How clinically relevant are prostate cancer cell lines?
A comprehensive characterisation and multiomics comparison.

Zahra Ahmed[1,2], Warda Mosabbir[2], Devansh Tandon[2], Snehal Pinto Pereira[1], Umber Cheema[1], Marilena Loizidou[1], John Withington[1,3], Caroline Moore[1,3], Uzoamaka Okoli[1],Susan Heavey[1]

1. Division of Surgery and Interventional Sciences, University College London, London, United Kingdom
2. University College London Medical School, London, United Kingdom
3. University College London Hospitals NHS Foundation Trust

Word count:2616

Abstract

Cell line experiments arguably remain the most used tool in preclinical cancer research, despite their limitations. With almost 95% drugs entering human trials failing, and up to 90% preclinical research failing before even being tested in humans, we must shift the pre-clinical paradigm. A range of in silico, in vitro, in vivo and ex vivo approaches are gaining popularity, with the aim of potentially replacing cell line use. However, we cannot ignore the plethora of historical data from cell lines, nor write off their future use– especially within advanced bioengineered models. Therefore, we must question if and how cell lines hold clinical relevance. This study evaluates the clinical characteristics of 46 prostate cancer cell lines against worldwide data and investigates the biological features of seven cell lines in depth, comparing them to over 10,000 well characterised human cases from 24 studies in nine countries. Clinical features compared included age, ethnicity, Gleason grade, cancer type, treatment history and multiomics variables included mutations, copy number alterations, structural variants, microsatellite instability, mRNA and protein expression, and tumour mutational burden. We found that the most used cell lines accurately represent a minute proportion of prostate cancer patients. Furthermore, we recommend a pipeline for tailoring selection of clinically relevant cell lines with the ultimate aim of increasing the scientific methodology behind choosing a cell line.

1. Introduction

In drug development, an estimated 0.1% drugs which begin in preclinical trials are eventually approved (1) and oncology drugs have the highest attrition in clinical trials(2), due to incomplete proof of concept with human results failing to confirm preclinical findings (3, 4). The need to bridge the preclinical-clinical boundary has been stressed by the Academy of Medical Sciences and the Association of the British Pharmaceutical Industry (ABPI)(5), with proposed solutions including exploration of new preclinical and experimental models (6, 7).

Cell lines are frequently used for basic science, diagnostic and drug development; despite their shortcomings. The traditional use of cell lines in 2D demonstrates no heterogeneity, tumour microenvironment (TME) or immune influence (8). Furthermore, there is often cross-contamination (9). Despite this, cell lines have the advantage of not requiring biobank tissue or cell number restrictions.

Different PCa cell line models exist which are each derived from one patient and subsequently biologically altered during immortalisation (Supplementary Table 1).

Publicly available clinical datasets of PCa patients represent an opportunity to investigate the proportion of patients with specific PCa phenotypes and for analysis, visualization tools like cBioPortal encompass a range of clinical cohorts (Supplementary Table 2).

Limited PCa cell lines have been characterised using multiomics, but have not been compared comprehensively against patients, making them well characterised but not well validated. Important features to consider when assessing the clinical relevance of cell lines include donor ethnicity, age, cancer type, Gleason grade, metastatic stage, patient survival and androgen sensitivity.

A cell line cannot represent all patients but should hold clinical and biological relevance to the research question being asked. This study aimed to characterize and compare cell lines with publicly available patient data to determine how clinically and biologically relevant PCa cell lines and in what contexts. By tailoring the selection of cell line panels to match PCa

target populations more accurately, we propose that a greater proportion of preclinical research can be utilised in practice.

2. Methods

2.1 Data collection, curation and comparison

Clinical and biological data was collected from a range of sources, overviewed in the Supplementary Methods and Supplementary Tables 3-6.

Seven cell lines from the 46 identified were analysed in depth using multiomics data from the CCLA Broad study and COSMIC. The most frequent mutations, CNAs and SNVs from the clinical cohort and cell lines were compared.

The seven cell lines from CCLE were compared against the KEGG PCa gene set (M13191,(10, 11, 12)) which acted as a control, and visual representations of gene interactions for altered genes were generated.

After identifying key gene alterations in the seven cell lines, the number of patients with any alteration the same was identified (Supplementary Table 7). Patients with increasing numbers of alterations the same were stratified, demonstrating increased biological similarity.

Using patient IDs with biological similarity to cell lines (Supplementary File 1), features of "cell line-like patients" were compared to one another, the overall clinical cohort and available global data. Based upon clinical and biological findings from this study, a guide to choosing relevant cell lines for preclinical research projects was composed.

2.3 Statistical analyses

Statistical analyses and graphs were plotted using GraphPad Prism and SPSS. Categorical variables were presented as frequencies and percentages, and continuous variables were checked for normality and compared using two way-ANOVA tests. For binary data, Chi Squared was used. Median survival was calculated using log-rank Mantel-Cox test and plotted on Kaplan-Meier curves. Significance was defined as $p<0.05$, alpha error 0.05.

3. Results

Collected clinical and biological data utilised in this study is shown in Figure 1.

3.1 Clinical relevance of prostate cancer cell lines

3.1.1 History of immortalization

From a literature search, 46 PCa cell lines were identified and screened. PC-3 was the most cited and patent referenced, followed by LNCaP (Figure 3A). The three most used cell lines were immortalised over 40 years ago.

3.1.2 Anatomical and geographical location of sample collection

Half PCa cell lines were from metastases (Supplementary Table 8, Figure 1B), most commonly from lymph nodes. 22Rv1 was the most used prostate-derived cell line.

Geographically, sample collection was concentrated mostly in Europe and North America. (Figure 1C).

3.1.3 Age at sample collection

For PCa patients worldwide, age at sample collection most commonly occurred at age 60-64 (Figure 1D). For cell lines, almost 40% were also collected between 60-64, but the range was broader (29 to 83, Supplementary Table 8). The clinical cohort demonstrated an older average and normally distributed age (Supplementary Table 9).

3.1.3 Ethnicity

Nomenclature for ethnicity was in line with guidance from the Office for National Statistics(13).Worldwide, Black patients accounted for 15.2% PCa diagnoses but only 9% of the total clinical cohort (Figure 3A). Seven cell lines were from Black patients and only two were from Asian patients (KuCaP13 and PSK-1). 22Rv1 comprised almost equal proportions of White and Black ancestry.

3.1.4 Cancer type

Adenocarcinoma accounted for over 95% PCa types worldwide and 99.08% of the clinical cohort. Cell lines represented a range of cancer types (58.1% adenocarcinoma, Figure 3B).

3.1.5 Cancer stage and metastatic stage

Over 90% cell lines were derived from patients with stage four PCa and 75% patients had known metastases (N=23). The clinical cohort and worldwide patients mostly had lower stage PCa (Figure 3C) without metastases (82.98% non-metastatic, Figure 3D).

3.1.6 Gleason grade

Worldwide, 96.73% patients had Gleason grade 6 or 7. Almost half of the clinical cohort had Gleason grade >7 (48.06%, N=1103, Figure 3E). Gleason grade of donors from whom cell lines were established was only available for 15 cell lines and differed widely.

3.1.7 Treatment History

Prior treatment history for the clinical cohort varied but was largely unknown and cell line donors received mostly no treatment or androgen deprivation therapy (Supplementary Table 10).

3.2 The biological relevance of prostate cancer cell lines

3.2.1 Microsatellite Instability (MSI)

MSI was rare in the clinical cohort (0.43%) but more common in cell lines, with seven exhibiting MSI (63.64%, Figure 3F).

3.2.2 Fraction genome altered and mutation count

Fraction genome altered, and mutation count, were plotted against Gleason grade. LuCaP35 was from Gleason 9 PCa and had a higher fraction genome altered compared to the clinical cohort with Gleason 9 (0.75 LuCaP35 vs median 0.12 clinical cohort, Figure 3G).

22Rv1 (also Gleason 9) had a mutation count 84 times higher than the median clinical cohort with Gleason 9 (N=298, Figure 3H).

3.2.3 Mutations, Copy Number Alterations (CNAs) and Structural Variants (SVs)

The most frequent mutations, CNAs and SVs in the clinical cohort were compared against cell lines with multiomics data (Figure 3I-K). Cell lines typically lacked the most common CNAs and SVs from patients.

3.2.4 Gene alterations and interactions

Multiomics data from cell lines was compared against the KEGG PCa gene set to evaluate how many genes were altered in each cell line (Figure 4A). LNCaP-FGC had the most alterations in genes from the KEGG PCa gene set, PC-3 had the least (37 and 10 respectively) and 24% genes were not altered in any cell line.

3.2.5 Molecular pathways

Figure 4B details the top 50 pathways with the most alterations in genes from the KEGG PCa gene set compared with the number of genes altered in cell lines in these pathways. NCIH660 had a high proportion of genes altered in the peptidyl-serine modification and regulation of the innate immune system (21 and 20% genes, Figure 4C) ; 22Rv1 also had genes altered in the peptidyl-serine modification pathway (18%). Visualizations for remaining cell lines including gene interactions and pathways are shown in Supplementary Figures 1-5.

3.3 Identification and characterisation of cell line-like patients

3.3.1 Sub-cohort identification

Identification of sub-cohorts of patients that were wholly representative of specific cell lines in terms of molecular alterations was largely unsuccessful. VCaP had the most patients with the same alterations present, 17 of which had 5-6 alterations the same out of the 89 genes in the KEGG PCa gene set. Other cell lines had less than 0.07% patients with 5-6 alterations the same (range 0 to 3).

3.3.2 Survival

Patients with at least one alteration the same as a cell line had significantly lower percentage survival at 50 months compared to the overall cohort ($p<0.001$ log rank Mantel-Cox, Figure 4E&F). Most patients with 2-4 alterations the same as cell lines had a lower 50 month percentage survival compared to control, except 22Rv1 (Figure 4G).

3.2 Creation of cell line selection tool

Figure 5 demonstrates a three-step process guiding choice of PCa cell line. Step one considers key clinical factors relevant to almost all research questions, which should be used to shortlist cell lines of desired cancer stage, treatment status and androgen sensitivity. Step two narrows choice of cell line, by considering clinical factors including: cancer type,

anatomical site of sample collection, patient age and race. Finally, step three caters for more specific factors.

