviated to clone details (Figure 1A). Antibodies were optimised on human tissue, HCT116 *ARID1A* isogenic (mutant and wild-type) cell lines, ES2 (*ARID1A* wild-type), TOV21G (*ARID1A* mutant) ovarian clear cell carcinoma cell lines and in HCT116 *ARID1A* isogenic xenograft models (Figure 1B-D). These cell lines are known to either express or not express ARID1A as previously evaluated by western blot (Supplementary Figure S1 and [24]). Antibodies were diluted accordingly to ensure a contrast between mutant and wild-type cell lines. ARID1A immunoreactivity was detected in the nucleus, in both malignant epithelial tumour and stromal cells, which were used as a positive internal control in all samples. Of all three antibodies tested, D2A8U showed the strongest immunoreactivity in the cell line models (Figure 1C), with background staining most visible with HPA005456.

We next evaluated ARID1A protein expression in a cohort of 45 gynaecological cancers with all three antibodies (Figure 2 and Table 1). All 3 antibodies demonstrated ARID1A immunoreactivity and performed well on archival tissue, with the oldest block evaluated was from 2005 (3705-0481) and the most recent was from 2016 (666179) (Table 2 and Supplementary Table S3). One case (3705-0541) was not fixed appropraiately at the time of resection and we were unable to process it for IHC, although we were able to extract good quality DNA. Twenty-four cases scored a maximum immunoreactive score of 12 with EPR13501, compared to 16 with D2A8U and 13 with HPA005456. Four cases scored 0 with EPR13501, 3 cases scored 0

with D2A8U and none scored 0 with HPA005456 (Table 2). The scoring concordance of the antibodies between the pathologists varied, with EPR13501 showing the best inter-rater agreement of 0.78, followed by D2A8U (0.67) and HPA005456 (0.57, Fleiss' kappa statistics[31]) (Supplementary Table S4).

Endometriosis related carcinomas show enrichment of *ARID1A* mutations

ARID1A mutations were next evaluated using a 59-gene targeted DNA capture panel, followed by next generation sequencing with a median depth of 718X for *ARID1A*, (IQR=370X-1060X) (Figure 2, Supplementary Table S2*)*. We were able to extract and sequence good quality DNA from archival FFPE blocks in 44 out of 45 cases (exception: 3705-0051). Twenty-five *ARID1A* mutations were identified in 16 cases, of which 21 mutations were validated in 14 cases, and were spread throughout the gene, consistent with previous studies[1, 2] (Figure 3A, Supplementary Table S5). The frequency of mutations according to histological subtype were 39% (7/18) in OCCC, 57% (4/7) in endometrioid adenocarcinoma of the ovary, 17% (1/6, a pelvic carcinosarcoma case) in carcinosarcomas and 40% in endometrial carcinoma, (2/5, both endometrioid endometrial carcinomas) (Table 2 and Figure 3B), in keeping with reported frequencies in the literature[2, 6, 34]. We identified 9 frameshift mutations, 11 nonsense mutations and 1 mis-sense mutation (Table 2 and Figure 3A and 3C-D). Six cases had more than 1 mutation (Table 2 and Figure 3C). The variant allele frequency (VAF) ranged from 0.07 to 0.59, with a number of the mutations showing a low VAF, suggesting subclonal tumour populations (e.g. case 3705-0464 with 2 mutations, with VAFs of 0.07 and 0.14). Of note, even with a low VAF, protein loss occurred (e.g. case a carcinosarcoma case, 3705-0481 with a frame shift mutation (p.Pro1135fs) with a VAF of 0.19, scores of 0 (EPR13501), 0 (D28AU) and 2 (HPA005456)), (Figure 3C). Two cases (3705-0553,

666179) had mutations that were not validated on the Ion Torrent (Supplementary Table S5) due to coverage issues.

