nature portfolio

Corresponding author(s): Gavin Ha, Andrew Hsieh, Hung-Ming Lam, and Omar Mian

Last updated by author(s): Jul 29, 2026

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

Illumina NovaSeq 6000, NovaSeq X Plus, NovaSeq X, HiSeq 2500, and HiSeqX sequencers with on-instrument control and Real-Time Analysis (RTA v2.7.3) software was used for data collection (primary base calling) of sequencing datasets.
IncuCyte software (version 2023A Rev1, Sartorius) was used for data collection (live-cell image acquisition) for the cell confluence and death experiments.

Data analysis

Software pipeline configurations and custom code written for this manuscript are accessible at https://github.com/GavinHaLab/BLCA-subtype-evolution-paper. Data analysis was performed using softwares, libraries, packages, and modules as listed in the methods section:
BWA-MEM (v0.7.17; https://github.com/lh3/bwa)
Picard (v2.18.29, v2.25.1; https://github.com/broadinstitute/picard)
Samtools (v1.10; https://github.com/samtools/samtools)
GATK (v4.1.8; https://gatk.broadinstitute.org/hc/en-us)
Snakemake (v5.19.2; https://snakemake.github.io/)
Java (v11.0.2; https://www.java.com/en/)
Mutect2 (v4.1.8; https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2)
Strelka2 (v2.9; https://github.com/Illumina/strelka)
VarScan2 (v2.4.4; https://dkoboldt.github.io/varscan/)
MuSE (v1; https://github.com/wwylab/MuSE)
ANNOVAR (v2020-06-07) (https://www.openbioinformatics.org/annovar/annovar_download_form.php)
Manta (v1.6.0; https://github.com/Illumina/manta)
SvABA (v1.0.1; https://github.com/walaj/svaba)
TITAN (v1.15.0; https://github.com/gavinha/TitanCNA)

ichorCNA (v0.3.4; https://github.com/broadinstitute/ichorCNA)
PyClone-VI (v0.1.1; https://github.com/Roth-Lab/pyclone-vi;commit ID: a090529)
Bam-readcount (v 1.0.1 ;https://github.com/genome/bam-readcount, commit ID: c7c76e6)
LICHeE (v1.0, https://github.com/viq854/lichee; commitID:26c2a70)
cloneMap (https://github.com/amf71/cloneMap; commitID: 7628d70)
MACHINA (v1.2, https://github.com/raphael-group/machina; commit ID: ac71282)
Comut (v0.0.3; https://github.com/vanallenlab/comut)
Graphviz (v2.40.1; https://graphviz.org/)
GRIDSS2 (v2.13.2; https://github.com/PapenfussLab/gridss)
JaBaA (v.1.1, https://github.com/mskilab-org/JaBbA; commitID: e3dc184)
gGnome (v0.1)
AmpliconArchitect (v.1.2.2, https://github.com/AmpliconSuite/AmpliconSuite-pipeline; commitID: 7ae7f32)
dNdScv (v.0.010, https://github.com/im3sanger/dndscv)
SVclone (v.1.1.3, https://github.com/mcmero/SVclone; commit ID: 93dafb2)
SigProfilerAssignment (v0.1.1; https://github.com/AlexandrovLab/SigProfilerAssignment; commit ID: 327fb93)
FastQC (v0.12.1; https://github.com/s-andrews/FastQC)
STAR2 (v2.7.3a; https://github.com/alexdobin/STAR)
Htseq-count (v0.11.1; https://htseq.readthedocs.io/en/release_0.11.1/count.html)
DESeq2 (v1.48.1; https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
Umap (v0.2.10.0; https://cran.r-project.org/web/packages/umap/index.html)
Lme4 (v1.1-36; https://cran.r-project.org/web/packages/lme4/index.html)
lmerTest (v3.1-3; https://cran.r-project.org/web/packages/lmerTest/index.html)
Seurat (v5.2.1; https://satijalab.org/seurat/)
GSVA (v1.52.3; https://github.com/rcastelo/GSVA)
consensusMIBC (v1.1.0; https://github.com/cit-bioinfo/consensusMIBC)
CIBERSORTx (https://cibersortx.stanford.edu/)
Fgbio (v2.1.1; https://github.com/fulcrumgenomics/fgbio)
Griffin (v0.2.0; https://github.com/GavinHaLab/Griffin; commit ID: 6dfe39d)
CellRanger (v7.1.0; https://github.com/10XGenomics/cellranger)
SingleR (v2.6.0; https://bioconductor.org/packages/release/bioc/html/SingleR.html)
Tumor Immune Cell Atlas (https://singlecellgenomics-cnag-crg.shinyapps.io/TICA/)
Fgsea (v1.30.0; https://github.com/alserglab/fgsea)
entropy (v1.3.1; https://cran.r-project.org/web/packages/entropy/index.html)
nichenetr (v2.1.7; https://github.com/saeyslab/nichenetr)
Numbat (v1.4.2; https://github.com/kharchenkolab/numbat)
R (v4.4.3; https://www.r-project.org/)
Python (v3.19, v3.7.4; https://www.python.org/)
Pandas (v1.40; https://pypi.org/project/pandas/)
Ggplot2 (v3.5.1; https://cran.r-project.org/web/packages/ggplot2/index.html)
Rstatix (v0.7.2; https://cran.r-project.org/web/packages/rstatix/index.html)
Survival (v3.8-3; https://cran.r-project.org/web/packages/survival/index.html)
Survminer (v0.5.0; https://cran.r-project.org/web/packages/survminer/index.html)
Forestmodel (v0.6.2; https://cran.r-project.org/web/packages/forestmodel/index.html)
MESS (v0.6.0; https://cran.r-project.org/web/packages/MESS/index.html)
Gurobi version 11.0 with a WLS license.
MIM software v7.3.7 (Build O606-01, Cleveland) for CT scan analysis
ImageJ (v1.53k) for Western blot analysis
Imaris (v11.0.1) for γH2AX analysis
IncuCyte software (version 2023A Rev1, Sartorius) for cell growth and death analysis