4. Discussion

Traditional 2D monoculture cell lines are employed to investigate mechanisms and test drugs(14). In vitro PCa models have advanced to better emulate in vivo PCa by incorporating cells in various setups and 3D cell line-based models(15). Molecular characterization and limited detail of the clinical features of the patients PCa cell lines were derived from exists. However, there is limited literature evaluating the extent to which cell lines represent patients(16, 17) and no studies investigating which cell line is most suited for investigating patient subtypes. Therefore, this study aimed to contextualise PCa cell lines against large patient cohorts and worldwide data, and to characterise cohorts biologically similar to specified cell lines.

We found that many cell lines were established from metastasis and patients with rare presentations of PCa. From this alone, PCa cell lines are poorly clinically relevant. Additionally, there were no cell lines and patient datasets from countries with the highest PCa incidence. Black men often have aggressive PCa (18, 19, 20); With limited cell line models or patient data from Africa, it is likely that most PCa research excludes patients with the lowest survival. Furthermore, Black men are underrepresented in clinical trials attributable to factors relating to interaction with healthcare services, institutional racism and unconscious bias (21). Our results emphasise the need to implement strategies increasing representation of Black men in PCa research.

PCa cell lines often come from unique cancer types; PSK-1 came from a prostatic small-cell carcinoma patient (<2% PCas)(22) with Klinefelter syndrome (0.1-0.2% prevalence)(23). Whilst interesting clinically, these cell lines hold little relevance to understanding most PCas. Cell lines are biased to be derived from aggressive tumours, from their greater propensity to grow indefinitely. Although advanced stage cancers are of interest due to unmet clinical need, in future, more cell lines representative of early stage PCa are needed.

Seven cell lines profiled for multiomics data shared gene alterations from the KEGG PCa gene set, however, 24% genes were not altered in any cell line. This could mean that failed preclinical drug trials in these cell lines are due to lack of alterations in pathways, wherein the

drug's mechanism is targeted. Evaluation of cell line choice should be performed to ascertain whether the drug may work in a more suitable model. Most cell lines do not have detailed multiomics data available, which may bias researchers to select cell lines with more data available in relation to their genes of interest, rather than other potentially more relevant cell lines that lack data.

NCIH660 showed significant gene interactions between CDK1, CCNE2 and INS, perhaps suggesting it is a good option to investigate pathways involving these genes or for patients with alterations in these genes.

PC-3 was the least biologically altered, perhaps suggesting it is representative of patients with less altered genomes and earlier stage PCa.

To identify patients biologically-like cell lines, a tenuous link was suggested: as the number of gene alterations the same increased, so did similarity. From the clinical cohort, VCaP had the most patients with gene alterations the same, potentially suggesting it represents more PCa patients. However, the key finding here was that similarity was lower than expected overall.

Many cell lines had high MSI(MSI-H), which positively correlates with treatment response to immune checkpoint inhibitors in advanced PCa (24, 25). Furthermore, patients with high TMB have greater likelihood of immune activation(26) and response to immunotherapeutics. Therefore, cell lines with high MSI and TMB may be suited to investigate new checkpoint inhibitors.

The characteristics of patients biologically similar to 22Rv1 were different to other cell line-like patients (Supplementary File 2) with increased survival rates. 22Rv1 exhibits MSI (27) and patients with 2-4 alterations the same as 22Rv1 had the highest proportion of MSI-H and high TMB of all groups (10 and 9.02% patients). This could suggest that 22Rv1 is most suitable for immunotherapeutic experiments.

The multitude of PCa cell lines available, and strikingly high attrition rate of pre-clinical PCa research, render it pivotal to follow a systematic, evidence-based approach when selecting cell lines for pre-clinical research. Figure 5 serves as a tool for this, and should be used to whatever degree deemed appropriate.

A study of this kind requires compromises to be made where global data is not available. Sources for Gleason grade and ethnicity were from single-country studies but included due to large patient cohorts. Another limitation of this study is that the number of cell line-like patients may have been skewed by alteration type and genes profiled for in each dataset. This was controlled for by including studies with a range of alterations (Supplementary Table 11).

Only seven PCa cell lines had multiomics data available, however, cell lines lacking multiomics data showed clinical and biological relevance. These cell lines should be comprehensively characterised to see if they hold relevance and if not, more representative cell lines are required.

Looking forward, 2D cell lines are likely to be continually used in 3D culture, especially as prostate-specific treatment studies have shown 3D in vitro models to be more biomimetic of in vivo tumours (28).

5. Conclusion

Cell lines are valuable in PCa research and are used for basic science research, development of advanced models, development of drugs and diagnostics, as well as training of junior scientists. However, it is pivotal that cell lines are applicable to the largest cohort of PCa patients possible and there are available niche cell lines representing patient subtypes. This study contextualised commonly used PCa cell lines against patients in clinical cohorts and worldwide. Cell line-like patients were characterised to inform researchers which cell line is most suited to their targeted cohort and experiment.

Acknowledgments

Thank you to the Wolfson Foundation and the Royal College of Physicians for funding this project.

References

1. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. Translational Medicine Communications. 2019;4(1):18.
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nature Reviews Drug Discovery. 2004;3(8):711-6.
3. Perrin S. Preclinical research: Make mouse studies work. Nature. 2014;507(7493):423-5.
4. Begley CG, Ellis LM. Raise standards for preclinical cancer research. Nature. 2012;483(7391):531-3.
5. Sciences TAoM. Bridging the preclinical-clinical boundary. 2018 09/03/2018.
6. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery. 2010;9(3):203-14.
7. INC. ILC. Preclinical Safety vs Efficacy Research [cited 2023 05/05/2023]. Available from: https://www.itrlab.com/preclinical-safety-vs-efficacy-research/.
8. Namekawa T, Ikeda K, Horie-Inoue K, Inoue S. Application of Prostate Cancer Models for Preclinical Study: Advantages and Limitations of Cell Lines, Patient-Derived Xenografts, and Three-Dimensional Culture of Patient-Derived Cells. Cells. 2019;8(1).
9. van Bokhoven A, Varella-Garcia M, Korch C, Hessels D, Miller GJ. Widely used prostate carcinoma cell lines share common origins. Prostate. 2001;47(1):36-51.
10. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
11. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947-51.
12. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587-d92.
13. Statistics OfN. Ethnic group, national identity and religion 2021 [Available from: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion.
14. Namekawa T, Ikeda K, Horie-Inoue K, Inoue S. Application of Prostate Cancer Models for Preclinical Study: Advantages and Limitations of Cell Lines, Patient-Derived Xenografts, and Three-Dimensional Culture of Patient-Derived Cells. Cells. 2019;8(1):74.
15. Faustino-Rocha AI, Jota-Baptista C, Nascimento-Gonçalves E, Oliveira PA. Evolution of Models of Prostate Cancer: Their Contribution to Current Therapies. Anticancer Res. 2023;43(1):323-33.
16. Masters JRW. Human cancer cell lines: fact and fantasy. Nature Reviews Molecular Cell Biology. 2000;1(3):233-6.
17. Trastulla L, Noorbakhsh J, Vazquez F, McFarland J, Iorio F. Computational estimation of quality and clinical relevance of cancer cell lines. Molecular Systems Biology. 2022;18(7):e11017.
18. Mahal BA, Alshalalfa M, Spratt DE, Davicioni E, Zhao SG, Feng FY, et al. Prostate Cancer Genomic-risk Differences Between African-American and White Men Across Gleason Scores. European Urology. 2019;75(6):1038-40.
19. Ahaghotu C, Baffoe-Bonnie A, Kittles R, Pettaway C, Powell I, Royal C, et al. Clinical characteristics of African-American men with hereditary prostate cancer: the AAHPC study. Prostate Cancer and Prostatic Diseases. 2004;7(2):165-9.

20. Carter BS, Bova GS, Beaty TH, Steinberg GD, Childs B, Isaacs WB, et al. Hereditary Prostate Cancer: Epidemiologic and Clinical Features. Journal of Urology. 1993;150(3):797-802.
21. Lillard Jr JW, Moses KA, Mahal BA, George DJ. Racial disparities in Black men with prostate cancer: A literature review. Cancer. 2022;128(21):3787-95.
22. UK CR. Small Cell Prostate Cancer 2022 [updated 02/05/2022. Available from: https://www.cancerresearchuk.org/about-cancer/prostate-cancer/stages/types/small-cell-prostate-cancer#:~:text=Less%20than%202%20in%20every,common%20type%20of%20prostate%20cancer.
23. Los E FG. Klinefelter Syndrome. StatPearls. 2023.
24. Abida W, Cheng ML, Armenia J, Middha S, Autio KA, Vargas HA, et al. Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. JAMA Oncology. 2019;5(4):471-8.
25. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. New England Journal of Medicine. 2015;372(26):2509-20.
26. Fusco MJ, West H, Walko CM. Tumor Mutation Burden and Cancer Treatment. JAMA Oncology. 2021;7(2):316-.
27. Lu Y, Soong TD, Elemento O. A Novel Approach for Characterizing Microsatellite Instability in Cancer Cells. PLOS ONE. 2013;8(5):e63056.
28. Fontana F, Raimondi M, Marzagalli M, Sommariva M, Gagliano N, Limonta P. Three-Dimensional Cell Cultures as an In Vitro Tool for Prostate Cancer Modeling and Drug Discovery. Int J Mol Sci. 2020;21(18).

Figure Legend

Figure 1 Overview of the clinical and biological factors were collected and analysed in this study Raw data was collected for 7 major clinical factors. Raw data was collected for 6 overarching biological factors. The numbers of cell lines and size of clinical cohort for each factor is stated (if applicable).