ARID1A immunoreactivity and mutational status show between 96-100% concordance in gynaecological cancers

In order to assess if IHC could be used as a biomarker of mutation status, we next evaluated the concordance of ARID1A IHC with the validated mutations. Using Receiver Operating Characteristics (ROC) curve analysis we were able to define a cut-off for each antibody to reliably identify mutant cases. (Table 3). Concordance between mutated gynaecological carcinoma cases and the averaged IHC scores (Supplementary Table S3) were calculated, including a specific analysis for OCCC cases (Table 3). One out of the six carcinosarcomas, 3705-0500, was an outlier, as although it had two nonsense mutations (pTyr1377Ter, p.Glu1542Ter) at relatively high VAF's (0.4), there was only a slight reduction in protein expression, scoring 5 (EPR13501), 5 (D28AU) and 4 (HPA005456). Both subtypes of serous ovarian cancer, a mesonephric adenocarcinoma case and the small cell carcinomas of the ovary were ARID1A wild-type and showed high protein expression (with scores ranging between 11-12 (EPR13501), 11-12 (D28AU) and 6-12 (HPA005456) (Figure 4). Of all three antibodies tested HPA005456 showed the widest range of scores (Figure 4B). Seven OCCC cases were found to have ARID1A mutations; all of which showed a reduction in immunoreactivity (Figure 5). All OCCC ARID1A wild-type cases had detectable protein expression, with scores between 11-12 with EPR13501.

DISCUSSION

In this study, we assessed the concordance of three commercially available antibodies for the assessment of ARID1A protein expression as a surrogate

14

biomarker of mutation status. Overall, we found that IHC is an excellent surrogate biomarker of loss of function mutational status and were able to establish thresholds with each antibody to reliably identify mutant cases to be used in prospective patient assessment that improves upon published concordance rates.

Concordance rates for all cases were 100% (EPR13501), 100% D2A8U and 97% (HPA005456). The histograms for all antibodies show that overall there is a clear bi-modal distribution in immunoreactive scores between the mutant and wild-type cases, that is best exemplified with EPR13501 and D2A8U, with HPA005456 showing the greatest variation, perhaps due to the polyclonal nature of the antibody, whereas the D2A8U and EPR13501 are monoclonal. For instance, comparing OCCC cases alone, our concordance was 100% with all antibodies, compared to 73% (27/37 samples) as reported by Wiegand et al[1]. Given that the number of OCCC mutant cases we assessed is small due to the rarity of cases and fact that trials combine gynaecological histologies, we propose utilising the cut-offs derived from all gynaecological cases. As all three antibodies had excellent specificity and sensitivity, other factors come into consideration when deciding which antibody to recommend taking forward for potential clinical use. The inter-pathologists scores were most consistent with EPR13501, and the dilution factor of 1:1000 for EPR13501 compared to D2A8U's 1:250, which means that this would be more cost-effective with less batch-to-batch variation. D2A8U's staining was most intense in cell line models, but weaker in human tissue, whereas EPR13501 had the strongest intensity in human tissue out of all three antibodies. We investigated HPA005456 as it had been used in a number of recent papers[33, 35, 36], however, as it is a rabbit polyclonal antibody, it will not be available for investigators once the stock runs out and is therefore not a long term viable option.

One of the limitations of this study is that even though we identified 45 patient samples initially, we were only able to assess 43 cases. However, the rarity of these tumours in general renders larger series challenging. In order to correct for this, we were stringent with our analysis of the sequencing data, including only cases where mutations were identified by both sequencing platforms. Thus, we could be confident with the mutation calls that we made. Overall, we have shown that IHC is generally concordant with mutational status; we did observe some variation in the carcinosarcoma cases, which may, in itself, reflect the heterogeneity of these tumours and the limitations of using an immunoreactive score. For example, in case 3705-0500, this may be explained by the *ARID1A* mutation occurring after the ARID1A epitope sites and hence the detection of the truncated residual protein. We observed one endometrioid ovarian adenocarcinoma with a small area of subclonal loss (Supplementary Figure S2). The patient did not have an identifiable ARID1A mutation and scored a 12 (>80% of tumour staining strongly). How such patients respond to targeted therapy will only be able to be evaluated in the context of a clinical trial. Of note, we did not identify any somatic in-frame deletions. However, given reports that such mutations affect the subcellular distribution of ARID1A[37], it would be interesting to assess this with the optimal antibody.

