nature portfolio

Corresponding author(s): Sohrab Shah, Andrew McPherson

Last updated by author(s): May 4, 2025

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	scWGS pipeline: https://github.com/mondrian-scwgs/mondrian scRNA-seq pipeline: https://github.com/shahcompbio/scrna-pipeline WGS pipeline: https://github.com/shahcompbio/wgs
Data analysis	SIGNALS: https://github.com/shahcompbio/signals doubleTime: https://github.com/shahcompbio/doubleTime CINner: https://github.com/dinhngockhanh/CINner Figures: https://github.com/shahcompbio/spectrum_wgd_paper

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

The MSK SPECTRUM study is registered on dbGaP under accession number phs002857.v3.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002857.v3.p1).

A Synapse page for the MSK SPECTRUM WGD study is available to provide access to multi-modal datasets from one central location. This page can be accessed under accession number syn66366718 (https://www.synapse.org/Synapse:syn66366718/wiki/632295).

Data availability:
• scWGS:
- Raw sequencing reads are available for controlled access from the NCBI Sequence Read Archive via dbGaP.
- Processed objects are available from Synapse (https://www.synapse.org/Synapse:syn66366960/datasets/).
- Single-cell copy number and somatic mutation profiles are available as part of an HTML visualization hosted on GitHub: https://shahcompbio.github.io/spectrum_wgd_paper/cohort_visualization.html
• scRNA-seq:
- Raw and processed expression data are available from NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180661).
- Processed objects are available from Synapse (https://www.synapse.org/#!Synapse:syn33521743/datasets/ and https://www.synapse.org/Synapse:syn66477498/datasets/).
• Tumor-normal bulk WGS:
- Raw sequencing reads are available for controlled access from the NCBI Sequence Read Archive via dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002857.v3.p1).
- Somatic mutations and copy number data can be accessed from Synapse (https://www.synapse.org/#!Synapse:syn33521770/datasets/).
- Somatic mutations and copy number are available for visualization through cBioPortal (https://www.cbioportal.org/study/summary?id=msk_spectrum_tme_2022).
• Tumor-normal targeted panel sequencing (MSK-IMPACT):
- Somatic mutations and copy number are available for visualization through cBioPortal (https://www.cbioportal.org/study/summary?id=msk_spectrum_tme_2022).
• IF:
- Properties of identified objects are available from Synapse (https://www.synapse.org/Synapse:syn66366961/datasets/)
- Sample-level and ROI-level statistics, sample metadata, antibodies, and staining conditions for IF datasets are included as Supplementary Tables.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study focuses on high-grade serous ovarian cancer, which is inherently a female-specific disease as it affects the female reproductive system (ovaries and fallopian tubes).
Reporting on race, ethnicity, or other socially relevant groupings	No categorization variables such as race, ethnicity or other socially relevant groupings were used in this study.
Population characteristics	The study cohort (MSK SPECTRUM) includes 41 women with newly diagnosed, treatment-naive high-grade serous ovarian cancer (HGSOC). Patients between the ages of 39 and 81 at diagnosis (median age: 59 years). 12 out of 41 cases had BRCA1 mutations (29%) and 7 out of 41 cases had a BRCA2 mutation (17%).
Recruitment	All enrolled patients were consented to an institutional biospecimen banking protocol and a protocol to perform targetted panel sequencing (MSK-IMPACT). All analyses were performed per a biospecimen research protocol. All protocols were approved by the Institutional Review Board (IRB) of Memorial Sloan Kettering Cancer Center (MSKCC). Patients were consented following the IRB-approved standard operating procedures for informed consent. Written informed consent was obtained from all patients before conducting any study-related procedures. This study was conducted in accordance with the Declaration of Helsinki and the Good Clinical Practice guidelines (GCP).
Ethics oversight	Institutional Review Board (IRB) at Memorial Sloan Kettering Cancer Center (MSKCC).