Figure 2 The most commonly used prostate cancer cell lines were derived over 40 years ago, were not derived from the prostate and are from very few geographical locations: **A** Bubble plot showing the most commonly used prostate cancer cell lines with number of literature citations (y axis) against the year of sample collection (x axis) with the size of bubble representing the number of patents and colour representing primary tumour (orange) or metastatic (grey) origin. **B** Illustration showing the anatomical locations of prostate cancer cell line sampling, numbers represent how many cell lines are derived from the anatomical site and the size of bubble represents the number of literature citations for cell lines derived from the anatomical site. **C** Map showing where cell line (green pins) and human datasets from publicly available clinical cohort data used in this study (red pins) were sampled from against a background gradient of prostate cancer incidence from low (light shade) to high incidence (dark shade). **D** Line graph showing distribution of age at sample collection for cell lines (N = 26), clinical cohort (N = 4646) and worldwide prostate cancer patients. ⊥ = Global Cancer Observatory, International Agency for Research on Cancer, World Health Organization, Cancer Today Website

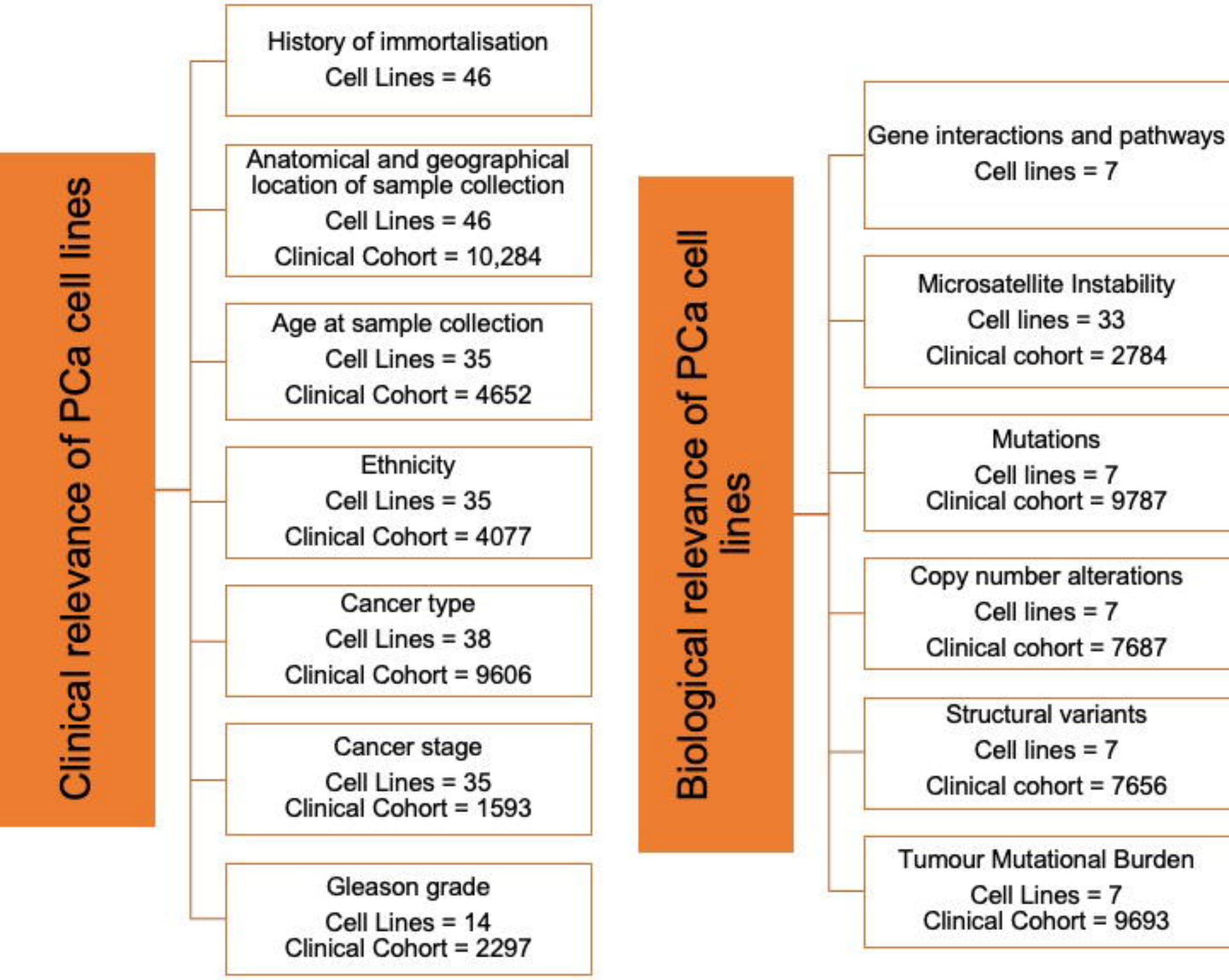
Figure 3 Cell lines and clinical cohorts have different clinical features to prostate cancer patients worldwide and the most common alterations in clinical datasets are not necessarily altered in cell lines : A Stacked bar chart showing the ethnicity distribution of original patients cell lines were derived from (N=35), 22Rv1 cell line, clinical cohort (N=4077) and worldwide prostate cancer patients (N=22977). Other races included Latino, Asian, Hawaiian, and Native American. **B** Stacked bar chart showing the cancer type of original patients cell lines were derived from (N=42), clinical cohort (N=9606) and worldwide prostate cancer patients. Other cancer types non-prostate carcinomas including bladder cancer**. C** Stacked bar chart showing the cancer stage of original patients cell lines were derived from (N=33), clinical cohort (N=1593) and worldwide prostate cancer patients (N=94811). **D** Stacked bar chart showing the proportion of patients with metastasis at sample collection of original patients cell lines were derived from (N=30), clinical cohort (N=2069) and worldwide prostate cancer patients (N=7810). **E** Stacked bar chart showing the distribution of Gleason grade of patients at sample collection of original patients cell lines were derived from (N=28), clinical cohort (N=2297) and worldwide prostate cancer patients (N=980). **F** Stacked bar chart showing proportion of patients with microsatellite instability for cell lines (N=11), clinical cohort (N=2784). **G** Scatter plot showing the fraction genome altered against Gleason grade for the clinical cohort and two cell lines with available data (LuCaP35 and 22Rv1)**. H** Scatter plot showing the mutation count against Gleason grade for the clinical cohort and one cell line with available data (22Rv1). **I** The most common mutations out of 9787 patients from clinical datasets were TP53, TTN, FRG1BP, SPOP, FOXA1 and MUC16- mutations in these genes occurred in at least 10% patients respectively. Cell lines which also had mutations in TP53 were VCaP, DU145, 22Rv1, PC-3, LAPC4 and LAPC9. The cell line with a mutations in FOXA1 was DU145. **J** The most common copy number alterations (CNAs) out of 7656 patients from clinical datasets were PTEN, AR and MYC- CNAs in these genes occurred in at least 8% patients respectively. Cell lines with CNAs in PTEN were NCIH660 and LAPC9. **K** The most common structural variants (SVs) out of 7687 patients from clinical datasets were all fusions in either : TMPRSS2 (intragenic), TMPRSS2-ERG, TMPRSS2-ETV1, TMPRSS2-ETV4, TMPRSS2-LINC00114 and ERG (intragenic) - SVs in these genes occurred in at least 0.5% patients respectively. Cell lines also with fusions in TMPRSS2-ERG were NCIH660 and VCaP. M stage = Metastatic Stage, MSI = Microsatellite Instability. o = The Journal of the American Medical Association worldwide racial demographics of prostate cancer study and The Surveillance, Epidemiology, and End Results Program database, ∂ = Ackerman et.al Prostate surgical pathology, ∧ = Weerakoon et. al The current use of active surveillance in an Australian cohort of men

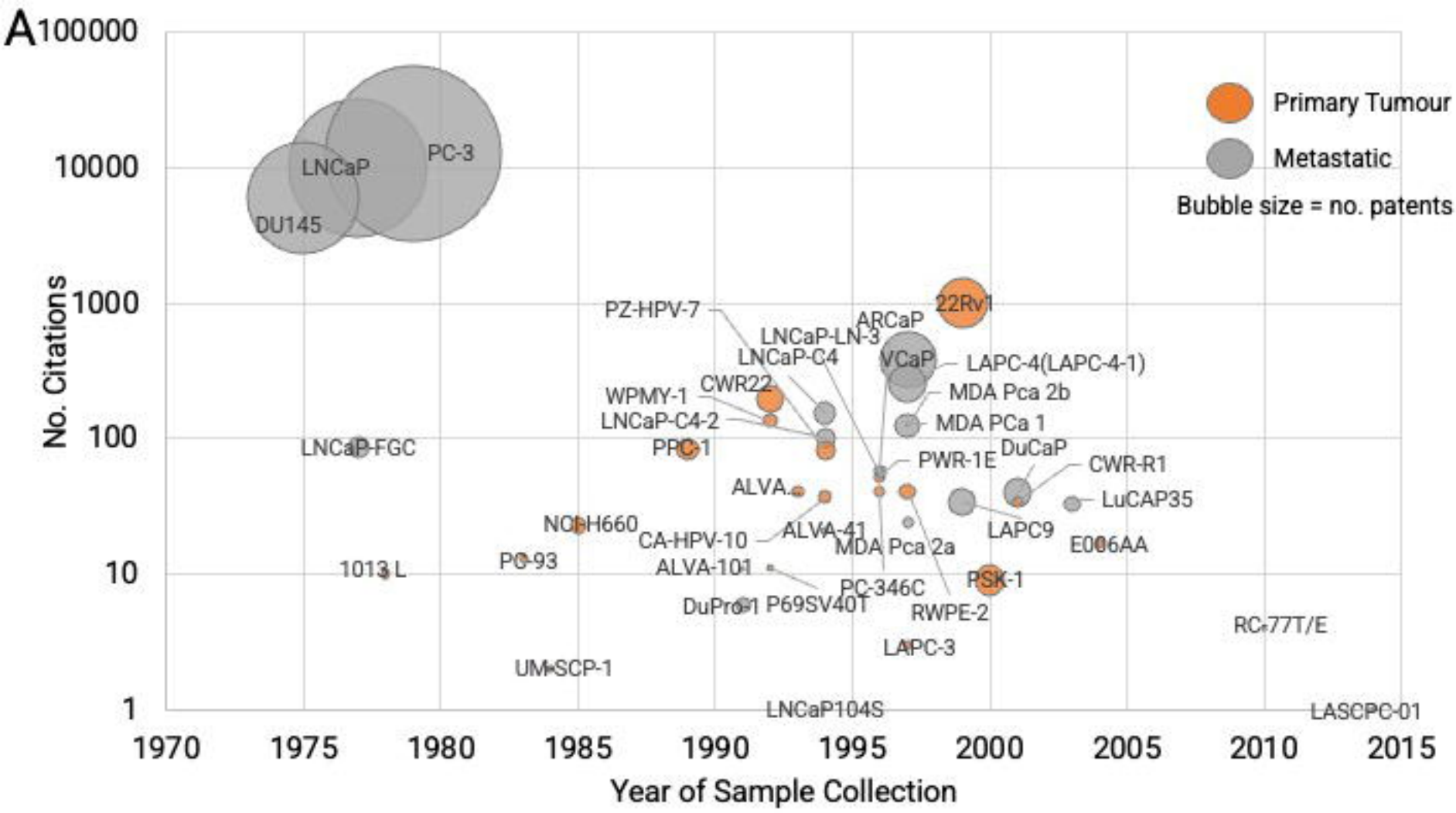
Figure 4 Cell lines do not share many of the same important gene alterations present in many prostate cancers; Comparing genes altered in cell lines at an individual and a pathway level: A Bar chart showing the frequency and alteration type in prostate cancer gene**s** from the well characterised KEGG prostate cancer gene set were altered in. **B** Heatmap showing the proportion of genes altered from the KEGG prostate cancer gene set in each cell line/ total number of genes in the molecular pathway, where blue is lowest (0) and red is highest (0.26). Top 50 pathways with genes altered from KEGG prostate cancer gene set as a control on the left. **C** Identifying any gene interactions between altered genes in NCIH660 and their important predicted pathways which the cell line could be used to investigate. Diagram created using GeneMANIA with the inner ring represents genes altered in this cell line and the outer ring represents predicted related genes. The top seven pathways are colour coded inside the gene bubbles and the different coloured lines represent networks. **D-F** Kaplan-Meier survival plots for cell line-like patients compared to overall clinical cohort

