Molecular subclasses of clear cell ovarian carcinoma and their impact on disease behavior and outcomes

Molecular subclasses of clear cell ovarian carcinoma and their impact on disease behavior and outcomes

Kelly L Bolton[1]*, Denise Chen[2]*, Rosario Corona de la Fuente[3]*, Zhuxuan Fu[4]*, Rajmohan Murali[5]*, Martin Köbel[6*], Yanis Tazi[5]*, Julie M. Cunningham[7]*, Irenaeus C.C. Chan[1], Brian J. Wiley[1], Lea A. Moukarzel[5*], Stacey J. Winham[7], Sebastian M. Armasu[7], Jenny Lester[8], Esther Elishaev[4], Angela Laslavic[4], Catherine J. Kennedy[9,10], Anna Piskorz[11], Magdalena Sekowska[11], Alison H. Brand[9,12], Yoke-Eng Chiew[9,10], Paul Pharoah[11], Kevin M. Elias[13], Ronny Drapkin[14], Michael Churchman[15], Charlie Gourley[15], Anna DeFazio[9,10,12,16], Beth Karlan[8], James D. Brenton[11], Britta Weigelt[5], Michael S. Anglesio[17], David Huntsman[17], Simon Gayther[3]*, Jason Konner[5]*, Francesmary Modugno[4]*, Kate Lawrenson[3]*, Ellen L. Goode[7]*, Elli Papaemmanuil[5]*

[1]Washington University School of Medicine, St Louis MO, USA; [2]Philadelphia College of Osteopathic Medicine, Philadelphia, PA; [3]Cedars-Sinai Medical Center, Los Angeles, CA, USA; [4]University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; [5]Memorial Sloan-Kettering Cancer Center, New York, NY, USA; [6]The University of Calgary, Calgary, Canada; [7]Mayo Clinic, Rochester, MN, USA; [8]David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California at Los Angeles, Los Angeles, CA, USA; [9]Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia; [10]Centre for Cancer Research, The Westmead Institute for Medical Research, Sydney, New South Wales, Australia; [11]University of Cambridge, Cambridge, UK; [12] The University of Sydney, Sydney, New South Wales, Australia; [13]Brigham and Women's Hospital, Boston, MA, USA; [14]University of Pennsylvania, Philadelphia, PA, USA; [15]University of Edinburgh, Edinburgh, UK; [16]The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW, Sydney, New South Wales, Australia; [17]University of British Columbia, Vancouver, BC, Canada.

Running Title: Molecular subclasses of clear cell ovarian carcinoma

*Denotes equal contribution

Correspondence to be addressed to:
Kelly L. Bolton
bolton@wustl.edu
Washington University, St Louis MO

Conflict of Interest

KLB receives grant funding from Bristol Myers Squibb and Servier, outside the scope of this study. BW reports ad hoc membership of the advisory board of Repare Therapeutics, outside the scope of this study. ADeF has received grant funding and honoraria from AstraZeneca, outside the scope of this study.

Statement of translational relevance

Clear cell ovarian cancer (CCOC) is the second most common subtype of epithelial ovarian cancer and when diagnosed at an advanced stage has a poor prognosis. The relationship between molecular profiles and clinical presentation or outcomes are still unknown but could help guide the development of personalized therapeutic approaches for CCOC. Here we profiled 421 primary CCOCs using deep targeted sequencing and whole transcriptome sequencing on a subset of 211. Clustering of cancer driver mutations and RNA expression converged upon two distinct subclasses of CCOC. The first was dominated by *ARID1A*-mutated tumors with enriched expression of canonical CCOC genes and markers of platinum resistance; the second was largely comprised of tumors with *TP53*-mutations and enriched for the expression of genes involved in extracellular matrix organization and mesenchymal differentiation. These two distinct molecular subclasses showed distinct clinical presentation and outcomes, with potential relevance to therapeutic responsiveness.

Abstract

Purpose: To identify molecular subclasses of clear cell ovarian carcinoma (CCOC) and assess their impact on clinical presentation and outcomes.

Experimental Design: We profiled 421 primary CCOCs that passed quality control using a targeted deep sequencing panel of 163 putative CCOC driver genes and whole transcriptome sequencing of 211 of these tumors. Molecularly-defined subgroups were identified and tested for association with clinical characteristics and overall survival.

Results: We detected a putative somatic driver mutation in at least one candidate gene in 95% (401 out of 421) of CCOC tumors including: *ARID1A* (in 49% of tumors), *PIK3CA* (49%), *TERT* (20%) and *TP53* (16%). Clustering of cancer driver mutations and RNA expression converged upon two distinct subclasses of CCOC. The first was dominated by *ARID1A*-mutated tumors with enriched expression of canonical CCOC genes and markers of platinum resistance; the second was largely comprised of tumors with *TP53*-mutations and enriched for the expression of genes involved in extracellular matrix organization and mesenchymal differentiation. Compared to the *ARID1A*-mutated group, women with *TP53*-mutated tumors were more likely to have advanced stage disease, no antecedent history of endometriosis, and poorer survival, driven by their advanced stage at presentation. In women with *ARID1A*-mutated tumors, there was a trend towards lower response rate to first-line platinum-based therapy.

Conclusions: Our study suggests that CCOC consists of two distinct molecular subclasses with distinct clinical presentation and outcomes, with potential relevance to both traditional and experimental therapy responsiveness.

Introduction

Historically, tumor treatment approaches have been dictated by tissue site; but large-scale molecular profiling efforts have shown that remarkable heterogeneity exists in the landscape of cancer driver genes and pathways within tumor types and even within histologic subtypes. This has been well characterized for many common tumors through multi-omic profiling[1] and characterization of the genetic determinants of tumor behavior and outcome has led to the development of personalized therapeutic approaches. Indeed, for some cancers, prognosis and therapeutic strategies are based primarily on their presence of genetic driver mutations identified in the tumor[2–7]. For several rare cancer types such as ovarian clear cell carcinoma (CCOC), no strong associations between molecular profiles and clinical presentation or outcomes are known and broad-acting platinum-based chemotherapy remains the standard of care.

When diagnosed at an advanced stage, CCOC has a worse outcome than other invasive ovarian cancers including the more common high-grade serous ovarian carcinoma (HGSOC) (median overall survival of 10 months)[8,9], presents at a younger age[10], and is less responsive to platinum-based therapy[11]. Relatively small studies suggest that CCOC possesses several driver events that are distinct from HGSOC. CCOC is thought to arise from endometriotic lesions with recurrent somatic mutations in *PIK3CA* and *ARID1A*, which are rare in HGSOC[12–15]. In addition, the existing data suggests that CCOCs are commonly *TP53*-wild-type (whereas HGSOC ubiquitously harbors TP53 mutations) and exhibits fewer structural rearrangements than HGSOC[13]. However, it is not known whether clinically meaningful molecular subtypes of CCOC exist.

In the current study, we performed comprehensive targeted sequencing and transcriptomic profiling of a large, multi-ethnic cohort of 421 primary CCOCs to identify disease subclasses with distinct biology and clinical behavior, which in turn may provide avenues for personalized therapeutic approaches.

Materials and Methods

Study Participants

Clinical data and therapy-naïve fresh frozen tumor material were utilized from women diagnosed with invasive CCOC and enrolled into research studies from the following sites: Memorial Sloan Kettering Cancer Center Gynecology Tissue Bank (MSK; New York NY, USA), Mayo Clinic (MAY; Rochester MN, USA), Addenbrooks Hospital (ADD; England), Cedars-Sinai Medical Center (WCP; Los Angeles CA, USA), University of Pittsburgh (PIT; Pittsburgh PA, USA), Gynaecological Oncology Biobank (GynBiobank) at Westmead Hospital (WMH; Sydney, Australia), University of Edinburgh (SCOT; Scotland), Canadian Ovarian Experimental Unified Resource (COEUR; Multiple sites, Canada), Brigham and Women's Hospital (BWH; Boston MA, USA), and University of Pennsylvania (UPA; Philadelphia PA, USA). Participants provided written informed consent. The studies were conducted in accordance with recognized ethical guidelines (e.g., Declaration of Helsinki, CIOMS, Belmont Report, U.S. Common Rule), and approved by local institutional review boards. Extraction of DNA/RNA was performed centrally at MSK (for cases from MSK, WCP, PIT, BWH and UPA) or locally (for cases from MAY, ADD, WMH and COEUR). For the cases which were extracted centrally at MSK, slides from frozen tissue sections were reviewed by a pathologist (R.M) and extraction of DNA/RNA was performed from tumor sections, selected based on high content (>80%) of clear cell carcinoma. In total, tumors from 447 women diagnosed with CCOC were analyzed. Race and menstruation status (pre vs. post-menopausal) was obtained through participant self-report. History of endometriosis was also obtained through self-report except at MSK where endometriosis was only available if mentioned on the pathology report. Tumor characteristics and clinical outcomes were obtained through medical record review.

Targeted DNA Sequencing and Analysis.

