Integrated molecular profiling of patient-derived ovarian cancer models identifies clinically relevant signatures and tumor vulnerabilities.

Michela Lupia, Valentina Melocchi, Francesca Bizzaro, Pietro Lo Riso, Elisa Dama, Micol Baronio, Alberto Ranghiero, Massimo Barberis, Loris Bernard, Giovanni Bertalot, Raffaella Giavazzi, Giuseppe Testa, Fabrizio Bianchi and Ugo Cavallaro

SUPPLEMENTAL MATERIAL

Table of contents:

Methods
References
Supplemental Figure and Table legends
Supplemental Figure S1-S7
Supplemental Tables S1-S9 (available as a separate file)
METHODS:

Spheroid formation assay

Single cell suspensions from primary cultures were seeded in Poly-(2-hydroxyethyl methacrylate) (P3932-25G, Sigma)-coated dishes at a density of 5x10^3 cells/ml in serum-free MEBM (Lonza) supplemented with 2mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin, 5 µg/ml insulin, 0.5 µg/ml hydrocortisone, 1 U/ml heparin, 2% B27, 20 ng/ml EGF, 20 ng/ml basic FGF. Spheroid formation was assessed 7 days after seeding.

Antibody List (Immunofluorescence)

Immunofluorescence was performed with anti-CD73 (undiluted, clone AA60E3.3), generated by E. Dejana (Milan) and produced by Cogentech, Milan; anti-CD44 (1:40, clone Hermes-3, from U. Günthert, Basel); anti-Vimentin (1:100, Abcam Cat. Ab92547); anti-CK8 (1:100, Troma-I, generated by R. Kemler and distributed by Developmental Studies Hybridoma Bank, University of Iowa); anti-CK7 (1:100, AbCam Cat. Ab68459) anti-EpCAM (1:100, AbCam Cat. Ab32392), anti-CA125 (1:50, Millipore Cat. CA1004). The staining was performed as previously described \(^1\).

Antibody List (Immunoblot)

Immunoblotting was performed with anti-Vinculin (Sigma Cat. V9131), phospho-AKT (Cell Signaling Cat. 3787S) and total AKT (Protein Atlas Cat. HPA002891). The signal was detected by the Clarity Western ECL Substrate (BioRad) and the images were acquired by ChemiDoc (BioRad) and analyzed using the Image Lab software (BioRad).

Histopathological analysis

After an overnight incubation at 37 °C, tumor sections (3µm) were deparaffinized using Leica ST5020 Multistainer according to the following protocol: Bio-clear (Bio-Optica
Ref. 06-1782D) twice for 5 minutes, absolute ethanol twice for 2 minutes, 95% ethanol for 2 minutes, 70% ethanol for 2 minutes. After two washes in distilled water, tissue sections were treated with hemalast reagent (Leica) for 30 seconds and with Hematoxylin (Leica) for 5 minutes. After one wash in water, sections were treated with differentiator (Leica) for 45 seconds. After one wash in water, sections were treated with bluing agent (Leica), treated with 80% ethanol for 1 minute and finally stained with eosin (Leica) for 1 minute. Hemalast, Hematoxylin, differentiator, eosin and bluing reagents were from Leica ST Infinity H&E Staining System (Leica, 3801698). Pictures of stained sections were acquired with the scanner Aperio ScanScope XT, 20X and 40X objective.

**RNA-Seq analysis**

Poly-A enriched strand-specific libraries were generated with the TruSeq mRNA V2 sample preparation kit (#RS-122-2001, Illumina), ribosomal RNA depleted strand-specific RNA libraries with the TruSeq Stranded Total RNA LT sample preparation kit with Ribo-Zero Gold (#RS-122-2301and #RS-122-2302, Illumina), and transcriptome capture based libraries with the TruSeq RNA Access Library Prep Kit (#RS-301-2001, Illumina). Recommended amounts of starting material were as follows: 100 ng of input RNA for TruSeq, 100 ng for Ribo-Zero, and 10 ng of intact RNA or 20 ng of degraded RNA for RNA access. All protocols were performed following the manufacturer’s instructions. Libraries were sequenced by Illumina HiSeq2000 resulting in paired 50nt reads. Fastq files were aligned to the hg19 genome assembly using STAR.  