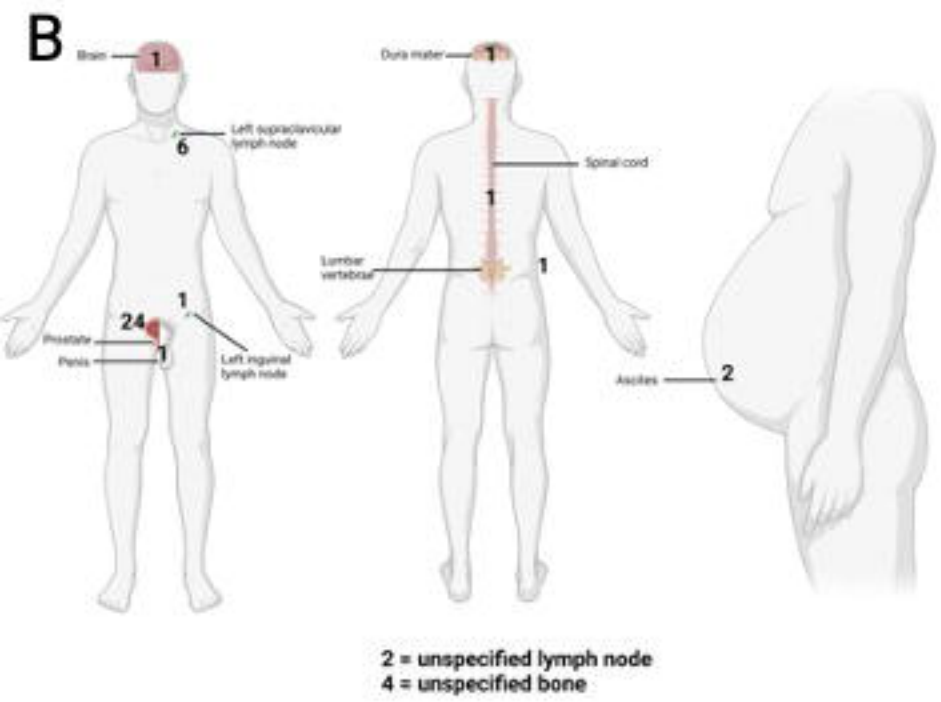
control (n=2323). **D**-Patients with at least one alteration the same as cell lines (overall), **E**-Patients with one alteration the same as cell lines, **F**-Patients with 2-4 alterations the same as cell line. MSI = Microsatellite Instability, KM = Kaplan-Meier

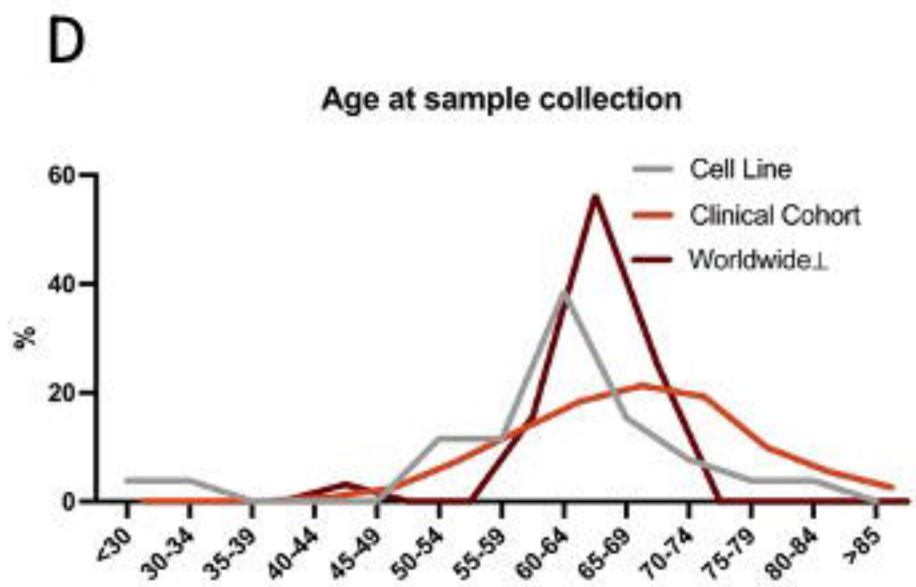
Figure 5 Selection of clinically relevant prostate cancer cell line(s) should be tailored to the research question. Researchers are encouraged to select a panel of cell lines to carry out experiments on. Each cell line can be selected based on a variety of factors, stratified into steps based on importance and impact. **A** Step 1: Cell lines can be selected based initially on their baseline clinical characteristics. Here, PCa cell lines are stratified based on cancer stage, the mode of treatment and whether the cancer is androgen sensitive or not. Pie charts highlight the discrepancy between the percentage of cell lines and percentage of the corresponding clinical cohort, based on donor characteristics at the time of data collection. **B** Step 2: Further clinical characteristics should be taken into consideration. Here, PCa cell lines are split into tables based on cancer type, anatomical location of sample collection, age of donor at time of sample collection and race of donor. **C** Step 3: Relevance to the research question should be used to further select cell lines for inclusion in a panel. Here, A flowchart categorises PCa cell lines based on whether the research question is generalised or specific in nature. The generalised side of the flowchart specifies 3 cell lines that may be suitable for predominantly broad studies. The specific side of the flowchart uses pie charts and tables to further stratify cell lines based on Gleason grade, microsatellite instability and specific gene alterations that are most popular in PCa research. A panel should include cell lines possessing an alteration of interest as well as cell lines that do not.

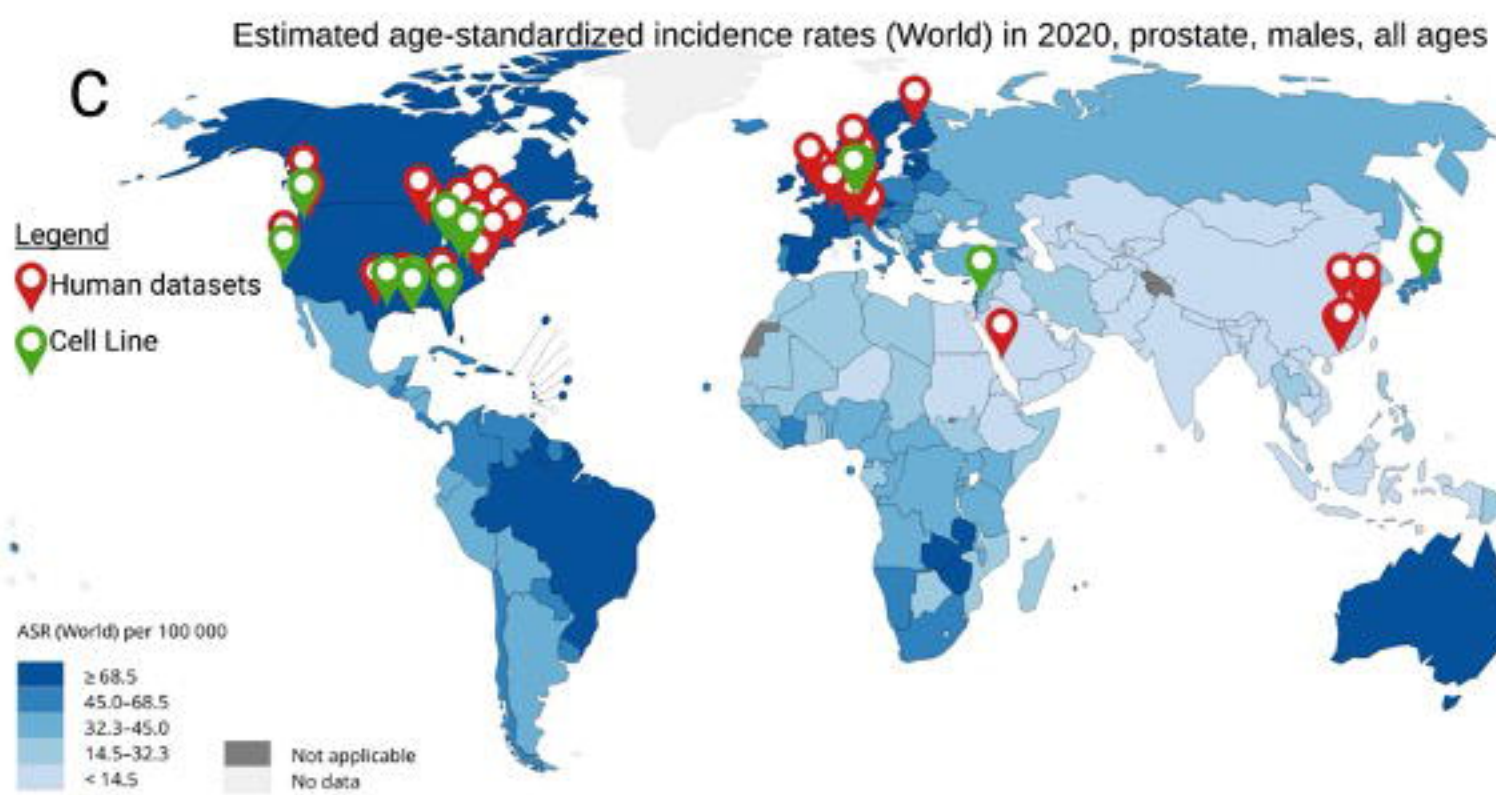
Total Cell Lines = 46
Total Clinical Cohort = 10,284









