nature portfolio

Corresponding author(s): David D. L. Bowtell, Dale W. Garsed

Last updated by author(s): Oct 3, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

WGS
Paired-end Whole Genome Sequencing (WGS) at 150bp was performed on HiSeq X Ten System machines.
Software used for WGS processing:
- FASTQC (v0.11.8)
- FastQ Screen (v0.11.4)
- ea-utils (v1.05)
- BWA mem (v0.7.17-r1188)
- Picard Tools (v2.17.3)
- GATK (v4.0.10.1)
- GATK (v3.8-1-0-gf15c1c3ef)
- FACETS (v0.6.1)

RNA-seq
Libraries were generated using Illumina Stranded mRNA Prep and 150 bp paired-end sequencing was performed on Illumina NovaSeq 6000 instruments.
Software used for RNAseq processing:
- FASTQC (v0.11.8)
- FastQ Screen (v0.11.4)
- ea-utils (v1.05)
- STAR(v2.6.0b)
- Picard Tools (v2.17.3)
- RSeQC (v2.6.4)

Methylation
Tumor DNA was bisulfite converted with the EZ DNA Methylation kit (Zymo Research) and assayed using the Infinium MethylationEPIC

BeadChip arrays according to manufacturer's instructions (Illumina).
Software used for Methylation data processing:
- minfi82 955 (v1.32.0)
- sva (v3.34.0)

SNP array
Tumor and matched normal DNA was assayed with the Infinium OmniExpress-24 BeadChip arrays, arrays scanned and data processed using Genotyping module 2.0.3 software in GenomeStudio 2.0.3 to calculate logR ratios and B-allele frequencies according to manufacturer's instructions (Illumina).
Software used for SNParray data processing:
- HYSYS github version b730498 (https://github.com/PapenfussLab/HaveYouSwappedYourSamples)
- qPure v1.1 (https://sourceforge.net/projects/qpure/)
- ASCAT v2.5.2 (https://github.com/VanLoo-lab/ascat)

Data analysis

WGS
- VarDictJava (v1.5.7 with –r=2 –Q=10 –V =0.05 –f=0.01)
- Mutect2 (v4.0.11.0 with defaults)
- Strelka (v2.9.9 with defaults)
- VarScan2 (v2.4.3 with --min-coverage 7 --min-var-freq 0.05 801 --min-freq-for-hom 0.75 --p-value 0.99 --somatic-p-value 0.05 --strand-filter 0 & SAMtools v1.9 for mpileup)
- vt (v0.57721)
- GATK ReadBackPhasing (v3.8-1-0-gf15c1c3ef with --phaseQualityThresh 10 --enableMergePhasedSegregatingPolymorphismsToMNP --min_base_quality_score 10 min_mapping_quality_score 10 --maxGenomicDistanceForMNP 2)
- GATK CombineVariants (v3.8-1-0-gf15c1c3ef with -genotypeMergeOptions UNIQUIFY -priority Strelka2, Mutect2, VarScan2, VarDictJava)
- GATK VariantAnnotator (v3.8-1-0-gf15c1c3ef with --reference_window_stop 1000 -A HomopolymerRun -A TandemRepeatAnnotator)
- BCFtools (v1.9)
- bam-readcount (v0.8.0 with -w 0 --min-mapping-quality 10 –
- min-base-quality 10 --max-count 100,000,000)
- Ensembl Variant Effect Predictor (VEP v92.4)
- Manta + BreakPointInspector (v1.5.0)
- GRIDSS (v2.0.1)
- Smoove (v0.2.2)
- SvABA (v134)
- StructuralVariantAnnotation (v1.3.1 with maxgap = 10, ignore.strand = FALSE)
- InteractionSet (v1.14.0)
- rtracklayer (v1.46.0)
- snp-pileup(v1.0 with --pseudo-snps 100 --min-map830 quality 10 --min-base-quality 10 --max-depth 5000 --min-read-counts 15,0)
- cnv_facets (v0.13.0 with --nbhd snp=500 --cval=50 1000 --depth=15 5000)
- CHORD (v2.00)
- csaw (v1.20.0)
- ICAMS v2.0.10.9001
- rstatix (v0.7.0)
- signature.tools.lib (v0.0.0.9000 with SignatureFit_withBootstrap [method = "KLD", nboot = 100, randomSeed = 42, threshold_percent = 2, threshold_p.value = 0.05])
- ConsensusClusterPlus (v1.50.0 with maxK = 10, reps=1000, pItem=0.9, pFeature=0.9, clusterAlg="pam", distance="pearson", innerLinkage="ward.D2", finalLinkage="ward.D2", seed=12345678)
- seriation (v1.3.0 with method="OLO")
- HLA-VBSeq (v11_22_2018)
- Jvarkit samviewwithmate (ec2c236)
- samtools (v1.9)
- deepTools (v3.0.0 with parameters --binSize 10 --minMappingQuality 10 --normalizeUsing CPM --skipNonCoveredRegions --samFlagExclude 1024 --outFileFormat bigwig)
- pVACtools pVACseq (v1.3.5)
- dNdScv (v0.0.1.0 with refdb = "hg19", sm = "192r_3w", max_muts_per_gene_per_sample = Inf, max_coding_muts_per_sample = Inf, use_indel_sites= FALSE)
- GRIN (v1.4)
- GISTIC2 (v2.0.23 with -savegene 1 -maxspace 1000 -ta 0.1 -td 0.1 -rx 0 -cap 3 -broad 0 –twoside 1 -res 0.05 -genegistic 0 -v 10)
- maftools (v2.2.10)
- Mutalyzer (v2.0.35)
- R (v3.6.3)
- Prism (v9.2.0)