STAR gene counts were normalized applying the median of ratios method implemented in DESeq2 R package. Briefly, the normalization process implies different steps: i) for each gene, a pseudo-reference sample is created and is equal to the geometric mean across all samples; ii) for every gene in a sample and for each sample, the ratios
sample/ref are calculated; iii) the median value of all ratios for a given sample is taken as the normalization factor (size factor) for that sample; iv) for each gene in each sample the normalized count values is calculated dividing each raw count value by the sample’s normalization factor.

Quantitative Traits analysis. Quantitative Traits (QT) consist of “measurable phenotypes that depend on the cumulative actions of many genes and the environment” (www.nature.com/subjects/quantitative-trait). The BRB ArrayTools_v4.6 software was used to identify QT using RNA-seq normalized data and Quantitative Trait Analysis tool (Pearson correlation, cutoff $p>0.8$; $p$-value $<0.05$, univariate test). Random variance model for t statistic was used to improve estimation of variances considering the small sample size of AS11 cohort. Heatmaps and clusters were generated by using Cluster 3.0 for Mac OS X (C Clustering Library 1.56) and Java TreeView version 1.1.6r4 (uncentered correlation and centroid linkage) using median centered gene expression data. Global geometric mean of gene expression (DESeq2 normalized reads) was computed in each QT (only expressed genes were considered i.e., with $>0$ reads) to provide a quantitative estimation of QTs expression across samples (Figure 2B). Analysis of enriched mechanisms in QT gene sets (Figure 2C) was performed by using Molecular Signature Database (MSigDB; www.gsea-msigdb.org/gsea/msigdb/index.jsp) with the CGP gene set collection (CGP collection, $N=3302$ gene sets) and q-value (FDR) $<0.05$ as cutoff for statistical significance of overlapping. Bubble plots were built using JMP 13 software (SAS).

TCGA HGSOC data analysis. We downloaded gene expression data, number of mutated genes and clinical data for a cohort of 307 patients with high-grade serous ovarian adenocarcinoma available in the TCGA data portal (http://cancergenome.nih.gov). Heatmaps and clusters were generated by using
Cluster 3.0 for Mac OS X (C Clustering Library 1.56) with uncentered correlation and centroid linkage, and Java TreeView (version 1.1.6r4; http://jtreeview.sourceforge.net) on median centered gene expression data (genes were retained in the clustering analysis if only their 25th of intensities were less than 100). Gene Set Enrichment Analysis (https://www.gsea-msigdb.org/gsea/index.jsp) was performed using Signal2Noise metric, 1000 random gene sets permutation, and median gene expression values for class comparison. Kruskal-Wallis test and Fisher’s exact test were used to assess association for continuous and categorical variables.

**Immune subpopulation deconvolution.** We used the CIBERSORT and MCP-counter R code 6,7 to calculate the absolute fraction of 22 different immune subpopulations within the bulk cancer population of our samples. Normalized RNA-seq data were used as the input.