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

Whole exome and whole genome sequencing (WGS) data are accessible at dbGAP under accession phs001797.v2.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001797). Bulk RNA-seq and single nucleus RNA-seq data are accessible in the Gene Expression Omnibus (GEO) under accession GSE302273 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE302273). The Hartwig Medical Foundation (HMF) WGS data and mutation call set were obtained via controlled access, available only with permission from the HMF data access request (https://www.hartwigmedicalfoundation.nl/en/data/data-access-request/). Publicly available bladder cancer datasets from The Cancer Genome Atlas (https://portal.gdc.cancer.gov/projects/TCGA-BLCA) were accessed through the Genomic Data Commons (GDC) Data Portal and analyzed in this study. Publicly available cisplatin sensitivity and transcriptomic data for bladder cancer cell lines analyzed as part of this study were obtained from the Cancer Dependency Map (DepMap release 23Q4, https://doi.org/10.25452/figshare.plus.24667905.v2), including the Genomics of Drug Sensitivity in Cancer (GDSC2) dataset (https://depmap.sanger.ac.uk/documentation/datasets/drug-sensitivity/). All other processed data necessary to reproduce the results are found in the Supplementary Tables and at https://github.com/GavinHaLab/BLCA-subtype-evolution-paper. All analyses were performed using the human reference genome GRCh38 (Broad version; https://console.cloud.google.com/storage/browser/gcp-public-data--broad-references/hg38/v0).

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	Patients of both male (n=15) and female (n=5) sexes were included in this study. The higher proportion of males in our cohort is consistent with the higher incidence of bladder cancer in males overall. Sex and gender were not considered in study design. Sex of patients was determined by self-reporting and verified in DNA sequencing analyses. No findings were exclusive to one sex; all analyses were performed on the combined group of patients due to the low number of females and focus on histological subtypes.
Reporting on race, ethnicity, or other socially relevant groupings	Race was self-reported and self-attributed by the patients and collected by clinicians. This is reported in Supplementary Table S2. There was minimal variability in race across the patient population (19 Caucasian, 1 Asian), so it was not controlled for as a confounding variable and no further downstream analysis was performed using this information.
Population characteristics	All population characteristics, including age, tumor characteristics, and treatment history are available in Supplementary Table 2.
Recruitment	Patients were asked by their primary oncologist if they were interested in participating in the study if they had urothelial carcinoma with three or more tumors in the body, with a deliberate enrichment for patients who had tumors with variant histological differentiation. This self-selects for patients with aggressive disease and those interested in contributing to biomedical research. Patients who consent for rapid autopsies tend to be very motivated and have a higher likelihood of participating in clinical trials. This did not impact the results of our study.
Ethics oversight	The University of Washington IRB 2341 approved the study protocol.