Although a number of smaller studies have assessed concordance between ARID1A IHC and mutational status in gynaecological carcinomas, the concordance rates are lower than we report here and the antibody details and scoring systems used are not fully reported [29][31-33]. For instance, L'heureux et al analysed archival tissue by IHC and mutational status as part of a 40-patient phase II clinical trial in OCCC with ENMD-2076, an oral multi-targeted kinase inhibitor[38]. They had paired data for 32 samples, with 19 ARID1A mutant cases and 13 wild-type cases, where the concordance of the mutational status with IHC was only 69%. Furthermore, the antibody and specific scoring details and nature of mutations were however not

16

detailed. Guan et al compared ARID1A IHC and mutational status in uterine carcinomas using the Sigma polyclonal HPA005456 antibody (targeting 1266-1370 amino acids), with negativity defined as absence of nuclear staining, and identified that only 50% (5/10) of cases with deleterious mutations showed complete lack of ARID1A expression. Our study only included gynaecological tumours and would need to be extended to other tumour types to ascertain its general applicability.

A recent sequencing study identified cancer-associated inactivating *ARID1A* mutations in deep infiltrating endometriosis, with loss of ARID1A immunoreactivity serving as a surrogate for *ARID1A* inactivating mutations using the HPA005456 antibody[36]. Using the monoclonal EPR13501 antibody, which had the best inter-rater agreement in our study, would allow comprehensive analysis of endometriosis to determine the clinical significance of ARID1A loss and be of potential diagnostic use.

In conclusion, we have systematically assessed a number of commercially available antibodies and identified EPR13501 as a robust biomarker of *ARID1A* status with a cut-off of <8 using our optimised scoring system. This will be useful for recruiting patients for clinical trials based on *ARID1A* mutational status. An international academic trial of ATR inhibition in combination with a PARP inhibitor in ARID1A-stratified gynaecological cancers that utilises our findings is planned to open in 2018 using this approach, allowing validation and evaluation of the IHC scoring system in the context of a prospective clinical trial.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

RN, CJL and SB conceived and designed the study. SK, KN, AA, DK, FD, AL, TJ, MH, KF and NM carried out the experiments and undertook IHC scoring. JC, TJ, MH, AGR and RN performed the bioinformatics analysis. SK, KN, AA, DK, FD, CJL, SB discussed and interpreted the results. SK, CJL, SB and RN wrote the first draft. All authors read and approved the final manuscript.

LIST OF ONLINE SUPPORTING MATERIAL

Supplementary Figure S1: ARID1A status of OCCC cell lines. Western blot of ARID1A protein expression in ES2 (OCCC *ARID1A* wild-type cell line) and TOV21G (OCCC *ARID1A* mutant cell line). ARID1A (250 kDa fragment), loading control β-Actin (42 kDa).

Supplementary Figure S2: Subclonal expression in a Grade 2 endometrioid adenocarcinoma of the ovary case (3705-0482). Images of area of subclonal expression with IHC scores and sequencing results. Top Row: Low power magnification, arrow shows area of negative protein expression surrounded by area of positive expression with all 3 antibodies (scale bar 1mm). Bottom Row: High power magnification, arrow shows area of negative protein expression surrounded by area of positive expression with all 3 antibodies (scale bar 100µm).

Supplementary Table S1: Extended table of clinicopathlogical features of patient cohort

Supplementary Table S2: Summary of *ARID1A* sequencing metrics from targeted capture panel

Supplementary table S3: Summary of individual and combined scores for ARID1A IHC

Supplementary Table S4: Inter-pathologist concordance metrics

Supplementary Table S5: Summary of *ARID1A* mutations identified and validation

REFERENCES

1. Wiegand KC, Shah SP, Al-Agha OM*, et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010; **363**: 1532-1543.
2. Jones S, Wang TL, Shih Ie M*, et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* 2010; **330**: 228-231.
3. Guan B, Mao TL, Panuganti PK*, et al.* Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *Am J Surg Pathol* 2011; **35**: 625-632.
4. Cerami E, Gao J, Dogrusoz U*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; **2**: 401-404.
5. Hembree TN, Teer JK, Hakam A*, et al.* Genetic Investigation of Uterine Carcinosarcoma: Case Report and Cohort Analysis. *Cancer Control* 2016; **23**: 61-66.
6. Jones S, Stransky N, McCord CL*, et al.* Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes. *Nat Commun* 2014; **5**: 5006.
7. Kobel M, Piskorz AM, Lee S*, et al.* Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. *J Pathol Clin Res* 2016; **2**: 247-258.
8. Anglesio MS, Wiegand KC, Melnyk N*, et al.* Type-specific cell line models for type-specific ovarian cancer research. *PLoS One* 2013; **8**: e72162.
9. Miller RE, Brough R, Bajrami I*, et al.* Synthetic Lethal Targeting of ARID1A-Mutant Ovarian Clear Cell Tumors with Dasatinib. *Mol Cancer Ther* 2016; **15**: 1472-1484.
10. Wu RC, Wang TL, Shih Ie M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* 2014; **15**: 655-664.
11. Kwon H, Imbalzano AN, Khavari PA*, et al.* Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* 1994; **370**: 477-481.
12. Imbalzano AN, Kwon H, Green MR*, et al.* Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 1994; **370**: 481-485.