**DNA-Seq analysis**

Briefly, 100 ng high-quality genomic DNA was used to prepare the Ion AmpliSeq™ Exome library (Ion AmpliSeq™ Exome RDY, Rev. C.0; MAN0010084; Thermo Fisher Scientific, Inc.). Ion Torrent adapters and amplicons were ligated with DNA ligase. Following AMPure bead purification (Beckman Coulter, Inc., Brea, CA, USA), the concentration and size of the library were determined using the Applied Biosystems® Qubit 3.0 fluorometer (Thermo Fisher Scientific, Inc.). Sample emulsion PCR, emulsion breaking, and enrichment were performed using the Ion PI™ Hi-Q™ Chef Kit (Thermo Fisher Scientific, Inc.), according to the manufacturer’s instructions. An input concentration of one DNA template copy per ion sphere particles (ISPs) was added to the emulsion PCR master mix and the emulsion was generated using the Ion Chef™ System (Thermo Fisher Scientific, Inc.). Template-positive ISPs were enriched, sequencing was performed using Ion PI™ Chip kit v3 chips on the Ion
Torrent Proton, and barcoding was performed using the Ion DNA Barcoding kit (Thermo Fisher Scientific, Inc.). The whole exome DNA-seq analysis yielded an average of ~32 million reads, with an average of 1.8 SNVs/Mb and ~70 de novo mutations.

**Variant calling and annotation.** Fastq data from the Proton runs were initially processed using Ion Torrent platform-specific pipeline software, Torrent Suite v5.0.2 (Thermo Fisher Scientific, Inc.), to generate sequence reads, trim adapter sequences, filter, and remove poor signal-profile reads. The resulting bam files were used to call somatic variants and annotate them applying the Ion Reporter AmpliSeq Exome Tumor-Normal pair workflow v5.10 (Thermo Fisher Scientific, Inc.), using patient-matched blood sample as the normal counterpart. Each SNV was assigned to one of the six basic mutational signatures described by Alexandrov et al. \(^8\) together with indels. The six substitution subtypes are: T>C (comprising A>G and T>C); C>T (comprising G>A and C>T); C>G (comprising G>C and C>G); T>G (comprising A>C and T>G); T>A (comprising A>T and T>A); C>A (comprising G>T and C>A). Standard error of the mean was calculated with R (https://www.R-project.org/).

**Analysis of xenograft-derived sequence read data.** Xenome software was used to estimate mouse reads contamination by analyzing fastq DNAseq data coming from all analyzed samples. From all mapped reads, we retrieve the percentage of reads belonging to human, mouse and both, and we consider as contaminated reads those mapping to mouse and both.

**Variant filtering.** Somatic mutations were filtered in order to obtain exonic and nonsynonymous variants. Among them we selected those that were present in at least one human sample to exclude xenograft-specific mutations, and those that were present in at least two samples in order to exclude possible sequencing errors.
**Somatic copy number alterations.** AmpliSeq Exome single sample workflow v5.10 of Ion Reporter was applied to call copy number variations (CNV) in order to better estimate them by using a panel of normal samples implemented into the workflow. To retrieve somatic CNV for each tumor sample we exclude those present also in the normal sample (blood). CNV were annotated with tumor suppressor genes (TSG), retrieved from the tumor suppressor gene database ⁹, oncogenes (ONC), retrieved from oncoprotein database ¹⁰, and DNA damage repair genes, retrieved from REPAIRtoire database.

**Mutation cluster identification.** We applied the run_analysis_pipeline of PyClone software ¹¹ using default parameters and default settings to cluster mutations according to their variant allele frequency and their genotype. The output of PyClone was subjected to the clonevol package ¹² of R to identify possible clusters of mutations within our samples and to draw fish plot. Mutated genes were annotated with information of the Cancer Gene Census database of the COSMIC project.

**Cancer relevant pathways.** KEGG database was used to map all mutated genes within cellular pathways. Among the identified pathways, we selected the top 10 pathways, and mutated genes were annotated accordingly. For what concerns PI3K pathway, most important genes were selected and mutation, CNV and gene expression data were used to generate Oncoprints using cBio portal (https://github.com/cBioPortal/cbioportal/). Ingenuity Pathway Analysis (IPA; QIAGEN) was used to define interactions between genes belonging to the PI3K pathway. All mutated genes were submitted to PANTHER classification system in order to identify enriched or depleted pathways within Gene Ontology biological process database.
The sequencing coverage and quality statistics for each sample are summarized in Table S9.