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	Sample size was determined by the number of patients willing to consent to be a part of the rapid autopsy program from 2015-2020 and the availability of funding for this program. No statistical method was used to predetermine sample size.
Data exclusions	Four DNA sequencing samples 17-047pD10, 18-101M3, 18-120pA14 and 19-044pB11 were not included in any analysis involving somatic mutations, as they had tumor fraction lower than 20% and highly noisy genome wide copy number profiles. Additionally, for any CNA specific downstream analysis the samples 17-030pM9 and 19-022pJ13-B were excluded due to noisy copy number profiles and can lead to inaccurate founder, shared and private copy number status. One RNA sequencing sample (18-016G2) was excluded from all RNA-seq analysis due to a low RIN score of the extracted RNA. The 8 FFPE RNA-seq samples were excluded from analyses involving a small subset of genes (FANC scores, GSVA scores of snRNA-seq identified pathways) due to the artifacts often present in FFPE tissue. Hypermutated patient 17-047 was excluded from select analyses of mutation and CNA burden, as well as from associations between pathogenic alterations and time to death or metastasis, due to being an extreme outlier in mutational burden and a statistical outlier for overall survival. Patient 16-070 was excluded from analysis of survival from first metastasis to death due to being a statistical outlier for this survival metric.
Replication	All attempts at experimental replication were successful. Western blotting was performed in three to four biological replicates to confirm consistency. Cell growth assays were performed in three to four experimental/biological replicates, with five technical replicates per condition. Cell death assays were performed in two to four experimental/biological replicates, with five technical replicates per condition. Quantification of all biological replicates was shown in the manuscript and statistically analyzed to confirm successful replication of the experimental results. γH2AX staining was performed in one experimental replicate, with hundreds to thousands (n = 537-1395) of nuclei analyzed per condition to confirm consistency of the results.
Randomization	There was no allocation of samples or participants into different experimental groups in this study given its retrospective nature.
Blinding	Blinding is not relevant to our analyses as we only studied genomic and molecular characteristics of the samples and patients and did not apply experimental interventions.

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

rabbit anti-FANCD2 (Abcam; 108928; 1:1,000)

mouse anti-FANCF (Santa Cruz ; sc-271952; 1:500)

mouse anti-Tubulin (Sigma-Aldrich; T8203; 1:1,000)

rabbit anti-Vinculin (Cell Signaling Technology; 13901S; 1:1,000)

goat anti-rabbit (Invitrogen; 31460; 1:5,000)

goat anti-mouse (Invitrogen; 31430; 1:5,000)

mouse anti-γH2AX (Sigma, 05-636; 1:4,000)

Alexa Fluor 633 goat anti-mouse secondary antibody (ThermoFisher; A-21052; 1:1,000)

Validation

All antibodies were validated by their respective manufacturers, as stated below:

Anti-FANCD2 antibody [EPR2302] (ab108928) is a rabbit recombinant monoclonal antibody and is validated for use in Western Blot (WB) in Human, Mouse, Rat samples. The specificity of Anti-FANCD2 antibody [EPR2302] (ab108928) has been confirmed by Western blot testing in FANCD2 Knockout HAP1 cells.

FANCF Antibody (D-2) is a mouse monoclonal IgG1 kappa light chain antibody that detects FANCF protein of mouse, rat, and human origin by western blotting (WB), immunoprecipitation (IP), immunofluorescence (IF), and enzyme-linked immunosorbent assay (ELISA).

Mouse anti-Tubulin (Sigma-Aldrich; T8203) has been used in various immunochemical techniques including immunoprecipitation, immunocytochemistry, and immunohistochemistry and gone through Sigma's independent enhanced validation.

Cell Signaling Technology Vinculin (E1E9V) Rabbit Monoclonal Antibody (13901S) recognizes endogenous levels of total vinculin protein. This antibody also reacts with metavinculin, a 145 kDa splice variant of vinculin. This antibody is for Western Blotting, IHC, and flow use in human, mouse, rat, and monkey.

Goat anti-rabbit (Invitrogen; 31460) has been validated for use in Western blot, immunohistochemistry (IHC), ELISA, and immunoprecipitation (IP) applications.

Goat anti-mouse (Invitrogen; 31430) has been validated for use in Western blot, immunohistochemistry (IHC), ELISA, and immunoprecipitation (IP) applications.

Anti-phospho-Histone H2A.X (Ser139), clone JBW301 (Sigma, 05-636) is a well published Mouse Monoclonal Antibody validated in ChIP, ICC, IF, WB. This purified mAb is highly specific for phospho-Histone H2A.X (Ser139) also known as H2AXS139p.

Alexa Fluor 633 goat anti-mouse secondary antibody (ThermoFisher; A-21052) is used for applications of IHC, ICC/IF, and flow cytometry. To minimize cross-reactivity, these goat anti-mouse IgG (H+L) whole secondary antibodies have been affinity purified and cross-adsorbed against bovine IgG, goat IgG, rabbit IgG, rat IgG, human IgG, and human serum. No nonspecific staining was observed with the secondary antibody alone, or with an isotype control.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

The cell lines used in this study were SW780, HT1376, BFTC905, UBLC1, and HEK293T. SW780, HT1376, BFTC905, and UBLC1 were derived from female patients with bladder cancer. HEK293T cells originated from human embryonic kidney cells. All cell lines were obtained from the American Type Culture Collection (ATCC).

Authentication	The cell lines were authenticated using Short Tandem Repeat (STR) profiling by Promega.
Mycoplasma contamination	All cells tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A