13. Nagl NG, Jr., Patsialou A, Haines DS, *et al.* The p270 (ARID1A/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest. *Cancer Res* 2005; **65**: 9236-9244.

14. Guan B, Rahmanto YS, Wu RC, *et al.* Roles of deletion of Arid1a, a tumor suppressor, in mouse ovarian tumorigenesis. *J Natl Cancer Inst* 2014; **106**.

15. Chandler RL, Damrauer JS, Raab JR, *et al.* Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun* 2015; **6**: 6118.

16. Takai E, Totoki Y, Nakamura H, *et al.* Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. *Sci Rep* 2015; **5**: 18425.

17. Petros Grivas. et al. Circulating tumour (ct)-DNA alterations in advanced urothelial carcinoma; Association with outcomes and evolution with therapy. *ASCO Genitourinary Cancers Symposium*. Orlando: J Clin Oncol 2017:abstract 334.

18. Rebecca J Nagy. et al. Circulating cell-free DNA profiling of patients with advanced urothelial carcinoma of the bladder. *ASCO Annual Meeting*. Chicago: J Clin Oncol 2016.

19. Guru Sonpavde. et al. Circulating tumor (ct)-DNA alterations in metastatic castration-resistant prostate cancer (mCRPC): Association with outcomes and evolution with therapy. *ASCO Genitourinary Cancers Symposium*. Florida: J Clin Oncol 2017.

20. Sadakatsu Ikeda. et al. Analysis of tissue and circulating tumor DNA by next-generation sequencing of hepatocellular carcinoma. *ASCO Gastrointestinal Cancers Symposium*. San Francisco: J Clin Oncol 2017.

21. Bitler BG, Aird KM, Garipov A, *et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat Med* 2015; **21**: 231-238.

22. Bitler BG, Wu S, Park PH, *et al.* ARID1A-mutated ovarian cancers depend on HDAC6 activity. *Nat Cell Biol* 2017; **19**: 962-973.

23. Helming KC, Wang X, Wilson BG, *et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat Med* 2014; **20**: 251-254.

24. Williamson CT, Miller R, Pemberton HN, *et al.* ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. *Nat Commun* 2016; **7**: 13837.

25. Shen J, Peng Y, Wei L, *et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov* 2015; **5**: 752-767.

26. OLAParib COmbinations (OLAPCO) (NCT02576444). https://clinicaltrials.gov/ct2/show/NCT02576444 2017.

27. A Randomised Phase II Study of Nintedanib (BIBF1120) Compared to Chemotherapy in Patients with Recurrent Clear Cell Carcinoma of the Ovary or Endometrium (EudraCT Number: 2013-002109-73). https://www.clinicaltrialsregister.eu/ctr-search/search?query=2013-002109-73 2016.

28. Dasatinib in Treating Patients With Recurrent or Persistent Ovarian, Fallopian Tube, Endometrial or Peritoneal Cancer (NCT02059265). https://clinicaltrials.gov/show/NCT02059265 2017.

29. Maeda D, Mao TL, Fukayama M, *et al.* Clinicopathological significance of loss of ARID1A immunoreactivity in ovarian clear cell carcinoma. *Int J Mol Sci* 2010; **11**: 5120-5128.

30. Samartzis EP, Samartzis N, Noske A, *et al.* Loss of ARID1A/BAF250a-expression in endometriosis: a biomarker for risk of carcinogenic transformation? *Mod Pathol* 2012; **25**: 885-892.