References:

1. Francavilla C, Lupia M, Tsafou K, Villa A, Kowalczyk K, Rakownikow Jersie-Christensen R, Bertalot G, Confalonieri S, Brunak S, Jensen LJ, Cavallaro U, Olsen JV. Phosphoproteomics of Primary Cells Reveals Druggable Kinase Signatures in Ovarian Cancer. Cell Rep 2017;18:3242–56.

2. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.

3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

4. Simon R, Lam A, Li M-C, Ngan M, Menenzes S, Zhao Y. Analysis of Gene Expression Data Using BRB-Array Tools. Cancer Inform 2007;3:117693510700300.

5. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003;19:2448–55.

6. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7.

7. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218.

8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415–21.

9. Zhao M, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res 2013;41:D970-976.

10. Liu Y, Sun J, Zhao M. ONGene: A literature-based database for human oncogenes. J Genet Genomics 2017;44:119–21.

11. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A, Shah SP. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014;11:396–8.
12. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, Maher CA. ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol* 2017;28:3076–82.

13. Lupia M, Angiolini F, Bertalot G, Freddi S, Sachsenmeier KF, Chisci E, Kutryb-Zajac B, Confalonieri S, Smolenski RT, Giovannoni R, Colombo N, Bianchi F, et al. CD73 Regulates Stemness and Epithelial-Mesenchymal Transition in Ovarian Cancer-Initiating Cells. *Stem Cell Reports* 2018;10:1412–25.
Supplementary Figure and Table legends

**Figure S1.** STR profile electropherograms of As11 patient’s ascites (top) and the As11 established cell line (bottom).

**Figure S2.** A) Immunofluorescence staining on ascites cells of As11 patient (AS) and on derived primary cell lines at early (PR) passages. The expression of specific ovarian cancer markers (PAX8 and WT1) was investigated (in red). Nuclei were counterstained with DAPI (in blue). Scale bars, 50 μM. B-D) qRT-PCR expression analysis of a panel of stem cells and EMT markers in OCSC vs. PR samples (B), in OCSC+PIK3 inhibitor (LY294002) vs. OCSC+DMSO (as control) (C), in PR+PIK3 inhibitor vs. PR+DMSO (as control) (D). Y-axes, expression fold change (FC). Error bars refer to standard deviations of PCR triplicates.

**Figure S3.** A) Representative images of immunofluorescence analysis on the following As11-derived samples: freshly isolated ascites aggregates (AS), primary cells (PR) at early passage (p3) and the established cell line (CL) at late passage (p21). Cells were stained for the OCSC markers CD44 and CD73 \(^{13}\) and the mesenchymal marker vimentin. Nuclei were counterstained with DAPI. Primary cells showed higher levels of CD44, CD73 and vimentin as compared to the stable cell line, in agreement with the gene expression analysis. Scale bar, 20 μM. B) CIBERSORT analysis of infiltrated immune cell populations in As11 set. Stacked-bar charts colors are as per the legend. C) MCP-counter analysis of infiltrated immune and D) stromal cell (endothelial and fibroblast) populations in As11 set.
**Figure S4. A)** Hierarchical cluster analysis of the TCGA HGSOC cohort of patients using QT-A, -B or -C genes. The cluster of tumors with a higher expression of genes in the considered QT is highlighted. Tumors belonging to the highlighted cluster were labeled accordingly (i.e. QT-A-like, QT-B-like and QT-C-like). Color bar explains the heatmap colors i.e. the log ratio changes of expression over median centered expression values of genes. Bar plots underneath the heatmaps represent samples “QT score” based on the log2 (median centered) sum of ranks assigned to each patient after sorting each gene by ascending log2 ratio changes of expression. **B)** GSEA plots of QT-A, -B and -C across the HGSOC subtypes. NES, normalized enrichment scores. q-values, significance of the enrichment (False Discovery Rate; 1000 random permutations). In red, significant QT positively enriched in the specific HGSOC subtype.