We performed targeted sequencing of 163 putative CCOC driver genes (Supplementary Table 1) in DNA samples from the 447 tumor and blood-derived DNA from 16 unmatched controls using a custom Nimblegen capture-based panel. Genes were selected based on a combined analysis of 105 clear cell somatic sequencing studies including: (1) whole genome sequencing of 31 CCOCs from Wang et al.[13];

(2) whole exome sequencing of eight cases from Jones et al[12]; (3) targeted sequencing of 26 CCOCs using a panel of 465 known cancer drivers (MSK-IMPACT)[16]; and targeted or whole exome sequencing of 40 CCOCs from project GENIE[17] . Included in our panel were 119 genes where somatic mutations have been identified in two or more CCOCs; 41 established cancer driver genes based on the COSMIC Cancer Gene Census[18] mutated in one CCOC and three genes in the SWI/SNF complex (*SMARCB1*, *SMARCC1*, *SMARCC2*)[14] that have been implicated in CCOC biology[19] (Supplementary Table 1). We also included on the sequencing panel highly polymorphic single nucleotide variants distributed every 3MB throughout the genome to capture large copy number deletions/amplifications.

Of 447 tumor samples, 421 (94%) passed quality control. As a technical set of normal samples (panel of normals), we included DNA extracted from the blood of ten healthy, cancer free individuals. Two tumor samples failed due to low coverage, 12 due to sample contamination and 12 due to duplication. The median sequencing coverage per sample was 539x. Raw sequence data were aligned to the human genome (NCBI build 37) using BWA[20]. Variant calling for single nucleotide variants was performed using Mutect2[21], Strelka[22] and CaVEMan[23] and for insertions/deletions using Pindel[24], Mutect2[21] and Strelka[22]. We considered mutations to be true if they: (1) passed at least two variant callers; (2) were present at a variant allele fraction of greater than 2%; (3) were present in gNOMAD[25] whole exome sequencing data with a maximum population frequency of less than 0.001; (4) had a variant allele frequency (VAF) at least two times greater than the median VAF in a panel of normal samples; and (5) were present in none of the panel of normal samples at a VAF of 2% or greater. We further excluded mutations in low complexity regions (DUST[26] score >7). Mutations in known cancer hotspots that met all other requirements but failed due to low complexity or to only being passed by one variant caller were retained for consideration. We calculated a microsatellite instability score for each tumor using MSI sensor[27]

We used Bayesian Dirichlet processes to establish classification rules that partitioned tumors into subgroups, minimizing overlap between categories. The Dirichlet process defines an infinite prior distribution for the number and proportions of clusters in a mixture model, fitted with the use of the Markov chain Monte Carlo method[28]. Our method was based on an implementation of the Dirichlet

process mixture model available at https://github.com/nicolaroberts/hdp using a non-hierarchical Dirichlet process. We used 5,000 burnin iterations and subsequently sampled 10,000 realizations at intervals of 20 iterations. From this collection of data, we computed the optimal number of clusters, requiring that 90% of the samples were assigned a cluster.

Whole Transcriptome Sequencing and Analysis

RNA-Seq libraries were prepared for 211 cases from total RNA derived from the same tumor section using poly(A) enrichment of the mRNA. 100 bp paired-end libraries were sequenced on Illumina's HiSeq at a targeted depth of 40 million reads per sample. We performed alignment using STAR[29] (version STAR_2.5.1b) against the reference genome hg38 (GENCODE v26). Reads were summarized using featureCounts[30] (version 1.5.0-p1). RNA clusters were defined using hierarchical clustering using the top 500 most variable protein coding genes (clustering parameters: method = ward.D2, distance = canberra). Differentially expressed genes between RNA cluster 1 and RNA cluster 2 samples were obtained using the R package DESeq2[31] (version 1.28.1) with collection site and RNA cluster as part of the design formula. Pathway enrichment analysis was performed using Metascape[4] (version 3.5), looking for enrichment of GO and KEGG terms, Hallmark, Reactome and BioCarta Gene Sets, and Canonical Pathways. The top 500 most overexpressed genes in RNA cluster 1 ($\log_2$ fold change < 1 and FDR < 0.05) and the top 500 most overexpressed genes in RNA cluster 2 were used as input for Metascape[32].

Outcome Analyses

Survival data was available for 350 cases. Survival time was calculated from the date of diagnosis to last follow-up and allowed for left truncation for cases who were consented following diagnosis. We right censored at five years from diagnosis to reduce non-ovarian cancer related deaths. Race, age at diagnosis (continuous and quadratic, assigned as site median for three cases), tumor stage, extent of residual disease and study site were considered as covariates using a Cox Proportional Hazards model. Proportionality of hazards was examined using Schoenfeld residuals. In addition, contingency analysis was done on tumor mutational status and tumor cluster with primary treatment response

(complete response or partial response compared to stable or progressive disease) stratified by tumor stage and vital status up to five years using a chi-square test.

Data Availability Statement

The somatic variant calls and normalized RNAseq intensity data, code and deidentified clinical data is available here: https://github.com/kbolton-lab/Bolton_OCCC . This will enable all the figures and tables to be re-generated and also provide data for others for future analyses. We will also make the BAMs/FASTQs available to researchers through contacting Kelly Bolton (bolton@wustl.edu).

RESULTS

Clinical characteristics

Key characteristics, other than race, of the 421 participants included in the study did not vary between study sites (Table 1). Compared to clinical characteristics reported in the literature for women with HGSOC[10,33], women with CCOC in this cohort were more likely to be of Asian ancestry (12% of individuals with non-missing race), have a history of endometriosis (13%) and present with early stage disease (69%).

Targeted DNA sequencing of candidate CCOC driver genes

In 163 candidate CCOC driver genes we identified 6,361 mutations. Of these, 1,488 mutations were classified as potentially pathogenic based upon annotation in OncoKB[34], frequency in COSMIC, frequency in previously published CCOC sequencing data[12,13,16], predicted pathogenicity based on PolyPhen[35] and SIFT[36], and prior evidence in the literature (Supplementary Table 2). At least one putative driver mutation was identified in 401 out of 421 tumors (95%) (mean number of mutations 3, range 1-25) (Figure 1a and c). The most commonly mutated genes were *ARID1A* (49%, N=205), *PIK3CA* (45%, N=188) and the *TERT* promoter (20%, N=84). The most frequently recurrent mutations were clonally dominant with a VAF >35% (e.g. *ARID1A* and *TP53*) suggesting that they represented early events while others (e.g. *CREBBP*) were more often sub-clonal, possibly representing secondary

events (Figure 1b). We detected a higher proportion (16%, N=71) of tumors with *TP53* mutations than has been described by some (9-15%)[13,37] but not all NGS studies (18%)[38]. This raises the possibility that some of the CCOCs in this cohort were misdiagnosed high-grade serous or endometrioid ovarian cancers. We explored this possibility in detail. First, we noted that 10 out of 71 *TP53* mutations (14%) were deeply sub-clonal (VAF<10%); previous studies may not have detected these mutations as they used lower-depth sequencing (Figure 1b). Second, we performed additional pathologic review to verify clear cell histology for a subset of the cases where formalin-fixed paraffin-embedded (FFPE) tissue sections were available. This included 14 (20%) of the *TP53*-mutated cases and 4 (15%) of the *BRCA1/2*-mutated cases where formalin-fixed paraffin-embedded (FFPE) tissue sections were available. On the basis of morphology combined with and immunohistochemical staining of Napsin A, p53, and WT1[39] (markers of HGSOC and not CCOC) it was determined that four out of 14 *TP53*-mutant cases (28%) (three endometrioid carcinomas and one HGSOC) were misclassified as CCOC. None of the *BRCA1/2*-mutated cases were misclassified. Thus, by extrapolation we estimate that approximately 19 of our 71 *TP53*-mutant tumors in this cohort were misclassified.

A subset of tumors (N=20) bore mutations in SMARCA4, a gene that is the sole driver mutation in ovarian small cell carcinoma hypercalcemic type (OSCCHT)[40–42]. However, unlike OSCCHT, in our CCOC cases we observed SMARCA4 to be most commonly co-mutated with either ARID1A (50%) or PIK3CA (35%). Similar to our analysis of TP53 mutated cases we performed central pathology review of a subset (N=8) of the SMARCA4 mutated cases. All of these cases showed typical CCOC morphology and were positive for clear cell markers such as PAX8 (8/8 diffuse), and Napsin A (5/8 diffuse, 2/8 focal) or HNF1B (5/5 diffuse). We conclude that there was no evidence for these cases being misclassified OSCCHT. Whether SMARCA4 has a similar driver capacity in CCOC compared to OSCCHT requires further study.

Most cases (75%) had at least one large-scale copy number event with the most frequently recurrent events reflecting common cancer-driver aneuploidies including 8q amplification[19] (Supplementary Figure 1). Cases with *TP53* mutations had more whole chromosome or arm-level aneuploidies (mean =12) compared to wild-type tumors (mean = 8) (Supplementary Figure 2). *TP53*-mutant/*ARID1A*-mutant

tumors showed less genomic instability (mean number of aneuploidies=7) compared to *TP53*-mutant/*ARID1A*-wildtype tumors (mean number of aneuploidies=13). We detected recurrent fusions in *TGM7* (N=5) as previously shown by Earp *et al*[43]. In addition, recurrent fusions involving *BCAR4* (N=6), *ITCH (*N=6*)* and *DCAF12* (N=5) were observed. These are known cancer fusion partners but have not been reported in CCOC before. (Supplementary Figure 3).