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

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	Quality-filtered study datasets (MSK SPECTRUM cohort): • scWGS: 41 patients, 70 samples, 30,260 cells • scRNA-seq: 32 patients, 125 samples, 751,300 cells • IF: 37 patients, 102 samples, 20,988,413 primary nuclei, 896,042 ruptured micronuclei • Bulk tumor WGS: 33 patients, 33 samples • Bulk normal WGS: 41 patients, 41 samples • Tumor-normal targeted panel sequencing (MSK-IMPACT): 41 patients, 41 samples
Data exclusions	Low-quality cells were removed from scWGS and scRNA analyses, and fields of view with scant tissue were removed from IF analyses, as described in the methods.
Replication	Genomic profiling estimates of ploidy were consistent between scWGS, bulk WGS and targeted panel sequencing (MSK-IMPACT). Cellular ploidy reflected the number of WGD events per cell detected by DLP+, and correlated with both cell size measured through the optical components of DLP+, and mitochondrial copy number, providing orthogonal validation based on known correlates of nuclear genome scaling. Rates of chromosomal instability were found to be concordant as measured by scWGS and IF. Specifically, the rate of micronucleus rupture measured by IF was significantly elevated in WGD tumors. Cell cycle alterations resulting from increased chromosome missegregation in WGD-low tumors were replicated in diploid TP53-/- RPE1 cells, which were treated with nocodazole, reversine and DMSO control in validation experiments to induce varying levels of chromosomal instability.
Randomization	Patients were stratified into groups based on prevalence of malignant WGD cells detected by scWGS, as described in the manuscript.
Blinding	Group allocation was done by assigning patients to WGD classes based on data analysis described in the manuscript.

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
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Antigen Antibody Clone Manufacturer Titration DAPI D9542 Sigma 5.0 µg/ml CD8 790-4460 Ventana Roche 0.07 µg/ml panCK M3515 DAKO 1 to 500 cGAS 7997 Cell Signaling 1.25 µg/ml p53 ab32389 Abcam 0.005 µg/ml STING 13647 Cell Signaling 0.075 µg/ml
Validation	Primary antibody staining conditions were optimized using 4 µm formalin-fixed, paraffin-embedded tissue sections and serial antibody titrations to determine the optimal antibody concentration.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Telomerase-immortalized retinal pigment epithelial cell lines (RPE-1) were purchased from the American Type Culture Collection (ATCC, CRL-4000). TP53-knockout RPE-1 cells were a gift from the Maciejowski Laboratory at MSKCC, and originally also obtained from the ATCC. Fallopian tube epithelial (FNE1) cells were a generous gift from Dr. Tan Ince at Weill Cornell.
Authentication	All cell lines used in this manuscript were authenticated by ATCC which used morphology, karyotyping and PCR-based

Authentication	techniques.
Mycoplasma contamination	All cell lines were periodically tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Not applicable.
Study protocol	The clinical study was conducted under the purview of MSKCC institutional tissue banking protocol 06-107 titled "Storage and research use of human biospecimens". Data generation and data analysis were carried out under MSKCC protocol 15-200 titled "Chemotherapy, somatic mutations, neoantigens, and the immune environment in ovarian cancer". Protocol 15-200 operates by using specimens banked under protocol 06-107. Due to MSKCC standard operating policies, internal protocol documents cannot be shared publicly but are available upon request to the IRB.
Data collection	Patients were enrolled on the study at MSKCC in New York, USA. Sample collection took place between January 2019 and March 2021.
Outcomes	No patient outcomes were reported.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Viably frozen dissociated cells used for scWGS were thawed and stained with a mixture of GhostRed780 live/dead marker (TonBo Biosciences) and Human TruStain FcX™ Fc Receptor Blocking Solution (BioLegend). The stained samples were then incubated and stained with Alexa Fluor® 700 anti-human CD45 Antibody (BioLegend). Post staining, they were washed and resuspended in RPMI + 2% FCS and submitted for cell sorting. The cells were sorted into CD45 positive and negative fractions by fluorescence assisted cell sorting (FACS) on a BD FACSAria™ III flow cytometer (BD Biosciences). Positive and negative controls were prepared and used to set up compensations on the flow cytometer. CD45- cells were sorted into tubes containing RPMI + 2% FCS for scWGS. scRNA-seq data analyzed in this study was previously generated from fresh tissue using a similar sorting protocol described in Vázquez-García, Uhlitz et al. Nature (2022).
Instrument	BD FACSAria™ III
Software	BD FACSDiva™ Software
Cell population abundance	After gating for live cells, the fraction of CD45+ and CD45- cells was determined out of the total number of live cells.

Gating strategy | Preliminary FSC/SSC gating of live cells was followed by gating of CD45+ and CD45- populations of live cells.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.