31. J.J R. Online Kappa Calculator (Computer Software). 2008 [cited 2017 4th Sepetember]; Available from: http://justus.randolph.name/kappa

32. Gibson WJ, Hoivik EA, Halle MK, *et al.* The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat Genet* 2016; **48**: 848-855.

33. Coatham M, Li X, Karnezis AN, *et al.* Concurrent ARID1A and ARID1B inactivation in endometrial and ovarian dedifferentiated carcinomas. *Mod Pathol* 2016; **29**: 1586-1593.

34. Cancer Genome Atlas Research N, Kandoth C, Schultz N, *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497**: 67-73.

35. Guan B, Mao T-L, Panuganti PK, *et al.* Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *The American journal of surgical pathology* 2011; **35**: 625-632.

36. Anglesio MS, Papadopoulos N, Ayhan A, *et al.* Cancer-Associated Mutations in Endometriosis without Cancer. *N Engl J Med* 2017; **376**: 1835-1848.

37. Guan B, Gao M, Wu CH, *et al.* Functional analysis of in-frame indel ARID1A mutations reveals new regulatory mechanisms of its tumor suppressor functions. *Neoplasia* 2012; **14**: 986-993.

38. Stephanie Lheureux. et al. Phase II clinical and moilecular trial of oral ENMD-2076 in clear cell ovarian cancer (CCOC), A Study of the Princess Margaret Phase II Consortium. *ASCO Annual Meeting*. Chicago: J Clin Oncol 2017.

Figure legends

Figure 1. ARID1A antibody optimisation in tumour cell lines and tumour xenograft models

(A) Schematic of ARID1A protein illustrating epitope regions raised against all 3 antibodies studied. The lollipop plot shows mutation loci of two cell lines used in antibody optimisation: HCT116 *ARID1A* isogenic mutant cell line Q456*/Q456*, nonsense mutation (depicted in orange) and TOV21G, a compound heterogeneous cell line with mutation loci: TOV21G p.548fs and p.756fs, frameshift mutations (depicted in mauve). (B) Immunoreactivity of ARID1A detected by all three antibodies in HCT116 *ARID1A* isogenic cell lines embedded in FFPE blocks. HCT116 -/- (*ARID1A* mutant) shows loss of ARID1A immunoreactivity whereas nuclear immunoreactivity was preserved in the HCT116 +/+ (*ARID1A* wild-type) cell line. EPR13501, rabbit monoclonal, antigen retrieval using microwave and dilution 1:1000. D2A8U, rabbit monoclonal, antigen retrieval with pre-treatment module and dilution 1:250. HPA005456, rabbit polyclonal, antigen retrieval with pre-treatment

module and dilution 1:400. Scale bar is equal to 100µm. (C) ARID1A immunoreactivity in ovarian clear cell carcinoma cell lines and HCT116 +/+ and -/- cells engrafted as xenografts was detected with all three antibodies in ovarian clear cell carcinoma cell lines. TOV21G (*ARID1A* mutant) shows loss of ARID1A immunoreactivity whereas immunoreactivity was preserved in the ES2 (*ARID1A* wild-type) cell line. Scale bar is equal to 100µm. **D.** ARID1A immunoreactivity was detected with all three antibodies in xenograft models of HCT116 +/+ and HCT116 -/-. The background staining seen in cell lines was reduced in xenografts. Scale bar is equal to 100µm.

Figure 2. Study workflow

Modified CONSORT diagram showing the processing of the 45 gynaecological cancer cases identified. After histopathology review a representative block was chosen. The same FFPE block was used for extracting DNA for next-generation sequencing (NGS) and immunohistochemistry (IHC). One case failed NGS quality control and one case was not suitable for IHC analysis. Three histopathologists independently reviewed and scored cases. These were then averaged and compared to mutational status to establish concordance (n=43).

Figure 3. Distribution of *ARID1A* mutations identified by targeted sequencing in gynaecological cancers.