We evaluated mutation status with respect to clinical and epidemiological factors including age, race, tumor and history of endometriosis. Compared to *ARID1A*-mutated tumors, patients with *KRAS* mutations were older at presentation (median age 53 vs. 67, p=0.03; Figure 2a). Individuals with a history of endometriosis were more likely to have *ARID1A*-mutated tumors (72% and 47% of patients with and without endometriosis respectively, $p=2x10^{-4}$) (Figure 2b). Advanced stage tumors were more likely to harbor *TP53* mutations than early-stage tumors (27% vs. 11% respectively, $p=2x10^{-4}$) (Figure 2c). Among *TP53* mutant tumors, a similar proportion (50% and 51%, respectively) were advanced stage with or without co-occurring *ARID1A* mutations. There was a trend towards a higher frequency of *ARID1A*-mutated tumors in women of east Asian descent but this was not significant (Figure 2d).

We next examined the relationship between mutational burden, cancer driver genes and patterns of genetic co-occurrence. Several genes harbored recurrent mutations within the same tumor (Supplementary Figure 4). This seen for both tumor suppressor genes (e.g. *ARID1A*) and specific oncogenes including *PIK3R1* and *PIK3CA*. Among tumors with multiple *PIK3CA* mutations*,* variants were more likely to occur in non-hotspot locations within the gene (Supplementary Figure 5)[44]. MSIsensor score was higher among individuals more than 10 driver mutations (N=12, 3%) and among those with *MSH2* and *MSH6* mutations (Supplementary Figure 6). We observed a statistically significant co-occurrence between mutations in *ARID1A, PIK3CA, TP53* and *BRCA1/BRCA2* Mutual exclusivity between somatic mutations of *ARID1A, TP53, PIK3CA* and *PIK3R1* (Supplementary Figure 7) suggests that these may represent distinct pathways to oncogenesis. The exclusivity between *TP53* and *ARID1A* mutation was stronger in the setting of multiple *ARID1A* mutations (OR=0.21; 95% CI 0.07-0.54, $p=2x10^{-4}$) compared to a single *ARID1A* mutations (OR=0.68, 95% CI 0.32-1.34, p=0.28). "We observed 54 mutations in genes known to be relevant to high penetrance genetic predisposition to

ovarian cancer including *PMS2*, *MSH6*, *MSH2*, *BRCA1* and *BRCA2.* Overall 52% of these mutations were present at a VAF in the tumor of >=35%. In the absence of matched normal tissue sequencing, we were not able to distinguish these from germline variants. Thus, it is possible that up to 26 cases (6% of the cohort) harbored a germline pathogenic variant in a known cancer susceptibility gene."

Because we observed clear patterns of exclusivity and co-occurrence between gene drivers, we used unsupervised clustering approaches to define non-overlapping subgroups of CCOC based on their mutational spectrum. We defined seven subgroups (Supplementary Figure 8) and compared the frequency of mutations between subgroups. Four clusters were characterized by having an *ARID1A* mutation; the first cluster (cluster A) was characterized by a single *ARID1A* mutation in combination with another disease defining mutation (e.g. *PIK3CA*, *TERT*, *TP53*, *KRAS*, *PTEN*, *PPP2R1A*, *PIK3R1*, *CREBBP* or *SPOP*) (N=86); the second (cluster B) with a single *ARID1A* mutation alone or in combination with non-disease defining mutation (N=19); the third (cluster C) with multiple *ARID1A* mutations combined with a *PIK3CA* mutation (N=81); and a forth (cluster D) with multiple *ARID1A* mutations and *PIK3CA* wild-type (N=25). Two clusters were *ARID1A* wildtype: Cluster E was defined by a *TP53* mutation (N=50); and cluster F by other non-*TP53* disease-defining mutations (N=104). A final cluster (cluster G) was characterized by mutations in *SMARCA4* (N=13); a mutation typically observed in small cell ovarian carcinoma[23]. The remaining tumors were undefined (N=57).

Similar to the patterns we observed when studying the association between individual mutations and clinical features, the *TP53*-mutated, *ARID1A*-wild-type cluster showed an enrichment of advanced stage disease while tumors belonging to the *ARID1A*-mutant clusters were more likely in individuals of Asian ancestry and those with a history of endometriosis (Supplementary Figure 9). Individuals in cluster G (*SMARCA4*-mutant tumors) had a non-significant trend towards a younger age at diagnosis (p=0.32).

Transcriptomic profiling of CCOC

Transcriptomic profiles were generated for 212 CCOC tumors in which targeted sequencing was also performed. Using unsupervised clustering informed by expression of the 500 most variable genes, we

identified two main RNA clusters (Supplementary Figure 10): Expression cluster 1 showed higher expression of genes previously reported as highly expressed in CCOC including *ANXA4* and *GPX3*, both of which are linked to platinum resistance[45,46]. Among the most highly expressed genes in cluster 1 compared to 2 also included *GPX3*[4,7], which is known to be overexpressed in endometriosis compared to normal endometrial tissue, and *EEF1A2*, known to be overexpressed in CCOC associated endometriosis but not benign endometriosis[48]. Genes that characterized this cluster were enriched in metabolic pathways including flavonoid glucuronidation ($p=10^{-15}$) and monocarboxylic acid metabolism ($p=10^{-13}$). Expression cluster 2 showed enriched expression of genes involved in extracellular matrix (ECM) organization ($p=10^{-22}$) and mesenchymal differentiation, including genes such as *ADGR2* and *PDCH19* (Supplementary Figure 10 and Figure 3b). Compared to cluster 1, expression cluster 2 also showed higher expression of *WT1* and lower expression of CCOC marker *HNF1B* , which are features classically associated with high-grade serous ovarian cancer[9] (Figure 3b). Expression cluster 2 was enriched with *TP53*-mutant tumors (55% of cases in cluster 2 compared to 10% in cluster 1). When comparing RNA expression and mutation clusters, cluster 2 was largely comprised of tumors belonging to mutation cluster E i.e *TP53*-mutant *ARID1A*-wildtype tumors (45% of cluster 2) and the undefined mutation cluster (33% of cluster 2) (Figure 3a).

Clinical Outcomes

There was no statistically significant association between overall survival and CCOC mutations when examined on a per-gene level in Cox proportional hazards models stratified by study site (Supplementary Table 3). We observed a non-significant trend towards improved survival for patients with *ARID1A* (HR=0.82, 95% CI 0.58-1.15, p=0.24) and *PTEN* (HR=0.52, 95% CI 0.24-1.12, p=0.10) mutant tumors. Because of the similarity of the *ARID1A*-mutant clusters in regards to clinical presentation and outcome, we combined these clusters for the purpose of survival analysis. Women with *TP53*-mutant, *ARID1A*-wildtype tumors had worse overall survival compared to those with *ARID1A*-mutant tumors (HR=1.72, 95% CI 1.06-2.81, p=0.03, Figure 4a). Similarly, RNAseq cluster 2 showed an increased risk of death compared to RNAseq cluster 1 (Figure 4b, Tumor Cluster 2 versus Tumor Cluster 1 HR 2.8, 95% CI 1.66 – 4.84; $p=1x10^{-4}$). Covariate adjustment for age, race, stage and

residual disease attenuated the estimated mutation and cluster-associated risk (Supplementary Table 4). To explore how these subgroups might influence therapy outcome, we studied the relationship between mutation status and response to first line therapy with platinum/taxane combination therapy. We limited this to women with advanced stage disease who successfully underwent debulking surgery followed by combination platinum/taxol therapy (N=36). Women with *ARID1A* wild-type, *TP53*-mutant tumors were more likely to have a complete response 75% (N=11) compared to *ARID1A*-mutant tumors (55%), although this was not statistically significant (p=0.33) in this small sample size.

DISCUSSION

Our results have several clinical implications. First, the results of both genomic and transcriptomic cluster associations with clinical presentation and outcome converged, suggesting two main subgroups of CCOC: The first subtype included *ARID1A*-mutant tumors (particularly double-mutant tumors) and other common CCOC mutations (e.g. *PIK3CA*, *TERT* etc) that showed enriched expression of metabolic pathways, presented with early stage disease and were more likely to have a history of endometriosis. We denote this group as “classic-CCOC”, which represented 83% of our cohort. The second CCOC subtype was dominated by *TP53*-mutant tumors that showed enriched expression of genes involved in extracellular matrix organization, mesenchymal differentiation and immune-related pathways. These cases presented with advanced disease and had worse survival. Interestingly, *TP53* mutations either in the presence or absence of co-occurring *ARID1A* mutations were associated with a higher degree of genomic instability and aggressive, advanced stage tumors. The worse survival for tumors in this “HGSOC-like” subgroup was largely explained by advanced stage and higher burdens of residual disease.