RNA-seq
- HTSeq (v0.10.0 with mode = "intersection-nonempty")
- edgeR (v3.28.1)
- limma (v3.42.2)
- DeepCC (v0.1.1)
- DESeq2 (v1.26.0)
- fGSEA (v1.15.1)
- CIBERSORTx web version as at 05/21/2020 (https://cibersortx.stanford.edu/)
- ConsensusClusterPlus (v1.50.0)
- Arriba (v1.1.0)
- R (v3.6.3)

Methylation
- limma (v3.42.2)

- R (v3.6.3)

Feature correlation
- polycor (v0.8-1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

ICGC dataset: Previously published WGS and RNA-seq data generated as part of the ICGC Ovarian Cancer project14 are available from the European Genome-phenome Archive (EGA) repository (https://ega-archive.org) as a single bam file for each sample type (tumor/normal), under the accession code EGAD00001000877 (https://ega-archive.org/datasets/EGAD00001000877). Due to the sensitive nature of these patient datasets, access is subject to approval from the ICGC Data Access Compliance Office (https://docs.icgc.org/download/data-access/), an independent body who authorizes controlled access to ICGC sequencing data. ICGC SNP array and methylation data sets have been deposited into the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession code GSE65821 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65821), without access restrictions. ICGC gene count level transcriptomic data has been deposited into the GEO under accession code GSE209964 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209964).
MOCOG dataset: WGS, RNA-seq and SNP array data from long-term survivors generated as part of the MOCOG study have been deposited in the EGA repository under accession code EGAS00001005984 (https://ega-archive.org/studies/EGAS00001005984). WGS and RNA-seq data are available as raw FASTQ files for each sample type (tumor/normal) and SNP array data are available as raw signal intensity files in text format for each sample type (tumor/normal). Access to patient sequence data can be gained for academic use through application to the independent Data Access Committee (dac@petermac.org). Responses to data requests will be provided within two weeks. Information on how to apply for access is available at the EGA under accession code EGAS00001005984. The MOCOG cohort raw methylation data sets have been submitted to the GEO under accession code GSE211687 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211687), with no access restrictions.
Uniformly processed somatic variant data from the ICGC and MOCOG cohorts is deposited in Synapse under accession code syn34616347 (https://www.synapse.org/#!Synapse:syn34616347), and processed expression and methylation data from both cohorts has been submitted into the GEO under accession code GSE211687 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211687), without access restrictions.
Population frequencies of genetic variants can be accessed via the Genome Aggregation Database (gnomAD) at https://gnomad.broadinstitute.org/. Supporting evidence for pathogenicity of genomic alterations can be accessed via ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), BRCA Exchange (https://brcaexchange.org/) and the TP53 Database (https://tp53.isb-cgc.org/). The Ensembl ranked order of severity of variant consequences is available at: https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html. Precomputed TCGA ovarian serous cystadenocarcinoma survival analysis data can be downloaded from OncoLnc (http://www.oncolnc.org/). Mutational signature reference databases can be accessed via COSMIC (https://cancer.sanger.ac.uk/signatures/) and Signal (https://signal.mutationalsignatures.com/). The LM22 signature matrix used for immune cell deconvolution can be downloaded here: https://cibersortx.stanford.edu/. The COSMIC Cancer Gene Census can be accessed here: https://cancer.sanger.ac.uk/census. MSigDB hallmark gene sets can be accessed here: https://www.gsea-msigdb.org/gsea/msigdb/. Illumina methylation probes that were filtered out due to poor performance (e.g. cross reactive or non-specific probes) can be found here: https://github.com/sirselim/illumina450k_filtering. Germline polymorphic sites for reference and variant allele read counts used in FACETS analysis can be found at ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/common_all_20180423.vcf.gz. The GTF used for annotation and RNA-seq counts is available here: ftp://ftp.ensembl.org/pub/grch37/release-92/. All other data are available within the article and its supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. A limited number of cases were available given the rarity of high-grade serous ovarian cancer long-term survivors, with fresh frozen tumor tissue and matched germline samples available.
Data exclusions	Two tumor samples were flagged as outliers with a high somatic mutation rate (>20 mutations/Mb): AOCS-076 due to cross sample contamination and AOCS-166 due to a germline mutation in the mismatch repair gene PMS2. The two samples were therefore excluded from further analyses.
Replication	Due to the rarity of tumor tissue from long-term survivors and the high cost of genomic assays, WGS, RNA-sequencing, SNP arrays and methylation arrays were performed once for each sample. Supporting evidence for variants detected by WGS were sought in orthogonal datasets, such as SNP array, clinical sequencing, RNA-sequencing or DNA sequencing of matched samples (e.g. germline and tumor) from the same patient (Supplementary Tables 3, 4 and 5). CCNE1 amplification status determined by WGS data was verified by FISH and immunohistochemistry data available from a previous study (Aziz et al 2018). CIBERSORTx immune scores were verified with CD8+ T cell scores determined by multi-color immunohistochemistry and automated image scoring from our previous study (Garsed et al 2018), which confirmed the identification of the two long-term survival immune clusters (IMM.1 and IMM3). Mutational signature analyses were informed