(A) A lollipop plot showing the distribution and classes of validated loss of function mutations in ARID1A detected by both sequencing platforms in 44 cases. Twenty-one *ARID1A* mutations were identified in 14 patients. The majority were frameshift mutations (9, in mauve), with 11 nonsense mutations (orange) and 1 missense mutation (green). (B) Bar chart showing the frequency of *ARID1A* mutations according to histology. An enrichment of *ARID1A* mutations was seen in

endometriosis-related tumours and carcinosarcomas; with 57% of ovarian endometroid (light blue (and 39% of OCCC (dark blue) cases having an *ARID1A* mutation. (C) A variant allele frequency plot shows each case with the type of validated mutation (colour of circle) and coverage (size of circle) with corresponding protein expression on IHC (boxes below). Six cases had more than 1 mutation in *ARID1A*. Case 3705-0541 was not suitable for IHC processing.

Figure 4. ARIDIA immunohistochemistry shows good concordance with mutational analysis for all three antibodies in a variety of gynaecological cancers

(A) ARID1A immunoreactivity in grade 2 endometrioid adenocarcinoma of the ovary. Case 3705-0460 with 1 mutation (p.Arg1505Ter), shows a lack of tumour cell staining with positive stromal staining (immunoreactive score with EPR13501: 1, D2A8U: 1 and HPA005456: 3). Case 3705-0529, *ARID1A* wild-type, shows positive nuclear staining with an immunoreactive score with EPR13501: 12, D2A8U: 12 and HPA005456: 9. Scale bar is equal to 100μm. (B) Histograms showing the distribution of all immunoreactive scores (n=37) with all 3 antibodies, annotated with mutational status. The majority of *ARID1A* mutant cases (red) show low immunoreactivity scores and *ARID1A* wild-type cases (blue) show high immunoreactivity scores. There is greatest variation in the HPA005456 scores.

Figure 5. ARIDIA immunohistochemistry shows good concordance with mutational analysis for all three antibodies in OCCC

ARID1A immunoreactivity in OCCC. Case 3705-0416 with 2 mutations (p.Gly191fs, p.Tyr485Ter), shows a lack of tumour cell staining with positive stromal staining (immunoreactive score with EPR13501: 0, D2A8U: 1 and HPA005456: 2). Case 3705-0514, *ARID1A* wild-type, shows positive nuclear staining with an

immunoreactive score with EPR13501: 12, D2A8U: 11 and HPA005456: 12. Scale bar is equal to 100µm. (B) Histogram showing immunoreactivity scores for OCCC cases shows a bimodal distribution between mutant cases that lead to loss of immunoreactivity and *ARID1A* wild-type cases

Tables

Table 1: Overview of patient characteristics in the study.

Patients were aged between 21 and 76 years of age, and comprised 8 gynaecological subtypes, the most frequent being clear cell carcinoma of the ovary, n=18.

Parameter	Total number (n)
Number of tumour samples	45
Diagnosis	
Carcinosarcoma (CS)	6
Endometrial Clear Cell (ECC)	1
Endometrioid Adenocarcinoma of the Endometrium (EAE)	4
Endometrioid Ovarian Carcinoma (ENOC)	7
Low Grade Serous Ovarian Carcinoma (LGSOC)	3
High Grade Serous Ovarian Carcinoma (HGSOC)	1
Small Cell Carcinoma of the Ovary, Hypercalcaemic Type (SCCOHT)	2
Mesonephric Adenocarcinoma of the Ovary (MAO)	1
Mesonephric Adenocarcinoma of the Ovary and Endometrium (MAOE)	1
Serous Ovarian Carcinoma/Carcinosarcoma (SOC/CS)	1
Ovarian Clear Cell Carcinoma (OCCC)	18
Grade	
I	6
II	8
III	31
FIGO stage	
I	14
II	14
III	13
IV	4
Median age, y (range)	57, (21-76)
Endometriosis	
Yes	16
No	28
N/A	1
Primary Specimen	
Yes	42
No	3

Table 2

Summary of averaged pathologist IHC scores and validated sequencing results for ARID1A. *Matched patient lung metastasis, ns= not specified, CS=carcinosarcoma of the ovary, ECC=Endometrial Clear Cell, EAE=Endometroid Adenocarcinoma of the Endometrium, ENOC=Endometrioid Ovarian Carcinoma, LGOSC=Low Grade Serous Ovarian Carcinoma, HGSOC= High Grade Serous Ovarian Carcinoma, SCCOHT=Small Cell Carcinoma of the Ovary, Hypercalcaemic Type, MAO=Mesonephric Adenocarcinoma of the Ovary, MAOE=Mesonephric Adenocarcinoma of the Ovary and Endometrium, UDC=Undifferentiated Component, SOC/CS=Serous Ovarian Carcinoma/Carcinosarcoma and OCCC=Ovarian Clear Cell Carcinoma.