Within both the “classic-CCOC” and “HGSOC-like” subgroups we noted a subset of individuals had tumor with mutations in genes known to be both somatic drivers of ovarian cancer and germline susceptibility genes including *PMS2*, *MSH6*, *MSH2*, *BRCA1*, and *BRCA2*. Due to the absence of matched normal samples, we were unable to fully distinguish whether these represented somatic or germline events and is a limitation of our study. Future studies estimating the frequency of CCOC

cases that arise in women with strong hereditary predisposition and who may be considered for risk reducing bilateral salpingo-oophorectomy should be prioritized[49].

There is increasing recognition that other histological types of ovarian carcinoma, including HGSOC and endometrioid carcinoma, can contain areas with clear cell change complicating the histologic diagnosis[50]. While a subset of cases in the "HGSOC-like" cluster are misclassified HGSOC, and is a weakness of our study, it is unlikely that this alone explains our findings. Firstly, all of our cases were morphologically diagnosed by expert gynecological pathologists and at some centers, this morphologic review was supplemented by immunohistochemistry for histotype-specific markers. Secondly, in a subset of *TP53*-mutant cases, we re-confirmed the diagnosis of CCOC using a combination of morphological and immunohistochemical features. Thus, our results suggest that a subset of *bona fide* CCOCs with HGSOC-like features exist. Our results also emphasize that expert histologic review of CCOC cases, particularly those who present with *TP53*-mutant, *ARID1A*-wildtype tumors, is warranted given similarities to the biology and behavior of HGSOC.

Gene expression profiles of the "classic-CCOC" and "HGSOC-like" CCOC subtypes we observed are similar to those reported by Tan el al[51] which also reported two clusters, the first enriched for genes in metabolic pathways and the second, a less common mesenchymal-like subgroup associated with late-stage disease. However, unlikely Tan et al., we observed differences in the frequency of *TP53*-mutated tumors across clusters. The source of this discrepancy is unclear and may include differences in sequencing technology (Tan et al. performed targeted sequencing using Ion Torrent) and patient characteristics (Tan et al., included only women of Asian ancestry which trend towards lower frequencies of *TP53*-mutated tumors in our analysis and which are known to have lower frequencies of endometrial ovarian cancer). The overlap between genes highly expressed in our "classic-CCOC" subgroup and those enriched in endometriosis provide further support for the likely transition from endometriosis to carcinoma in CCOC.

The greatest translational impact from these molecular CCOC subtypes is expected to lie in the development of therapeutic approaches tailored to the vulnerabilities of each group. Interestingly,

despite being aggressive on presentation, a trend was seen towards the “HGSOC-like” CCOC subgroup having higher response rates to first line platinum-based chemotherapy. Future studies are warranted to further explore whether genomic subtypes of CCOC predict response to platinum-based and other therapies as treatment data were limited here. The “classic-CCOC” subgroup dominated by mutations in the SWI/SNF pathway and markers linked to chemo-resistance may be of particular relevance to target for investigational first-line therapies. Recent data suggests that the SWI/SNF pathway plays a novel role in the regulation of anti-tumor immunity, and that SWI/SNF deficiency can be therapeutically targeted by immune checkpoint blockade[19]. Several studies are currently evaluating the role of immune check point inhibitors in CCOC including NCT03405454, NCT03425565. While a limitation of our study was that we were unable to assess MMR functional status, we did note a rare subset of tumors (3%) with higher mutational burden (>10 drivers) and MSIsensor score. The extent to which the subset of CCOCs with higher total mutation and with MMR deficiency show improved responsiveness to immune checkpoint blockade in ongoing clinical trials will be an important avenue of investigation. Additional targeted therapeutic strategies have been explored in preclinical settings including epigenetic synthetic lethality, some of which are entering into clinical trials. The PI3K inhibitor, alepelisib, is now FDA approved for HR-positive breast cancer and ongoing trials in additional *PIK3CA*-mutated cancers including CCOC are underway. Double *PIK3CA* mutations appear to hyperactivate PI3K signaling and enhance tumor growth and may confer increased responsiveness to PI3K inhibitors than those with a single mutation[52]. Thus, for CCOC cases harboring multiple *PIK3CA* mutations, PI3K inhibitors either alone or in combination with other agents may represent a promising approach.

The strengths of this study include the large sample size, use of multiple study sites, inclusion of women of European and non-European ancestry, and integration of genetic and transcriptomic markers of disease behavior and outcome. While this is the most extensive genomic study of CCOC to date, greater sample size with additional follow-up data will allow improved assessment and validation of these clinically relevant subtypes. Although future analyses would benefit from larger patient collections, our current results suggest that genomic classification may inform the future development of targeted therapeutics in CCOC.

Acknowledgements

Research reported in this publication was supported in part by a Cancer Center Support Grant of the NIH/NCI (Grant No. P30CA008748, MSK) and the Cycle for Survival including the Fatma Fund. B. Weigelt is funded in part by Breast Cancer Research Foundation and NIH/NCI (P50 CA247749 01) grants. KLB is funded by the Damon Runyon Cancer Research Foundation, the American Society of Hematology, the EvansMDS Foundation and the NCI (Grant 5K08CA241318). Additional support was provided by R21CA222867, R01CA248288, P30CA015083, and P50CA136393. M. Anglesio was funded through a Michael Smith Health Research BC Scholar Program award and the Janet D. Cottrelle Foundation Scholars Program (managed by the BC Cancer Foundation). This study used resources provided by the Canadian Ovarian Cancer Research Consortium's COEUR biobank funded by the Terry Fox Research Institute and managed and supervised by the Centre hospitalier de l'Université de Montréal. The Consortium acknowledges contributions of its COEUR biobank from Institutions across Canada (for a full list see http://www.tfri.ca/en/research/translationalresearch/coeur/). This work was supported by the Westmead Hospital Department of Gynaecological Oncology, Sydney Australia. The Gynaecological Oncology Biobank at Westmead (GynBiobank), a member of the Australasian Biospecimen Network-Oncology group, was funded by the National Health and Medical Research Council of Australia (Enabling Grants ID 310670 & ID 628903) and the Cancer Institute NSW (Grants 12/RIG/1-17 & 15/RIG/1-16). The Westmead GynBiobank acknowledges financial support from the Sydney West Translational Cancer Research Centre, funded by the Cancer Institute NSW. A. Piskorz, M. Sekowska, and J. Brenton were supported by Cancer Research UK grant 22905. Additional support was also provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

REFERENCES

1. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e18 (2018).
2. Dong, J., Li, B., Lin, D., Zhou, Q. & Huang, D. Advances in Targeted Therapy and Immunotherapy for Non-small Cell Lung Cancer Based on Accurate Molecular Typing. *Front. Pharmacol.* **10**, (2019).
3. Quintás-Cardama, A., Kantarjian, H. & Cortes, J. Imatinib and beyond—exploring the full potential of targeted therapy for CML. *Nat. Rev. Clin. Oncol.* **6**, 535–543 (2009).
4. Ortega, J., Vigil, C. E. & Chodkiewicz, C. Current Progress in Targeted Therapy for Colorectal Cancer. *Cancer Control* **17**, 7–15 (2010).
5. Kayser, S. & Levis, M. J. Advances in targeted therapy for acute myeloid leukaemia. *Br. J. Haematol.* **180**, 484–500 (2018).
6. Tangutoori, S., Baldwin, P. & Sridhar, S. PARP inhibitors: A new era of targeted therapy. *Maturitas* **81**, 5–9 (2015).
7. Vaishnavi, A., Le, A. T. & Doebele, R. C. TRKing Down an Old Oncogene in a New Era of Targeted Therapy. *Cancer Discov.* **5**, 25–34 (2015).
8. Lee, Y.-Y. *et al.* Prognosis of ovarian clear cell carcinoma compared to other histological subtypes: a meta-analysis. *Gynecol. Oncol.* **122**, 541–547 (2011).
9. Irodi, A. *et al.* Patterns of clinicopathological features and outcome in epithelial ovarian cancer patients: 35 years of prospectively collected data. *BJOG Int. J. Obstet. Gynaecol.* **127**, 1409–1420 (2020).
10. Wentzensen, N. *et al.* Ovarian Cancer Risk Factors by Histologic Subtype: An Analysis From the Ovarian Cancer Cohort Consortium. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **34**, 2888–2898 (2016).
11. Iida, Y., Okamoto, A., Hollis, R. L., Gourley, C. & Herrington, C. S. Clear cell carcinoma of the ovary: a clinical and molecular perspective. *Int. J. Gynecol. Cancer* ijgc (2020) doi:10.1136/ijgc-2020-001656.

12. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).

13. Wang, Y. K. *et al.* Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **49**, 856–865 (2017).

14. Anglesio, M. S. *et al.* Multifocal endometriotic lesions associated with cancer are clonal and carry a high mutation burden. *J. Pathol.* **236**, 201–209 (2015).

15. Murakami, R. *et al.* Exome Sequencing Landscape Analysis in Ovarian Clear Cell Carcinoma Shed Light on Key Chromosomal Regions and Mutation Gene Networks. *Am. J. Pathol.* **187**, 2246–2258 (2017).

16. Cheng, D. T. *et al.* Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J. Mol. Diagn. JMD* **17**, 251–264 (2015).

17. Consortium, T. A. P. G. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* **7**, 818–831 (2017).

18. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).