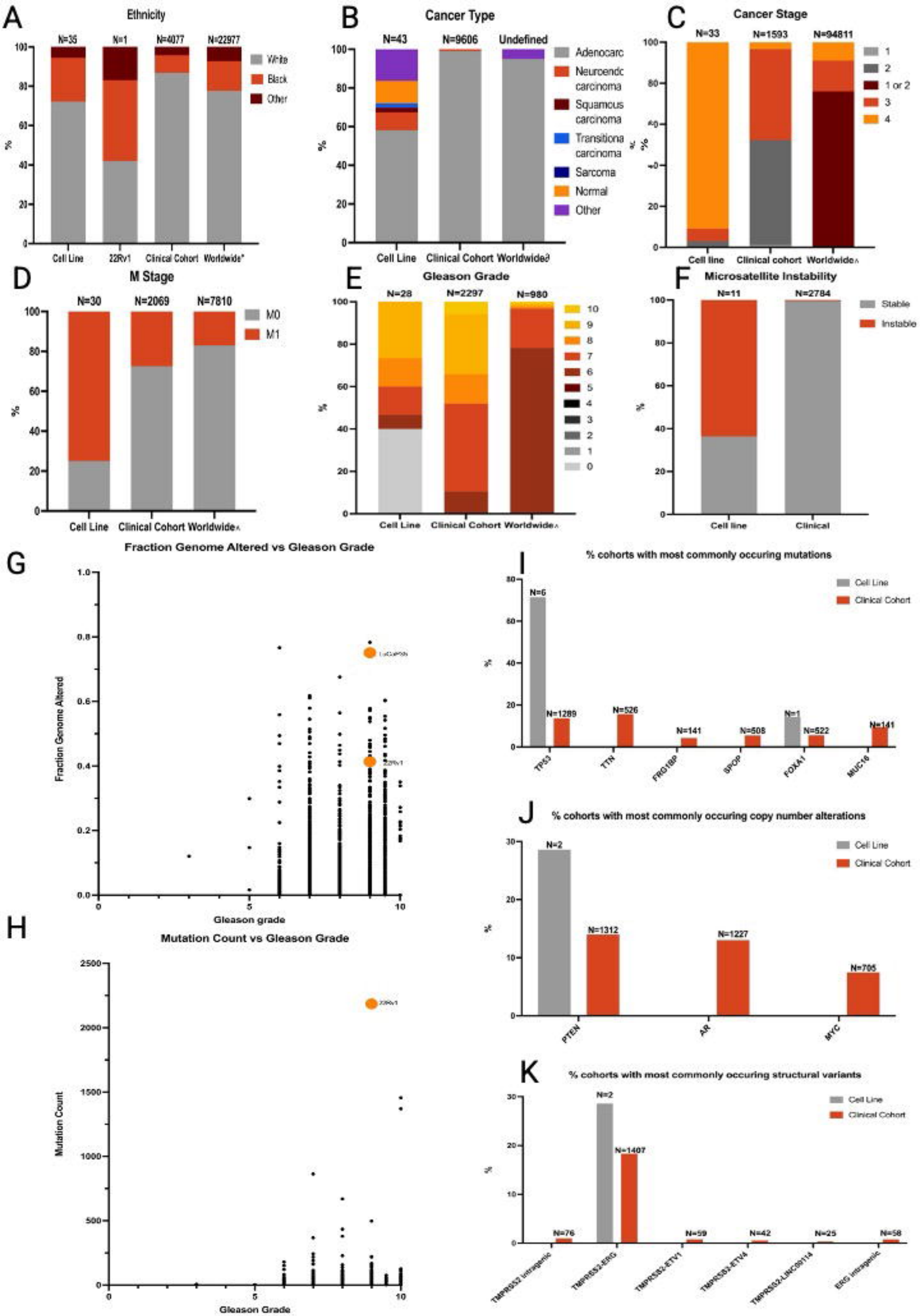


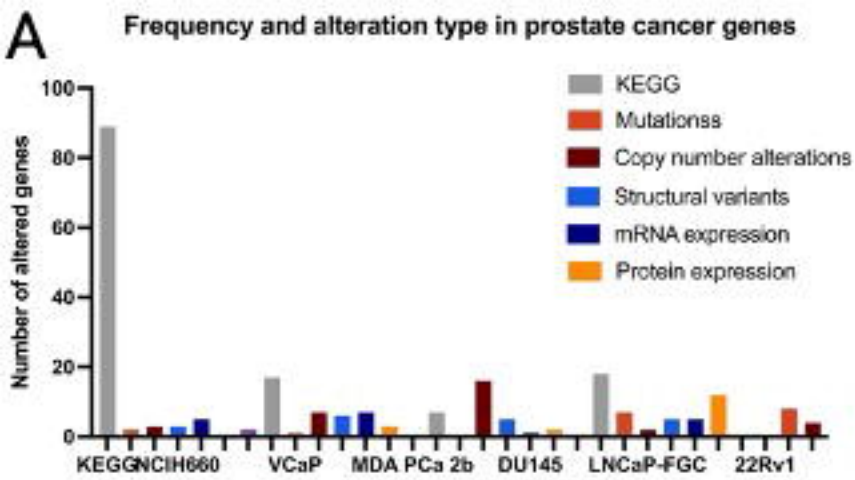
All rights reserved. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization / International Agency for Research on Cancer concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate borderlines for which there may not yet be full agreement.

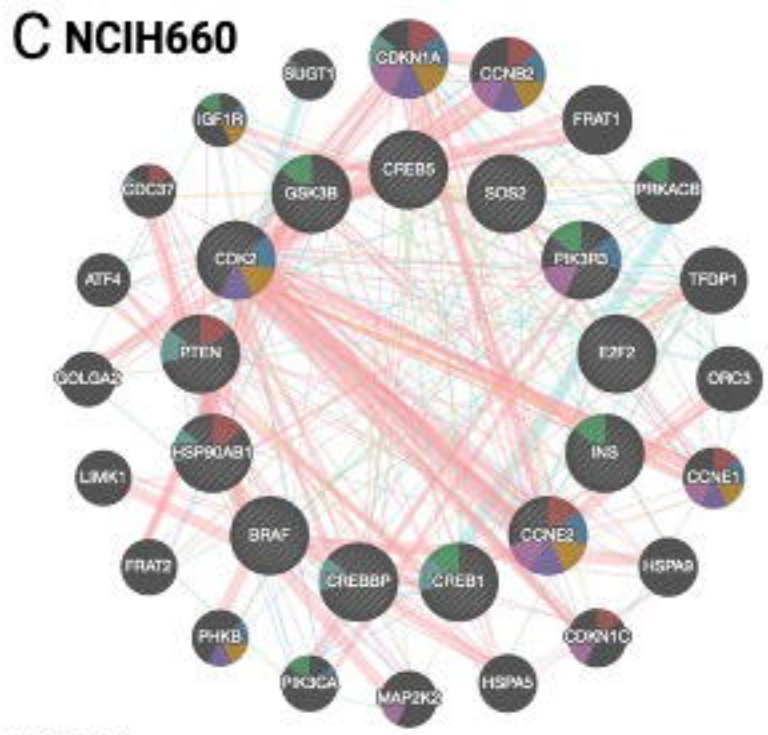
Data source: GLOBOCAN 2020
Map production: IARC
(https://gco.iarc.fr/today)
World Health Organization