Patient Identifier	Diagnosis	IHC ARID1A antibody EPR13501 average score (Pathologists 1,2, and 3)	IHC ARID1A antibody D2A8U average score (Pathologists 1,2, and 3)	IHC ARID1A antibody HPA005456 average score (Pathologists 1,2, and 3)	IHC: ARID1A protein loss	Sequencing: *ARID1A* mutation and protein change
3705-0449	CS	12	11	10	no	no
3705-0456	CS	10	11	9	no	no
3705-0470	CS	12	8	10	no	no
3705-0484	CS	12	12	12	no	no
3705-0500	CS	5	5	4	yes	p.Tyr1377Ter, p.Glu1542Ter
3705-0510	CS	12	12	8	no	no
666179	ECC	12	11	8	no	no
3705-0553	EAE (UDC)	9	6	7	no	no
3705-0341	EAO	12	11	9	no	no
3705-0481	EAO	0	0	2	yes	p.Pro1135fs
3705-0541	EAO	NA	NA	NA	NA	p.Gln428Ter
3705-0323	EAO	3	3	6	yes	p.Gln512Ter, p.Arg1899Ter
3705-0460	EAO	1	1	3	yes	p.Arg1505Ter
3705-0482	EAO	12	12	12	no	no
3705-0529	EAO	12	12	9	no	no
3705-0199	EAE	3	2	1	yes	p.Ile1975fs, p.Gln799fs
3705-	EAE	1	1	1	yes	p.Arg1504Ter

0438						
3705-0548	EAE	12	11	12	no	no
3705-0475	HGSOC	12	12	12	no	no
3705-0487	LGSOC	11	12	11	no	no
3705-0493	LGSOC	12	12	12	no	no
3705-0525	LGSOC	11	11	6	no	no
3705-0051	MAO	12	11	12	no	NA failed sequencing QC
3705-0308	MAOE	12	12	7	no	no
3705-0442	SOC/CS	12	12	10	no	no
3705-0142	OCCC	7	1	4	yes	p.Pro7fs
3705-0145	OCCC	0	3	2	yes	p.Ser614Ter
3705-0207	OCCC	12	10	8	no	no
3705-0346	OCCC	2	1	3	yes	p.Ala42fs, p.GLn723Ter
3705-0379	OCCC	12	11	12	no	no
3705-0383	OCCC	12	11	11	no	no
3705-0416	OCCC	0	1	2	yes	p.Gly191fs, p.Tyr485Ter
3705-0435	OCCC	11	12	8	no	no
3705-0453	OCCC	12	12	9	no	no
3705-0464	OCCC	1	1	1	yes	p.Gly1848fs, p.Ile2275Ser, p.Gln2070Ter

3705-0468	OCCC	12	11	12	no	no
3705-0497	OCCC	12	12	12	no	no
3705-0514	OCCC	12	11	12	no	no
3705-0540	OCCC	1	0	1	yes	p.Gln932fs
3705-0544	OCCC	12	12	12	no	no
3705-0545	OCCC	0	0	3	yes	p.Gly319fs, p.Leu1100fs
3705-0558	OCCC	11	11	11	no	no
3705-0435*	OCCC	12	12	12	no	no
3705-0466	SCCOHT	12	12	12	no	no
3705-0551	SCCOHT	12	12	12	no	no

Table 3: ROC curve analysis to define mutant immunoreactive score and concordance

Antibody	All gynaecological cases: ARID1A mutant score	Concordance (%)	OCCC cases ARID1A mutant score	Concordance (%)
EPR13501	<8	100	<9	100
	100% sensitivity and 100% specificity		100% sensitivity and 100% specificity	
D2A8U	<5.5	100	<6	100
	100% sensitivity and 100% specificity		100% sensitivity and 100% specificity	
HPA005456	<6.5	97	<6.5	100
	100% sensitivity and 96% specificity		100% sensitivity and 100% specificity	