19. Fukumoto, T., Magno, E. & Zhang, R. SWI/SNF Complexes in Ovarian Cancer: Mechanistic Insights and Therapeutic Implications. *Mol. Cancer Res. MCR* **16**, 1819–1825 (2018).

20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).

21. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019) doi:10.1101/861054.

22. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28**, 1811–1817 (2012).

23. Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr. Protoc. Bioinforma.* **56**, 15.10.1-15.10.18 (2016).

24. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).

25. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

26. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **13**, 1028–1040 (2006).

27. Niu, B. *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinforma. Oxf. Engl.* **30**, 1015–1016 (2014).

28. Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. Hierarchical Dirichlet Processes. *J. Am. Stat. Assoc.* **101**, 1566–1581 (2006).

29. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

30. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

32. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

33. Lisio, M.-A., Fu, L., Goyeneche, A., Gao, Z. & Telleria, C. High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int. J. Mol. Sci.* **20**, 952 (2019).

34. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, (2017).

35. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al* **0 7**, Unit7.20 (2013).

36. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

37. Kuo, K.-T. *et al.* Frequent Activating Mutations of PIK3CA in Ovarian Clear Cell Carcinoma. *Am. J. Pathol.* **174**, 1597–1601 (2009).

38. Friedlander, M. L. *et al.* Molecular Profiling of Clear Cell Ovarian Cancers. *Int. J. Gynecol. Cancer* **26**, 648–654 (2016).

39. Köbel, M. *et al.* An Immunohistochemical Algorithm for Ovarian Carcinoma Typing. *Int. J. Gynecol. Pathol.* **35**, 430–441 (2016).

40. Jelinic, P. *et al.* Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat. Genet.* **46**, 424–426 (2014).

41. Ramos, P. *et al.* Small cell carcinoma of the ovary, hypercalcemic type, displays frequent inactivating germline and somatic mutations in SMARCA4. *Nat. Genet.* **46**, 427–429 (2014).

42. Witkowski, L. *et al.* Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat. Genet.* **46**, 438–443 (2014).

43. Earp, M. A. *et al.* Characterization of fusion genes in common and rare epithelial ovarian cancer histologic subtypes. *Oncotarget* **8**, 46891–46899 (2017).

44. Saito, Y. *et al.* Landscape and function of multiple mutations within individual oncogenes. *Nature* **582**, 95–99 (2020).

45. Morimoto, A. *et al.* Annexin A4 induces platinum resistance in a chloride-and calcium-dependent manner. *Oncotarget* **5**, 7776–7787 (2014).

46. Y, S. *et al.* Glutathione peroxidase 3 is a candidate mechanism of anticancer drug resistance of ovarian clear cell adenocarcinoma. *Oncol. Rep.* **20**, (2008).

47. Tamaresis, J. S. *et al.* Molecular Classification of Endometriosis and Disease Stage Using High-Dimensional Genomic Data. *Endocrinology* **155**, 4986–4999 (2014).

48. Worley, M. J. *et al.* Molecular changes in endometriosis-associated ovarian clear cell carcinoma. *Eur. J. Cancer* **51**, 1831–1842 (2015).

527 49. Berek, J. S. *et al.* Prophylactic and risk-reducing bilateral salpingo-oophorectomy:
528 recommendations based on risk of ovarian cancer. *Obstet. Gynecol.* **116**, 733–743 (2010).
529 50. DeLair, D. *et al.* HNF-1β in Ovarian Carcinomas With Serous and Clear Cell Change. *Int. J.*
530 *Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **32**, 541–546 (2013).
531 51. Tan, T. Z. *et al.* Analysis of gene expression signatures identifies prognostic and functionally
532 distinct ovarian clear cell carcinoma subtypes. *EBioMedicine* **50**, 203–210 (2019).
533 52. Vasan, N. *et al.* Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3Kα
534 inhibitors. *Science* **366**, 714–723 (2019).
535
536

	ADD (N=28)	BWH (N=9)	COEUR (N=181)	MAY (N=38)	MSK (N=60)	PIT (N=24)	SCOT (N=22)	UPA (N=7)	WCP (N=28)
Age(y)									
0-40	0 (0%)	0 (0%)	3 (1.7%)	1 (2.6%)	1 (1.7%)	0 (0%)	0 (0%)	0 (0%)	2 (7.1%)
40-50	1 (3.6%)	0 (0%)	37 (20.4%)	3 (7.9%)	6 (10.0%)	4 (16.7%)	4 (18.2%)	2 (28.6%)	7 (25.0%)
50-60	8 (28.6%)	2 (22.2%)	81 (44.8%)	16 (42.1%)	28 (46.7%)	9 (37.5%)	7 (31.8%)	2 (28.6%)	14 (50.0%)
60-70	13 (46.4%)	7 (77.8%)	48 (26.5%)	9 (23.7%)	19 (31.7%)	5 (20.8%)	9 (40.9%)	2 (28.6%)	4 (14.3%)
70+	6 (21.4%)	0 (0%)	11 (6.1%)	9 (23.7%)	6 (10.0%)	6 (25.0%)	2 (9.1%)	1 (14.3%)	1 (3.6%)
Missing	0 (0%)	0 (0%)	1 (0.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Race									
White	16 (57.1%)	9 (100%)	0 (0%)	38 (100%)	44 (73.3%)	23 (95.8%)	0 (0%)	6 (85.7%)	23 (82.1%)
Asian	2 (7.1%)	0 (0%)	0 (0%)	0 (0%)	13 (21.7%)	0 (0%)	0 (0%)	0 (0%)	4 (14.3%)
Black	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.7%)	1 (4.2%)	0 (0%)	1 (14.3%)	1 (3.6%)
Other	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Unknown	10 (35.7%)	0 (0%)	181 (100%)	0 (0%)	0 (0%)	0 (0%)	22 (100%)	0 (0%)	0 (0%)
Endometriosis									
yes	0 (0%)	0 (0%)	13 (7.2%)	10 (26.3%)	6 (10.0%)	0 (0%)	2 (9.1%)	2 (28.6%)	7 (25.0%)
no	0 (0%)	9 (100%)	168 (92.8%)	26 (68.4%)	49 (81.7%)	0 (0%)	20 (90.9%)	5 (71.4%)	0 (0%)
unknown	28 (100%)	0 (0%)	0 (0%)	2 (5.3%)	5 (8.3%)	24 (100%)	0 (0%)	0 (0%)	21 (75.0%)
FIGO Stage									
I/II	17 (60.7%)	7 (77.8%)	128 (70.7%)	25 (65.8%)	42 (70.0%)	16 (66.7%)	14 (63.6%)	2 (28.6%)	15 (53.6%)
III/IV	5 (17.9%)	2 (22.2%)	46 (25.4%)	12 (31.6%)	17 (28.3%)	8 (33.3%)	7 (31.8%)	5 (71.4%)	13 (46.4%)
Missing	6 (21.4%)	0 (0%)	7 (3.9%)	1 (2.6%)	1 (1.7%)	0 (0%)	1 (4.5%)	0 (0%)	0 (0%)

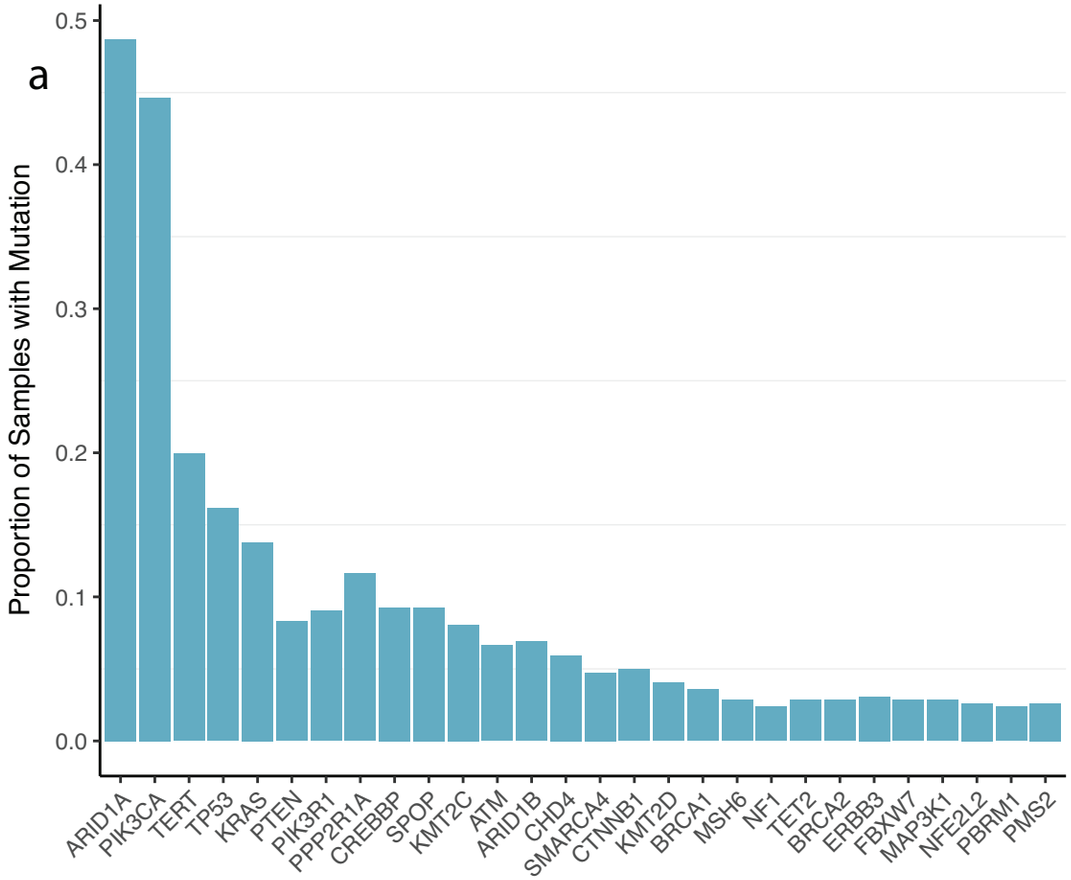
Table 1. Clinical Characteristics of CCOC cases sequenced using targeted panel

