nature research

Corresponding author(s): Xu Gao, Ting Wang & Yinghao Sun

Last updated by author(s): Jan 3, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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

The tumor and matched normal samples from 208 Chinese prostate cancer patients were collected by ourselves. Our cohort was consisted of 416 WGS, 268 RNA-seq, 374 WGBS, and 210 miRNA-seq datasets. Sequencing was performed using an Illumina HiSeq X TEN instrument running HiSeq Control software v3.5.0.
We also collected existing public datasets from 2,554 PCa representing 13 Western cohorts as well as our pilot Chinese cohort published in European Urology. All the access codes for the public datasets have been shown in Supplementary Data 1.
To obtain new FOXA1 binding sites experimentally validated in prostate tumor cell line, we collected the following ChIP-seq data from GSE123618: GSM3508089, GSM3508092, GSM3508095, GSM3508098, and GSM3508101.
No special software was used to download the public datasets.

Data analysis

Here we listed the versions of all the software we used and all the details can be found in the Methods section. All the code we used has been deposited in our supporting website (http://www.cpgea.com).
Sequencing reads were aligned to the Human Genome Reference Consortium build 38 (GRCh38) using BWA v0.7.8.
BWA v0.7.8 (BWA-mem);
TopHat v2.0.9;
Bowtie2 v2.0.6;
Bismark v0.16.3;
miRBase20.0;
mirdeep2 v1.1;
RepeatMasker v4.0.3;
Picard v1.111;
Genome Analysis Toolkit (GATK) v3.1;
GATK's HaplotypeCaller;
MuTect v1.1.4;
Strelka v1.0.13;
SOAPfuse v1.27;

ANNOVAR (Sun, 22 Mar 2015);
SomaticSignatures package v2.9.4;
Control-FREEC v6.7;
GISTIC 2.0 v6.2 GenePattern module;
MeerKat v0.189;
Shatterseek v0.4;
ChainFinder v1.0.1;
MutSigCV v1.4;
MuSiC v0.04;
funseq2 v2-1.6;
denovoTF;
Cuffdiff v1.3.0;
VMD v1.9.3;
FastQC v0.11.5;
Trimmomatic v0.36;
MethPipe v3.4.3;
DSS v2.14.0;
BEDTools v2.27.1;
GENCODE Release 27;
GeneHancer v4.7;
GREAT v3.0.0; and
iClusterplus v1.16.0

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All data, including raw data, mutation calls, and clinical information, were deposited to Genome Sequence Archive for Human (http://bigd.big.ac.cn/gsa-human/) at BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under project number PRJCA001124, GSA number HRA000099. The data deposited and made public is compliant with the regulations of Ministry of Science and Technology of the People's Republic of China.

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 sizes were determined in order to obtain more than 200 normal-tumor pairs. All primary tumor and matched normal tissue from 208 patients were used to generate sequencing data in this study. In all, 1,268 datasets including 416 WGS, 268 RNA-seq, 210 miRNA-seq, and 374 WGBS were used in this study.
Data exclusions	Two tumor cases exhibited extreme hypermutation phenotypes and were determined to contain mutations in DNA repair genes. Patient 13 who also had pre-surgery treatment history (Zoladex+bicalutamide) had a germline mutation in MSH6 (3991C>T, Arg1331Ter). Patient 502, who also had bone metastasis, had a somatic indel in MLH1 (Fig. 1). These outliers were excluded from the integrative analysis.
Replication	This is not applicable to our study.
Randomization	This is not relevant since we did not use different experimental groups or conditions in our study.
Blinding	This is not relevant since we did not use different experimental groups or conditions in our study.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

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	Our cohort was consisted of only Chinese patients with prostate cancer. Patient ages ranged from 50 to 88 years (median 69 years). Preoperative PSA levels ranged from 1.25 to 1270.14 ng/ml (median 38.8 ng/ml). The Gleason scores were distributed as follows: Gleason score 6:17 (8%); 7: 96 (46%); 8–10: 94 (45%). Organ-confined carcinoma (pT2) was found in 105 patients; 93 showed extra-prostatic tumor extension (pT3); 10 patients had advanced disease which invaded bladder, rectum or pelvic muscles (pT4). 15 patients (7%) had bone metastasis at the discretion of the referring urologist, prior to surgery. 11 percent of patients who underwent the lymph nodes dissection had positive lymph nodes invasion. Our cohort also contained 18 patients who exhibited some level of metastasis at diagnosis, and they were hormone sensitive. 16 patients were treated with ADT before surgery, mostly to reduce the size of their primary tumor. The clinical and pathological characterization of the final cohort were summarized in Supplementary Data 2.
Recruitment	After obtained the consent of patients, a series of 210 prostate tumor samples and their non-tumor counterparts were collected from patients surgically treated in the Urology Department. We focused on the primary tumors in our study.
Ethics oversight	Ethical committee approval for this study was obtained from Changhai Hospital (TMEC2014-001), the first hospital affiliated to Second Military Medical University. Written informed consent was obtained in accordance with Chinese legislation.

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