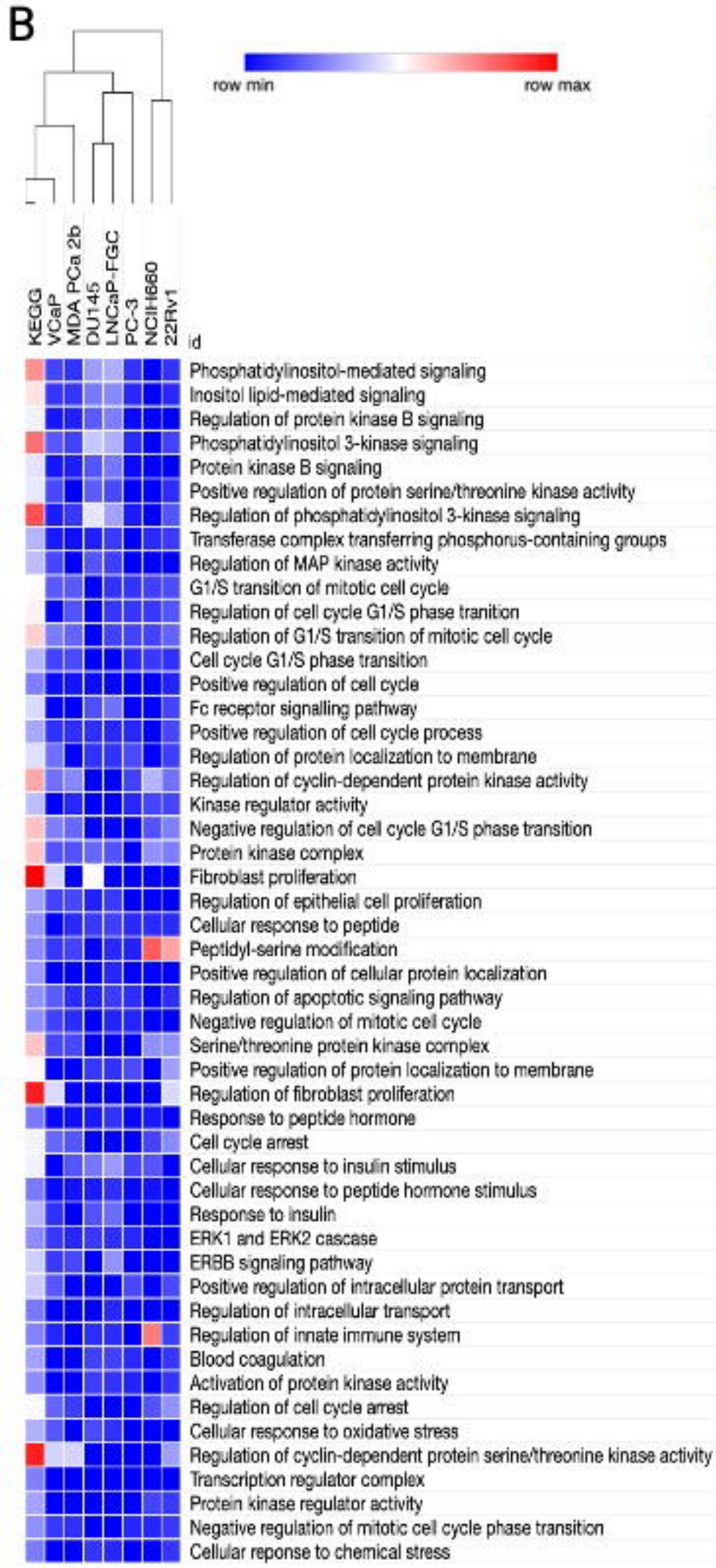
World Health Organization

© International Agency for Research on Cancer 2020
All rights reserved


A
Ethnicity
N=35
N=1
N=4077
N=22977
White
Black
Other
%
Cell Line
22Rv1
Clinical Cohort
Worldwide*
B
Cancer Type
N=43
N=9606
Undefined
Adenocarc
Neuroendc carcinoma
Squamous carcinoma
Transitiona carcinoma
Sarcoma
Normal
Other
Cell Line
Clinical Cohort
Worldwide
C
Cancer Stage
N=33
N=1593
N=94811
1
2
1 or 2
3
4
Cell line
Clinical cohort
Worldwide
D
M Stage
N=30
N=2069
N=7810
M0
M1
Cell Line
Clinical Cohort
Worldwide
E
Gleason Grade
N=28
N=2297
N=980
10
9
8
7
6
5
4
3
2
1
0
Cell Line
Clinical Cohort
Worldwide
F
Microsatellite Instability
N=11
N=2784
Stable
Instable
Cell line
Clinical
G
Fraction Genome Altered vs Gleason Grade
Fraction Genome Altered
Gleason grade
H
Mutation Count vs Gleason Grade
Mutation Count
Gleason Grade
I
% cohorts with most commonly occuring mutations
Cell Line
Clinical Cohort
N=6
N=1289
N=526
N=141
N=508
N=1
N=522
N=141
TP53
TTN
FRG1BP
SPOP
FOXA1
MUC16
J
% cohorts with most commonly occuring copy number alterations
Cell Line
Clinical Cohort
N=2
N=1312
N=1227
N=705
PTEN
AR
MYC
K
% cohorts with most commonly occuring structural variants
Cell Line
Clinical Cohort
N=2
N=1407
N=76
N=59
N=42
N=25
N=58
TMPRSS2 intragenic
TMPRSS2-ERG
TMPRSS2-ETV1
TMPRSS2-ETV4
TMPRSS2-LINC00114
ERG intragenic





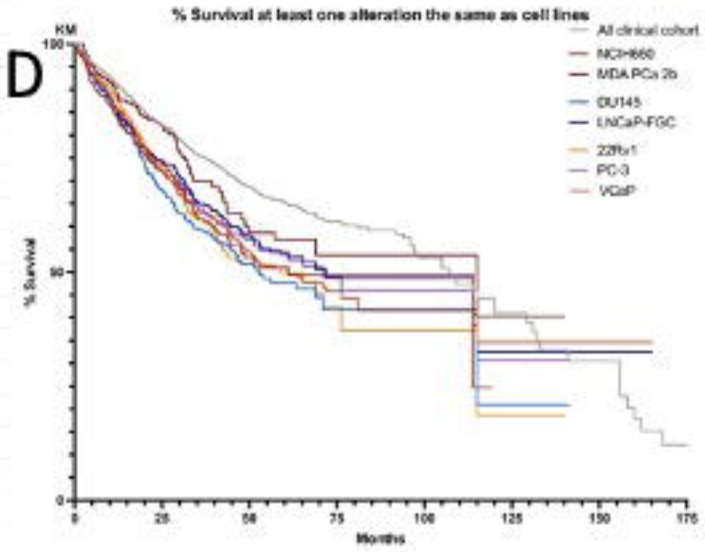


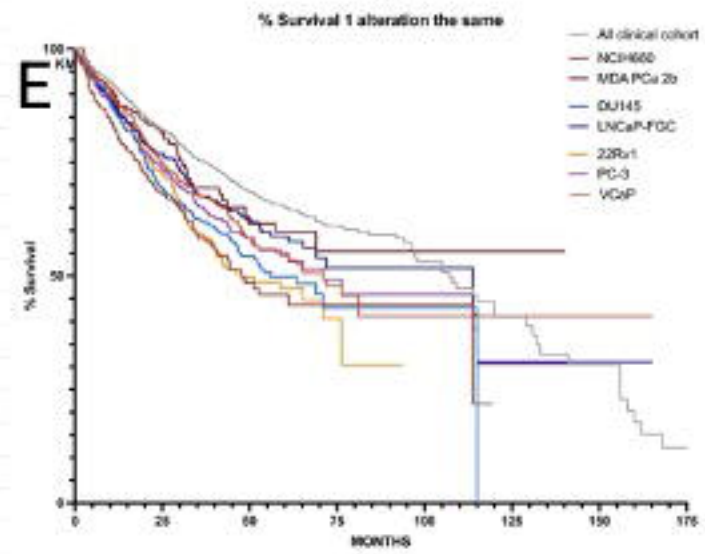
Functions

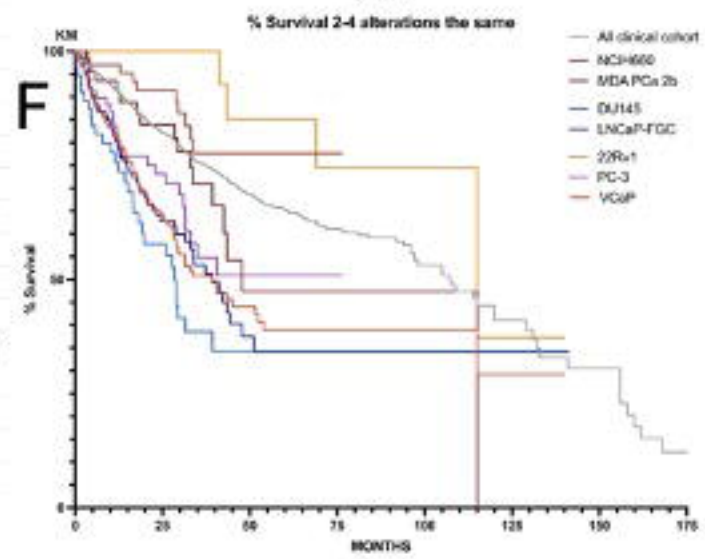
- regulation of cyclin-dependent protein kinase activity
- transferase complex, transferring phosphorus-containing groups
- protein kinase complex
- serine/threonine protein kinase complex
- kinase regulator activity
- regulation of protein stability
- response to peptide hormone

Networks

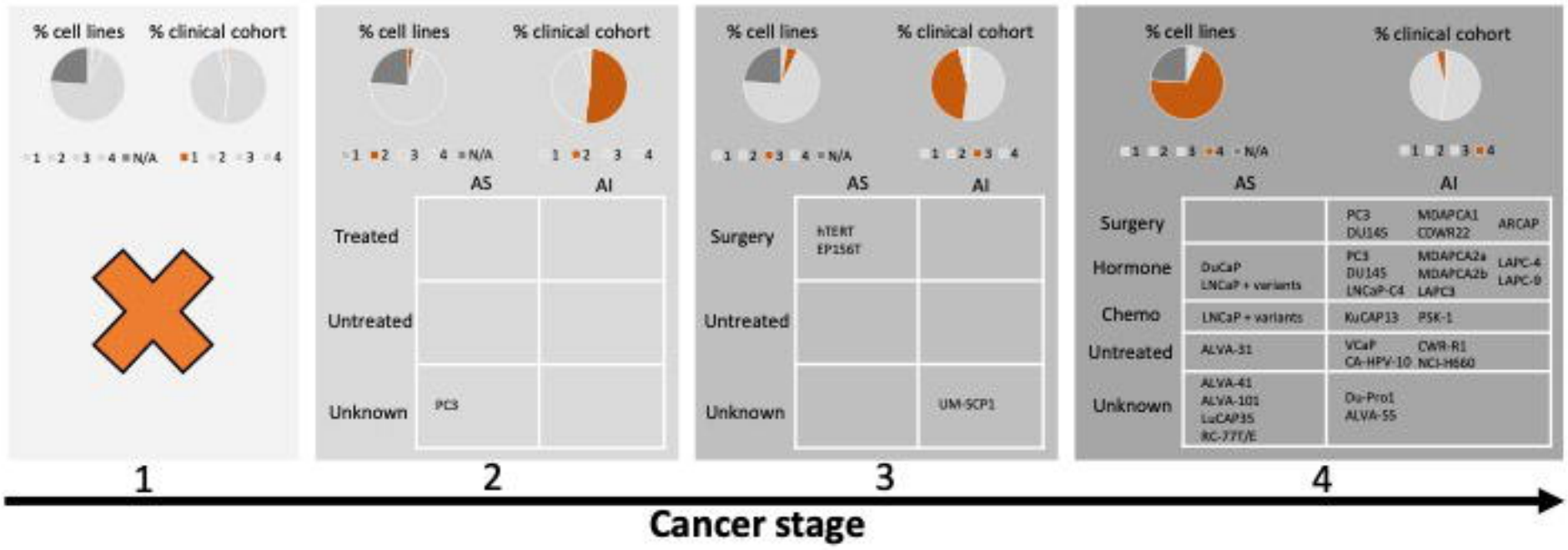
- Physical Interactions
- Co-expression
- Pathway
- Genetic Interactions
- Co-localization
- Predicted
- Shared protein domains