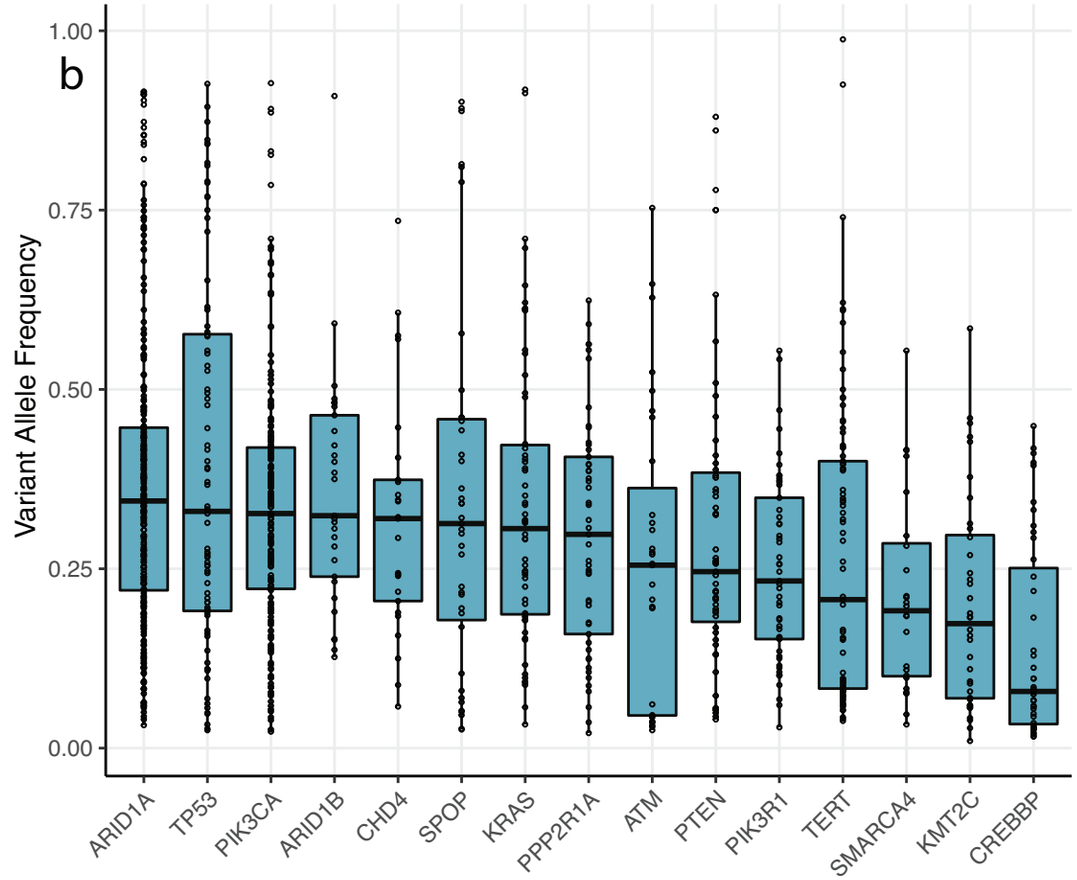
Figure 1. Mutational landscape of 401 Clear Cell Ovarian Carcinomas with a Detectable Mutation. (a) Proportion of patients with mutations in commonly mutated genes. (b) Mutation variant allele frequency (VAF) by genes mutated in at least 10% of individuals. (c) Number of mutated genes per individual. (d) Variant effect and nucleotide substitution change for single nucleotide variants.

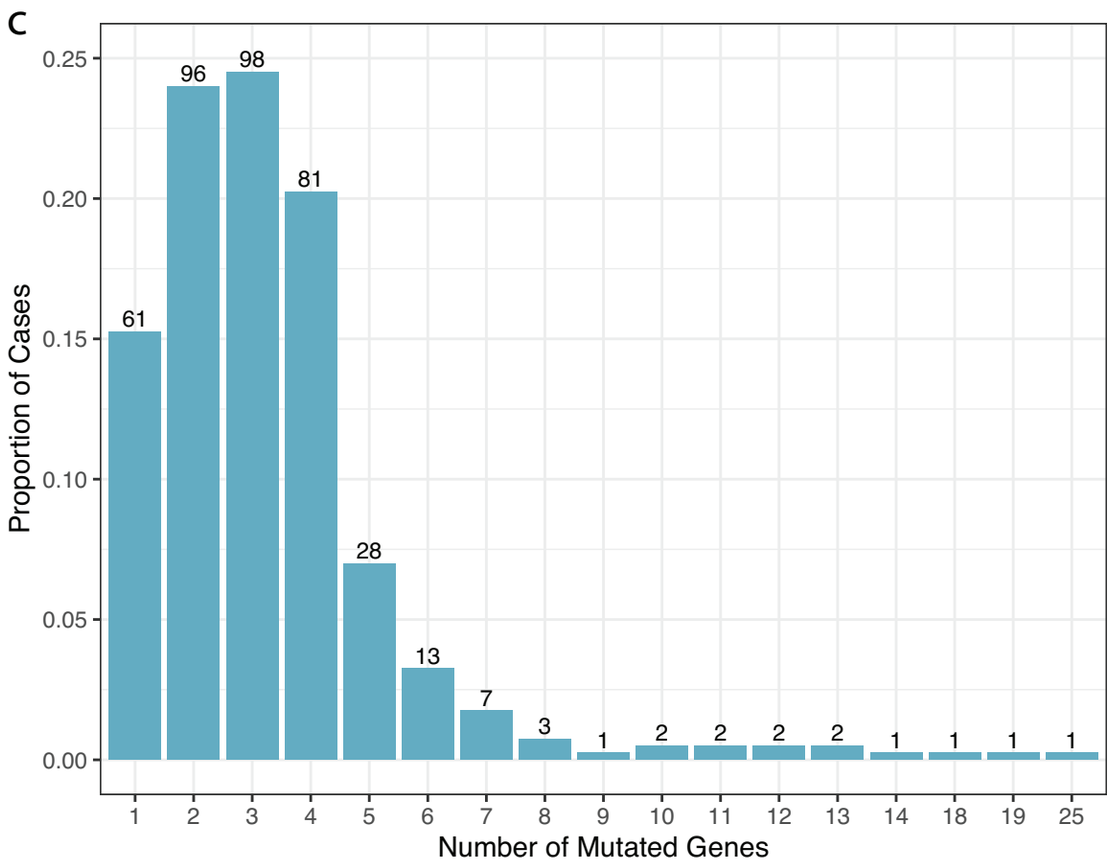
Figure 2. Frequency of somatic mutations by clinical characteristics including a) age at diagnosis, b) endometriosis, c) stage and d) race. Genes that were mutated in at least 20 individuals with non-missing values for the clinical characteristic were included. Shown are q-values (FDR corrected p-values) based on fisher's exact test. * $q<0.05$; ** $q<0.01$

Figure 3. The transcriptome of Clear Cell Ovarian Cancer Samples. (a) Sankey plot showing the correspondence of the samples annotations RNA clusters and DNA clusters. (b) Heatmap showing the normalized gene expression of the top 50 most differentially expressed genes between RNA cluster 1 and RNA cluster 2.

Figure 4. Association between CCOC molecular subgroups and all-cause mortality. Shown are the Kaplan Meier plots for the survival probability over five years following CCOC diagnosis stratified by (a) mutational clusters defined by ARID1A/TP53 mutation status, (b) RNAseq expression clusters.