by and/or consistent with previous analyses of breast, ovarian and pan-cancer genomic datasets (Alexandrov et al 2013, Patch et al 2015, Popova et al 2016, Alexandrov et al 2020, and Degasperi et al 2020). DNA methylation clustering identified a BRCA1-altered subset (MET.1) that overlapped with the poor outcome BRCA1 mutational signature subgroup (SIG.5), providing an independent marker of this group. To our knowledge, this is the first WGS, DNA methylation and RNA-seq analysis of HGSC long-term survivors, so there is currently no equivalent dataset available in which to validate the genomic and immune signatures, however where possible we sought validation of findings in an independent dataset (The Cancer Genome Atlas Research Network et al 2011).

Randomization

Randomization was not applicable as this study did not involve an intervention. Long-term survivors were compared with patients with similar known survival predictors: histology, grade, age at diagnosis, stage, treatment, etc. (Extended Data Fig. 1b).

Blinding

Blinding was not undertaken. Genomic and clinical data was uniformly processed and analyzed across all samples/cases.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics

Patients were diagnosed with advanced stage (FIGO Stage IIIC/IV), high-grade serous cancer (ovarian, fallopian tube or primary peritoneal carcinoma) confirmed by histopathology review, received primary chemotherapy incorporating a platinum-based compound, and met survival group criteria; either short-term survivors (death within 2 years); moderate-term survivors (death ≥2 and <10 years); or long-term survivors (overall survival ≥ 10 years). Population characteristics are provided in Extended Data Fig. 1 and Supplementary Table 1. All 126 patients are female and were aged 29-81 years.

Recruitment

Patients were recruited to the population-based Australian Ovarian Cancer Study (AOCS), Peter MacCallum Cancer Centre (Melbourne, Australia), the Gynaecological Oncology Biobank (GynBiobank), Westmead Hospital (Sydney, Australia) or the Mayo Clinic (Rochester, USA) under Research Ethics Committee approval for each study. Written informed consent was provided by all patients. Participation was voluntary, with no compensation provided to participants.
1. Patients over the age of 18 (and under 80 for AOCS) with invasive epithelial ovarian cancer (or primary peritoneal or fallopian tube cancer) were recruited through major cancer treatment centers across Australia (AOCS and GynBiobank) and in the USA (the Mayo Clinic), generally at the time of initial diagnosis.
2. Surgical resection samples: Tissue samples for research were taken from tissue removed during surgery and were excess to diagnostic requirements. Sampling of surgical tissue was undertaken by a trained pathologist to ensure that samples taken for the purposes of research did not affect histopathological assessment. Samples were snap-frozen and stored for future research, with patient consent.
3. Surgical, pathology, systemic treatment and clinical outcome data was collected from medical records during regular longitudinal clinical follow-up. Cases were selected for the current study based on pre-specified criteria, including advanced stage disease (FIGO Stage IIIC/IV), histology (HGSC), primary treatment (platinum-based chemotherapy), availability of biospecimens for analysis (snap frozen tumor tissue and blood), and survival within the three specified survival groups (death within 2 years; death ≥2 and <10 years; overall survival ≥ 10 years). This may represent a potential bias towards patients with tumors of sufficient size, and a high proportion of tumor cells within the research sample, to enable analyses.
AOCS recruited patients <80 years old, which could result in an increased proportion of long-term survivors compared to studies without an upper age limit, because the chance of surviving >10 years if diagnosed over 80 would be lower (due to older age), however ages are consistent between survival groups and not likely to impact the results. No other sources of selection bias were identified.

Ethics oversight

The Peter MacCallum Cancer Centre Human Research Ethics Committee (HREC), the Western Sydney Local Health District HREC, and the Mayo Clinic Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.