Step 1: choose cell line(s) based on key clinical factors



Step 2: consider further clinical factors relevant to study

Cancer type

Adenocarcinoma	LNCaP LNCaP-C4 LNCaP-C4-2 LNCaP-FGC LNCaP-LN3 LNCaP-P104S LuCAP35	PC3 DU145 MDA PCA1 MDA PCA2 MDA PCA2A PPC-1 E006AA	ALVA-101 LAPC-3 RC-77T/E ALVA-55 hTERT EP156T LAPC-9 ARCaP	CWR-R1 ALVA-31 CA-HPV-10 PC93 ALVA-41 PC-346C
Small cell carcinoma	PSK-1	NCI-H660		
Squamous cell carcinoma	UM-SCP-1			
Neuroendocrine	KuCaP13	LASPC-01		
Urothelial	1013-L			
Normal prostate	PZ-HPV-7 RWPE-2	P69SV40T WP-MY1	PNF-08	

Anatomical location

Prostate	1013L 22Rv1 PPC1 PZ-HPV-7 PWR-1E CWR22 CWR-R1 ALVA-31	RWPE-2 CA-HPV-10 PSK-1 PC-93 PC-346C E006AA NCIH660 P69SV40T	DuPro-1 WPMY-1 UM-SCP-1U LAPC-3 RC-77T/E LASCPC-01 hTERT EP156T PNF-08
Lymph node	LNCaP LNCaP-C4 LNCaP-C4-2	LNCaP-FGC LNCaP-LN3 LNCaP-P104S	LAPC-4 ALVA-55 LuCAP35
Metastases	PC-3 DU145 1013L MDA-PCA-1 MDA-PCA-2a MDA-PCA-2b	ARCaP DuCAP ALVA-41 ALVA-101 KuCaP13 LAPC9	

Race

White	LNCaP LNCaP-C4 LNCaP-C4-2 LNCaP-FGC LNCaP-LN3 LNCaP-P104S PC-3 DU145	1013L VCaP PPC-1 ARCaP PZ-HPV-7 PWR-1E DuCaP ALVA-31	RWPE-2 CA-HPV-10 ALVA-41 NCIH660 ALVA-101 DuPro-1 WPMY-1 UM-SCP-1U	ALVA-55 LASCPC-01 PNF-08 LuCAP35
Black	MDA-PCa-1 MDA-PCa-2a MDA-PCa-2b RC-77T/E	PC-93 E006AA P69SV40T		
Mixed	22Rv1	CWR22		
Asian	PSK-1	KuCaP13		

Age

<50	PSK-1	PC-93	
50-59	LNCaP LNCaP-C4 LNCaP-C4-2 LNCaP-FGC	LNCaP-LN3 LNCaP-P104S E006AA WPMY-1	VCaP RWPE-2 DuCaP
60-69	PC3 DU124 MDA-PCa-1 MDA-PCa-2a MDA-PCa-2b PPC1	PWR-1E ALVA-31 CA-HPV-10 ALVA_41 NCI-H660 P69SV40T	DuPro-1 RC-77T/E ALVA-55 KuCaP13 LuCAP35
>=70	ARCaP PZ-HPV-7	UM-SCP-1 LASCPC-01	

Step 3: additional factors for consideration

Generalised PCa studies

LNCaP - FGC

Most generalisable (biological)

- KEGG Prostate Cancer Geneset summarises well-recognised gene alterations in PCa
- Most congruence with KEGG (37/89 identical gene alterations)
- Note: despite this, similarity remains low

VCaP

Most generalisable (clinical)

- VCaP expressed the greatest congruence with the clinical cohort (n=8379)
- 1897 patients had >= 1 gene alteration the same as VCaP
- 17 patients had 5-6 gene alterations the same as VCaP
- Note: despite this, similarity remains low

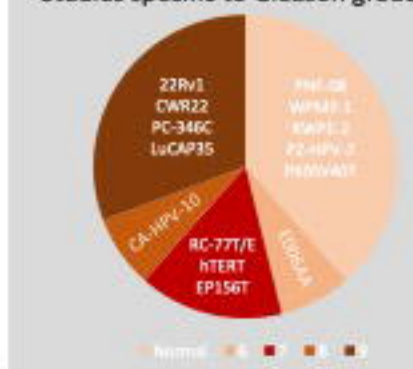
PC3

Most generalisable (validational)

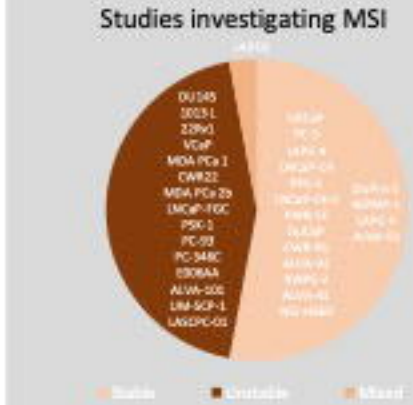
- More data exists for PC3 than any other PCa cell line, representing a wealth of knowledge for contextualisation of findings
- Most-cited PCa cell line on PubMed (15,673 citations)
- Most patented PCa cell line (37,891 patents on WIPO and 90,858 patents on Google)

Specific PCa studies

Studies specific to Gleason grade



Studies investigating MSI



Studies evaluating specific alterations	Cell line
BRCA2	LNCAP | LAPC4
MAPK pathway	DU145
PI3K pathway	DU145
PTEN	NCIH660 | LNCaP-FGC | PC3 | LAPC-9
TP53	VCaP | PC3 | DU145
MYC	NCIH660 | VCaP PC3 | MDAPCa2b LuCaP35
AR	LNCaP | MDAPCa 2b
TMPRSS2-ERG (SNV)	NCIH660 | VCaP

Cell Line	ATCC and/or Source	References
LNCaP	Julius Horoszewicz (Roswell Park Memorial Institute, Buffalo, NY, USA)	(1)
PC-3	ATCC (CRL-1435)	(2)
DU145	ATCC (HTB-81)	(3)
1013 L	Anita Blistrom (Active Biotech Research AB, Lund, Sweden)	(4)
22Rv1	ATCC (CRL-2505) James Jacobberger (Case Western Reserve University, Cleveland, OH, USA)	(5)
VCaP	ATCC (CRL-2876) Kenneth Pienta (University of Michigan, Ann Arbor, MI, USA)	(6)
MDA PCa 1	Nora Navone (University of Texas, Houston, TX, USA)	(7)
CWR22	NCI Division of Cancer Treatment and Diagnosis (DCTD) Tumour Repository	(8)
LAPC-4(LAPC(41))	Charles Sawyer (University of California, Los Angeles, CA, USA)	(9)
LNCaP-C4	ATCC (CRL-3313)	(10)
MDA Pca 2b	ATCC (CRL-2422), Nora Navone (University of Texas, Houston, TX, USA)	(11)
PPC-1	No authentic stock is known	(12)
LNCaP-C4-2	ATCC (CRL-3315)	(10)
ARCaP	Haiyen Zhau (University of Virginia, Charlottesville, VA, USA)	(7)
LNCaP-FGC	ATCC (CRL-1740)	(13)
PZ-HPV-7	ATCC (CRL-2221)	(14)
PWR-1E	ATCC (CRL-11611)	(15)
DuCaP	Kenneth Pienta (University of Michigan, Ann Arbor, MI, USA)	(16)
CWR-R1	Christopher Gregory (University of North Carolina, Chapel Hill, NC, USA	(17)
ALVA-31	No authentic stock is known	(18)
RWPE-2	ATCC (CRL-11610)	(19)
CA-HPV-10	ATCC (CRL-2220)	(14)
MDA Pca 2a	Nora Navone (University of Texas, Houston, TX, USA)	(11)
PSK-1	Chol Jang Kim (Shiga University of Medical Science, Ottsu, Japan)	(20)
PC-93	Gert Jan van Steenbrugge (Erasmus University, Rotterdam, Netherlands)	(21)
ALVA-41	No authentic stock is known	(22)
PC-346C	Wytske van Weerden (Erasmus University, Rotterdam, Netherlands)	(23)
E006AA	Walter Rayford (Louisiana State University-Health Sciences Center, New Orleans, LA, USA)	(24)
NCI-H660	ATCC (CRL-5813)	(25)
P69SV40T	No authentic stock is known	(26)
ALVA-101	Steven Loop (American Lake Veterans Administration Hospital, Tacoma, WA, USA)	(22)
DuPro-1	William Isaacs (Johns Hopkins University, Baltimore, MA, USA)	(27)
WPMY-1 normal prostatic myofibroblasts	ATCC (CRL-2854)	(28)
UM-SCP-1	Herb Barton Grossman (University of Texas, Houston, TX, USA)	(29)
LAPC-3	No authentic stock is known	(9)
RC-77T/E	Clayton Yates (Tuskegee University, Tuskegee, AL, USA)	(30)
ALVA-55	Steven Loop (American Lake Veterans Administration Hospital, Tacoma, WA, USA)	(31)
KuCaP13	No authentic stock is known	(32)
LASCPC-01	ATCC (CRL-3356) Owen N. Witte (University of California, Los Angeles, CA, USA)	(33)
hTERT EP156T	ATCC (CRL-3289)	(34)
LNCaP-LN-3	Korean Cell Line Bank 80018	(35)
PNF-08	Gerhard Unteregger (University of Saarland Medical School, Germany)	(36)
LAPC9	Noah Craft (Department of Medicine, Molecular Biology Institute, University of California, Los Angeles, USA)	(37)
LNCaP104S	John Kokontis (Ben May Institute, University of Chicago, Illinois, USA)	(38)
LuCAP35	Eva Corey (Department of Urology, University of Washington, Seattle, Washington 98195, USA	(39)
ACRJ-PC28	The University of the West Indies, Mona, Kingston, Jamaica, West Indies	(40)

cBioPortal Name	Number of samples	Number of patients
Race Differences in Prostate Cancer (MSK, 2021) (41)	2448	2435
Prostate Cancer (MSK, JCO Precis Oncol 2017) (42)	504	401

Supplementary Table 1: Prostate cancer cell lines analysed in this study.