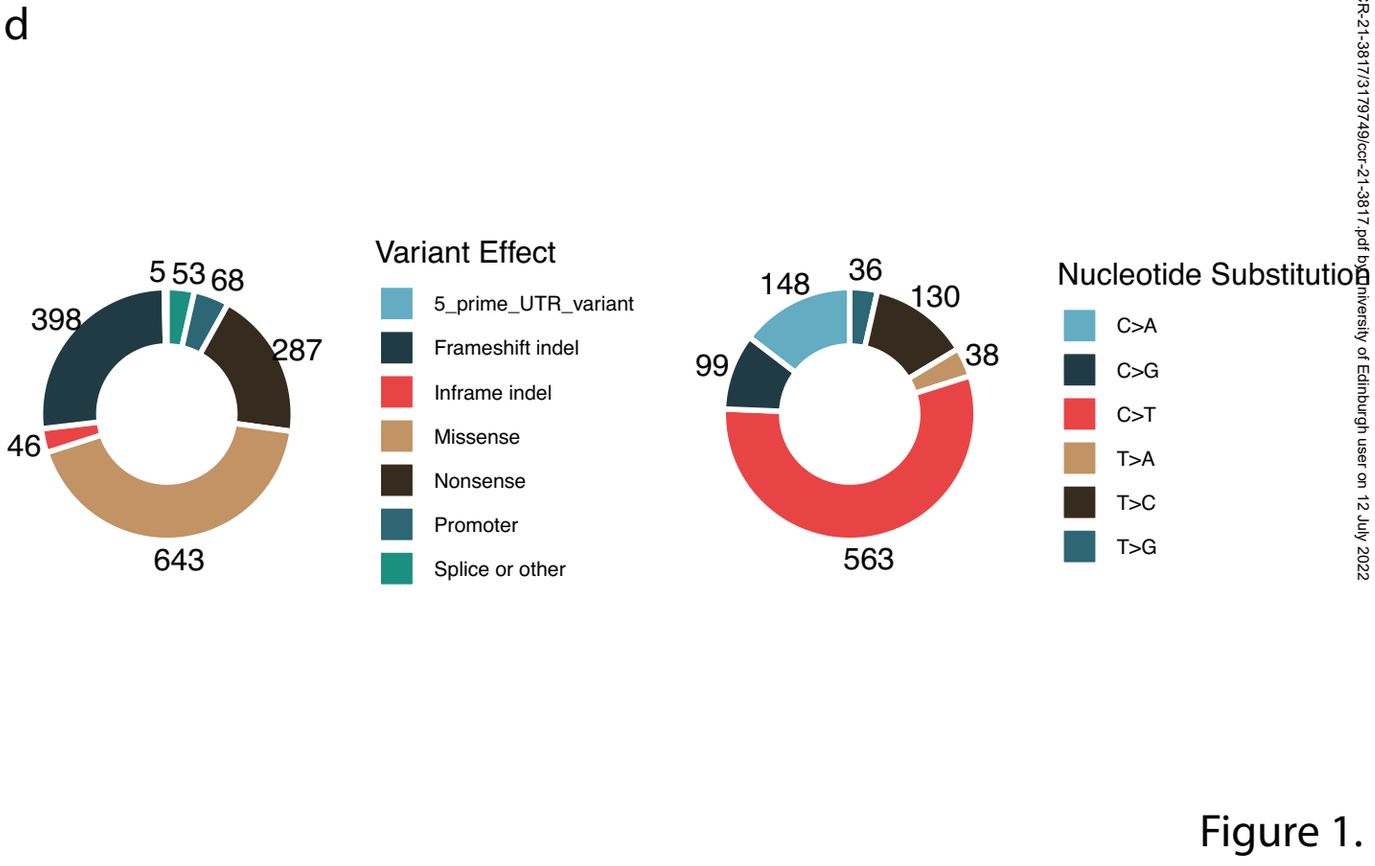


Figure 1.