**Figure S5.** Density plot of the selected 2053 mutations. Y-axis, density distribution of mutations. X-axis, mutation frequency expressed in percentage and calculated as the number of mutated reads over the number of total reads covering each position. Analyzed samples are indicated by relative color codes and are as per the legend. Horizontal bar represents the trend of genetic heterogeneity, from high to low.

**Figure S6. A)** Lollipop plots of PIK3R1 and PIK3R2-encoded p85 subunits found mutated in As11 samples. Domain structures and domain names are shown. Amino acid changes are reported on top of the lollipop. The A727T change (in red) was reported to be a germline SNP. Lollipop color indicates type of mutation as per legend. Y-axes, number of sample types with the displayed mutation. X-axes, amino acid position. **B)** Frequency of alterations in genes of the PI3K signaling pathways in the
TCGA cohort of HGSOC (the Oncoprint plot was generated through a cbioportal analysis (https://www.cbioportal.org/). The type of genetic alteration is indicated by relative color codes and are as per the legend. Percentages indicate the fraction of tumor samples in the TCGA HGSOC samples (provisional dataset, \(N=606\)) that carry genetic alterations of the indicated gene.

**Table S1.** Summary of the STR profiles of As11 cells.

**Table S2.** Coverage and quality control of DNA and RNA sequencing of the As11-set.

**Table S3.** List of genes included in the QT-A, -B and -C. Correlation coefficients (\(\rho\)) with QT are displayed as well as the relative parametric p-values.

**Table S4.** List of the top 10 overlapping C2-CGP and H signatures with protein-coding genes in QT-A, -B and -C. List of protein-coding genes in QT-A, -B, and -C is also shown. Relative p-values and FDR are displayed. CGP, chemical and genetic perturbations; H, hallmarks.

**Table S5.** TCGA-HGSOC cohort: patient and clinico-pathological features.

**Table S6.** Univariate and multivariable Cox regression analysis for overall survival (A) or for disease-free survival (B) in various subgroups of patients.
Table S7. A) Characteristics of the 2053 selected mutations. PyClone_cluster_ID, cluster ID of PyClone analysis; Not used, mutation not considered in the PyClone analysis. Allele_coverage, coverage of the reference allele; alt_allele_coverage, coverage of the mutated allele; frequency, mutation frequency; ref, reference allele; omim, Online Mendelian Inheritance of Man database; phylop, PhyloP score that measure the conservation of protein across a wide range of organisms in metagenomics 16S analyses; polyphen, sift, grantham and FATHMM, functional effects of missense mutations provided by SIFT, PolyPhen, Grantham and FATHMM predictors, sift score ranges from 0.0 (deleterious) to 1.0 (tolerated), polyphen score ranges from 0.0 (tolerated) to 1.0 (deleterious), grantham range from 5 (tolerated) to 215 (deleterious), FATHMM score ranges from 0.0 (tolerated) to 1.0 (deleterious); 5000Exomes, population frequency information from the 5000 exomes project. GO#, different Gene Ontology processes map to each mutation. Hallmark, presence of mutated genes in one of the hallmarks of cancer. Role_in_cancer, TSG refers to Tumor Suppressor Gene. B) TP53 status for all samples.

Table S8. Gene Ontology (GO) analysis for biological process enrichment of all mutated genes. Red, enriched biological processes. Blue, depleted biological processes.

Table S9. List of somatic (from chromosome 1 to 22) and germline (chromosome X) copy number variations. ONC, list of oncogenes; TSG, list of tumor suppressor genes.
As11 ascites

Allele Report

Sample 4: 5_G02.fsa Run date and time 01/16/2018-17:31:08 ->01/16/2018-18:14:05

As11 cell line (p.81)

Sample 1: IEO1_A01.fsa Run date and time 09/03/2021-13:34:55 ->09/03/2021-14:12:56
Supplemental Figure S3
Genetic Heterogeneity

Samples

Supplemental Figure S5

Mutation frequency (%)