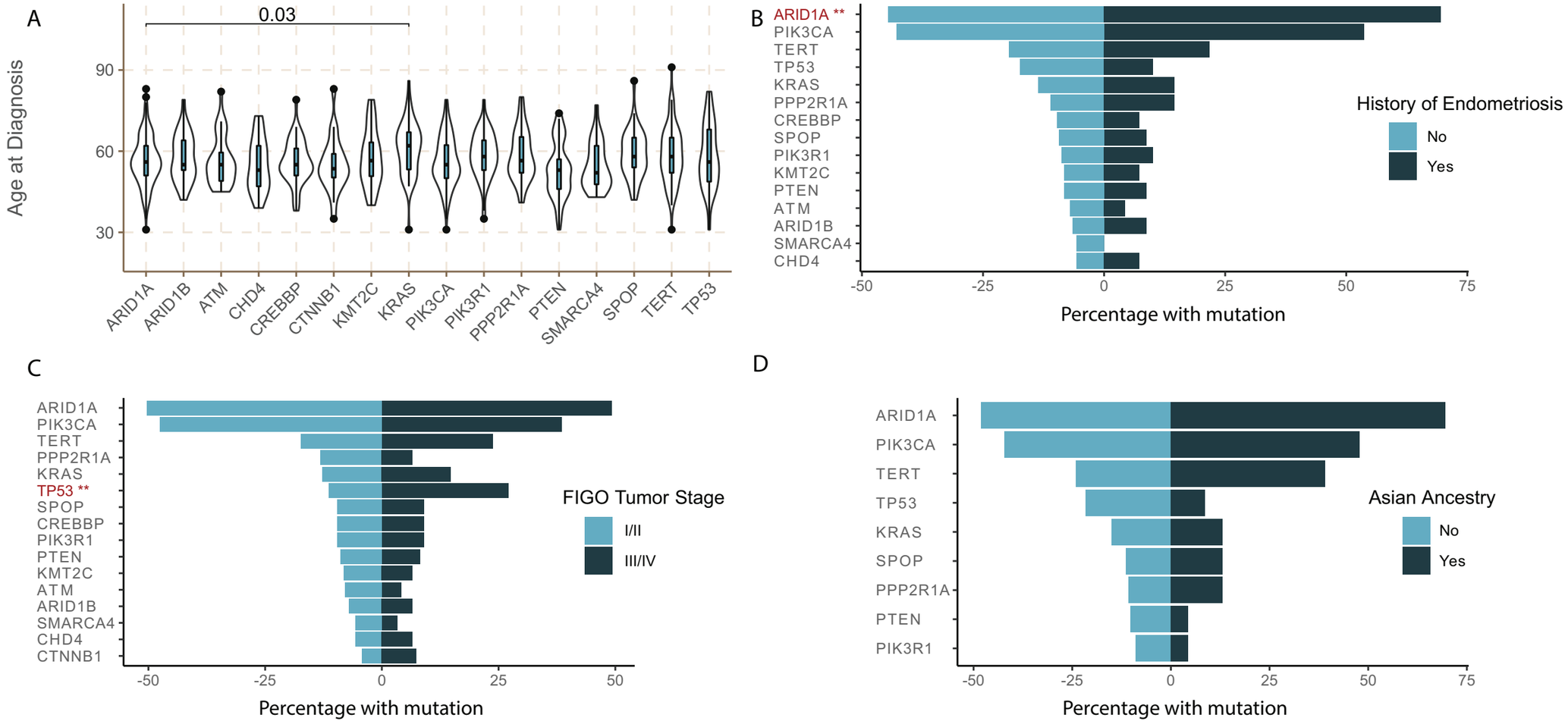


Figure 2

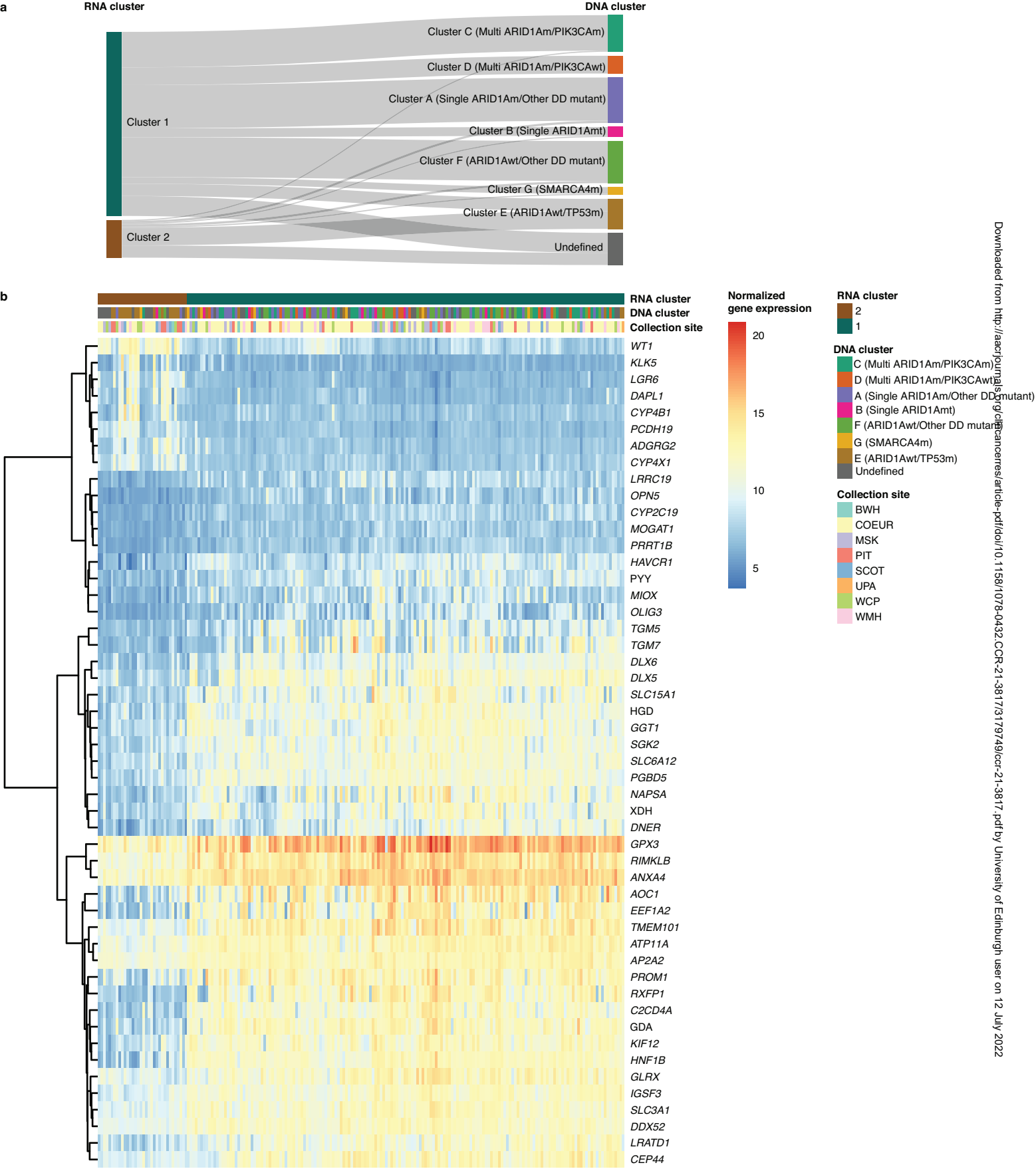
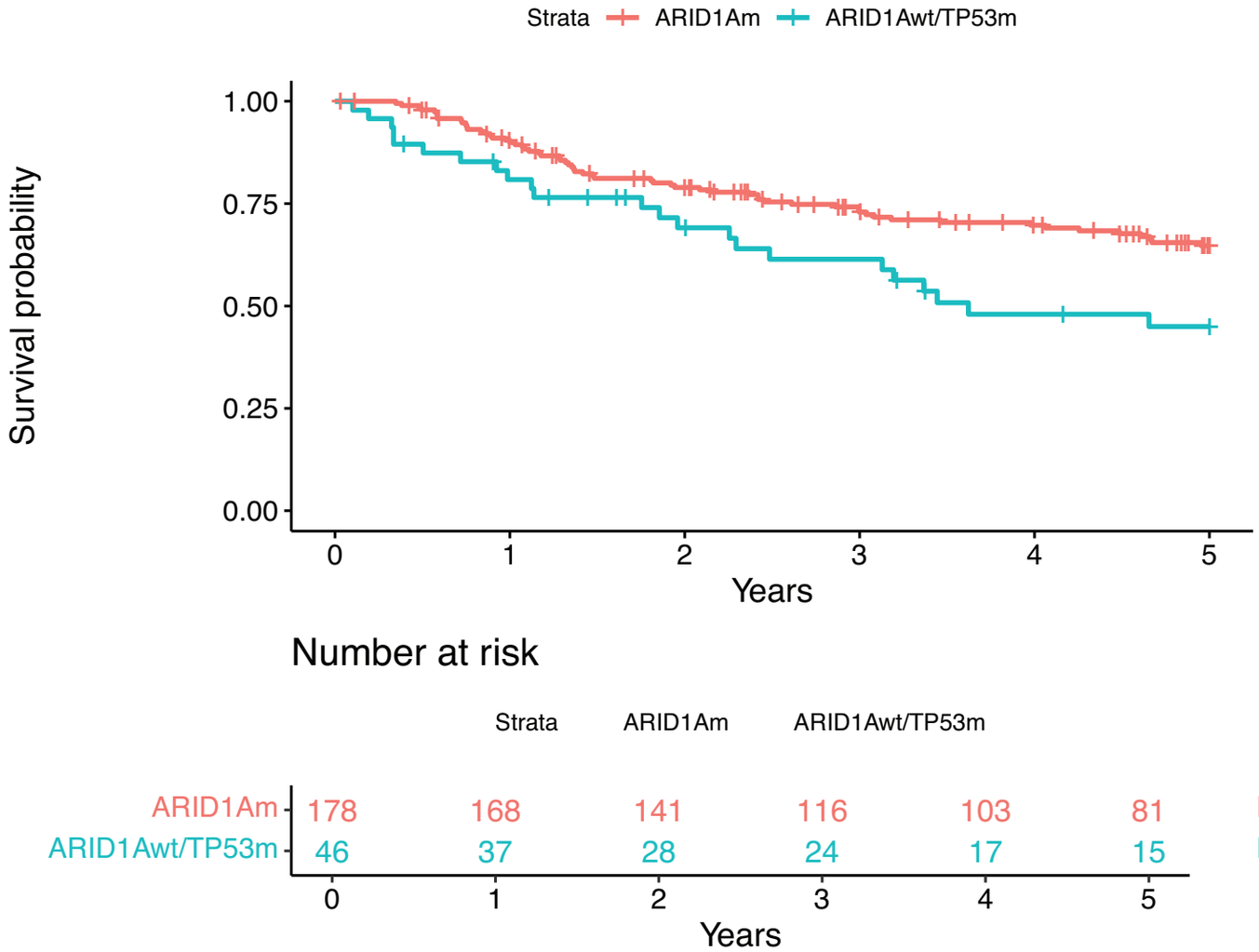


Figure 3



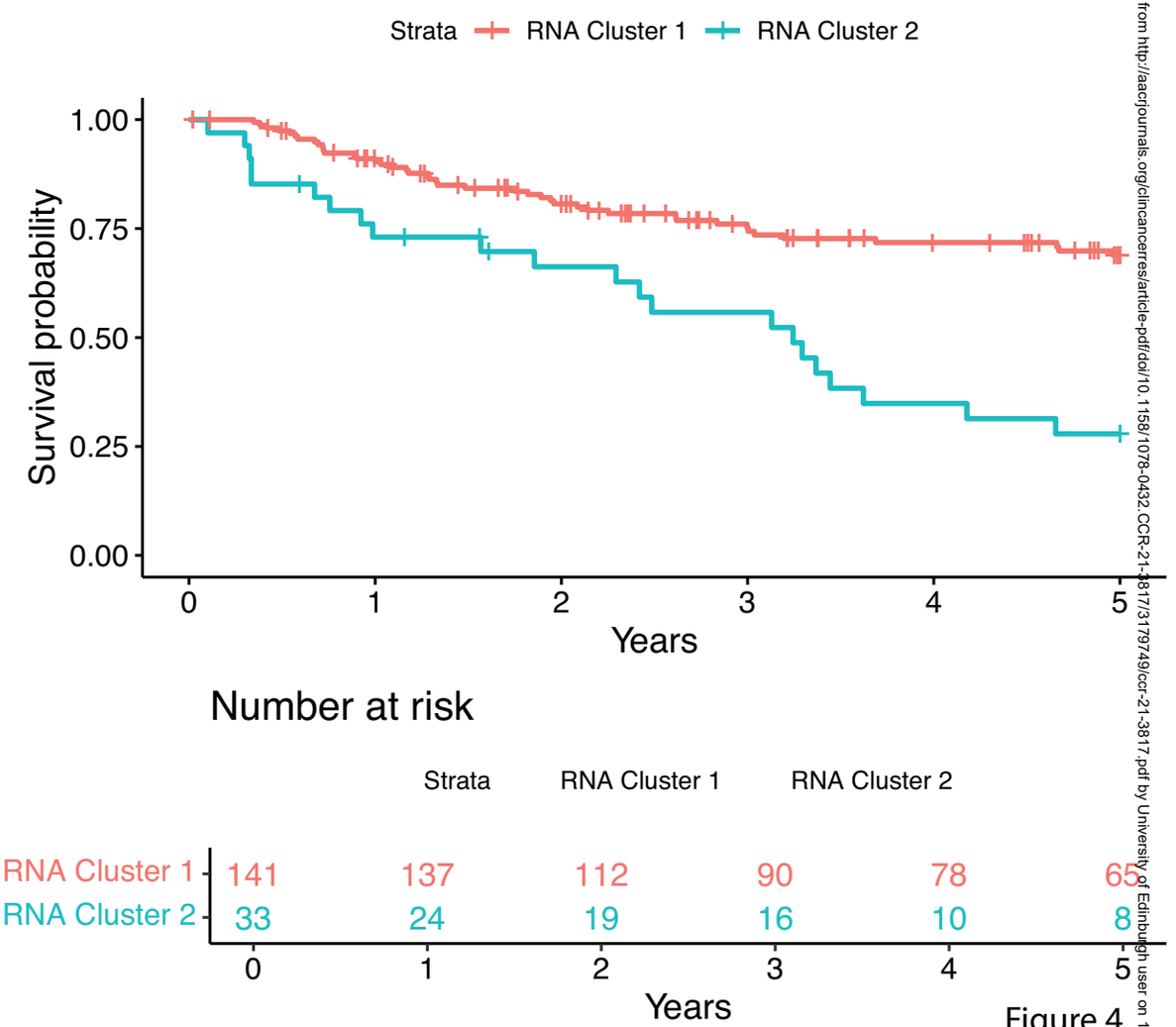


Figure 4