Variable	Source	Data available, n (%)
Number of citations	Ovid MEDLINE Web Of Science Core Collection	45 (100)
Number of patents	World Intellectual Property PATENTSCOPE database	45 (100)
Anatomical Location of	Original cell line papers (18, 19, 21-57),	45 (100)

Prostate Adenocarcinoma (TCGA, Firehose Legacy) (43)	501	500
Prostate Adenocarcinoma (MSK/DFCI, Nature Genetics 2018) (44)	1013	1013
Prostate Adenocarcinoma (MSK, Eur Urol 2020) (45)	1465	1465
Prostate Adenocarcinoma (MSK, Clin Cancer Res. 2022) (46)	1417	1417
Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019) (47)	444	429
Neuroendocrine Prostate Cancer (Multi-Institute, Nat Med 2016)	114	81
Metastatic castration-sensitive prostate cancer (MSK, Clin Cancer Res 2020) (48)	424	424
Prostate Adenocarcinoma (MSK, Cancer Cell 2010) (49)	240	238
Prostate Cancer (DKFZ, Cancer Cell 2018) (50)	324	292

Supplementary Table 2: Clinical cohorts accessed on cBioPortal used in this study.

sample collection	Cellosaurus website, ATCC website, COSMIC database	
Patient age at sample collection	Original cell line papers (18, 19, 21-57), Cellosaurus website, ATCC website, COSMIC database	27 (60)
Geographical location of sample collection	Original cell line papers (18, 19, 21-57)	45 (100)

Supplementary Table 3: Baseline characteristics investigated for each cell line, COSMIC = Catalogue of Somatic Mutations in Cancer

All	Year of sample collection Number of citations Number of patents Anatomical location of sample collection Cancer Type		DuCaP	Age at sample collection COSMIC data Ethnicity M stage	CCLE data MSI Gleason grade
LNCaP	Age at sample collection COSMIC data MSI Ethnicity M stage	CCLE data Gleason grade	CWR-R1	Ethnicity M stage	Age at sample collection COSMIC data CCLE data MSI Gleason grade
PC-3	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade	ALVA-31	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
DU145	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade	RWPE-2	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
1013 L	COSMIC data MSI Ethnicity	Age at sample collection CCLE data Gleason grade M stage	CA-HPV-10	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
22Rv1	COSMIC data CCLE data Ethnicity Gleason grade	Age at sample collection MSI M stage	MDA Pca 2a	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
VCaP	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade	PSK-1	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
MDA PCa 1	Age at sample collection MSI Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	PC-93	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
CWR22	COSMIC data MSI Ethnicity Gleason grade M stage	Age at sample collection CCLE data	ALVA-41	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
LAPC-4(LAPC-4-1)	COSMIC data M stage	Age at sample collection CCLE data MSI Ethnicity Gleason grade	PC-346C	Gleason grade M stage	Age at sample collection COSMIC data CCLE data MSI Ethnicity
LNCaP-C4	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	E006AA	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
MDA Pca 2b	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade	NCI-H660	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade
PPC-1	Age at sample collection COSMIC data Ethnicity	CCLE data MSI Gleason grade M stage	P69SV40T	Age at sample collection COSMIC data Ethnicity Gleason grade M stage	CCLE data MSI
LNCaP-C4-2	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	ALVA-101	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
ARCaP	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	DuPro-1	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
LNCaP-FGC	Age at sample collection COSMIC data CCLE data MSI Ethnicity M stage	Gleason grade	WPMY-1 normal prostatic myofibroblasts	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
PZ-HPV-7	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI	UM-SCP-1	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade
PWR-1E	Age at sample collection MSI Ethnicity Gleason grade M stage	COSMIC data CCLE data	PNF-08	Ethnicity Gleason grade M stage	Age at sample collection COSMIC data CCLE data
ALVA-55	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	LAPC9	COSMIC data M stage	Age at sample collection CCLE data MSI Ethnicity Gleason grade
KuCaP13	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI Gleason grade	LNCaP104S	Age at sample collection COSMIC data Ethnicity M stage	CCLE data MSI Gleason grade
LASCPC-01	Age at sample collection MSI Ethnicity M stage	COSMIC data CCLE data Gleason grade	LuCAP35	Age at sample collection COSMIC data Ethnicity M stage	CCLE data MSI Gleason grade
hTERT EP156T	Age at sample collection Gleason grade M stage	COSMIC data CCLE data MSI Ethnicity	LAPC-3	COSMIC data M stage	Age at sample collection CCLE data MSI Ethnicity Gleason grade
LNCaP-LN-3	Age at sample collection Ethnicity M stage	COSMIC data CCLE data MSI	RC-77T/E	Age at sample collection Ethnicity Gleason grade M stage	COSMIC data CCLE data MSI
ACRJ-PC28	Age at sample collection Ethnicity M stage Gleason grade	CCLE data COSMIC data MSI			

Supplementary Table 4 : Data available for each cell line

Variable	Source
Patient age at sample collection (radical prostatectomy)	Systematic Review on robotic-assisted radical prostatectomy (51)
Geographical distribution of prostate cancer (incidence)	GCO, IARC, WHO Cancer Today Website (52, 53, 54)
Ethnicity of prostate cancer patients	JAMA worldwide racial demographics of prostate cancer study and SEER database (55)
Gleason grade of prostate cancer patients	Weerakoon et. al The current use of active surveillance in an Australian cohort of men (56)
Prostate cancer type	Ackerman et.al Prostate surgical pathology (57)

Supplementary Table 5: Sources of worldwide Prostate Cancer data. GCO = Global Cancer Observatory, IARC = International Agency for Research on Cancer, WHO = World Health Organization. JAMA = The Journal of the American Medical Association, SEER = The Surveillance, Epidemiology, and End Results Program.

Variable	Variable Type	Variable options detailed

Cell Line	Patients with one	Patients with 2-4 alterations	Patients with 5-6 alterations
Mutation	Categorical	Missense mutation Truncating mutation Inframe mutation Splice	
Copy Number Alteration	Categorical	Amplification Deletion	
Structural Variant	Categorical	Fusion	
mRNA expression	Continuous	Low (Z score <2 relative to rest of cohort) High (Z score>2 relative to rest of cohort)	
Protein expression	Continuous	Low (Z score <2 relative to rest of cohort) High (Z score >2 relative to rest of cohort)	
Microsatellite Instability	Categorical	Stable Instable	

Supplementary Table 6: Biological variables investigated for each cell line.

	alteration the same N (%)	the same N (%)	the same N (%)
NCIH660	1299 (90.21)	141 (9.79)	0 (0)
VCaP	1897 (74.71)	625 (24.62)	17 (0.67)
MDA PCA 2b	661 (82.94)	134 (16.81)	2 (0.25)
DU145	1408 (87.95)	193 (12.05)	0 (0)
LNCaP-FGC	1738 (82.02)	380 (19.93)	1 (0.05)
22Rv1	1462 (84.95)	256 (14.88)	3 (0.17)
PC-3	1352 (83.92)	258 (16.02)	1 (0.06)

Supplementary Table 7: Number of patients with alterations the same as cell line

Cell Line	Age at sample collection	Anatomical location of sample collection	Ethnicity	Cancer type	Androgen independent or sensitive (AI/AS)
LNCaP	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma	AS
PC-3	62	Lumbar vertebrae	White	Prostate adenocarcinoma	AI
DU145	69	Brain	White	Prostate adenocarcinoma	AI
1013 L	NA	Prostate	White	Non-acinar urothelial prostate carcinoma	AI
22Rv1	NA	Prostate	Mixed	Prostate adenocarcinoma	AI
VCaP	59	Spinal cord	White	Prostate adenocarcinoma	AI
MDA PCa 1	63	Ascites	Black	Prostate adenocarcinoma	AI
CWR22	NA	Prostate	Mixed	Prostate adenocarcinoma	AI
LAPC-4(LAPC-4-1)	NA	Lymph node	NA	Prostate adenocarcinoma	AI
LNCaP-C4	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma	AI
MDA Pca 2b	63	Bone	Black	Prostate adenocarcinoma	AI
PPC-1	62	Prostate	White	Prostate adenocarcinoma	AI
LNCaP-C4-2	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma	AI
ARCaP	83	Ascites	White	Prostate adenocarcinoma	AI

LNCaP-FGC	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma			AS
PZ-HPV-7	70	Prostate	White	Prostate adenocarcinoma			AS
PWR-1E	67	Prostate	White	Bladder cancer			AS
DuCaP	59	Dura mater	White	Prostate adenocarcinoma			AS
CWR-R1	NA	Prostate	NA	Prostate adenocarcinoma			AI
ALVA-31	62	Prostate	White	Prostate adenocarcinoma			AS
RWPE-2	54	Prostate	White	Normal prostate			AS
CA-HPV-10	63	Prostate	White	Prostate adenocarcinoma			AI
MDA Pca 2a	63	Ascites	Black	Prostate adenocarcinoma			AI
PSK-1	29	Prostate	Asian	Small cell carcinoma			AI
PC-93	30.5	Prostate	Black	Prostate adenocarcinoma			AS
ALVA-41	62	Bone	White	Prostate adenocarcinoma			AS
PC-346C	NA	Prostate	NA	Prostate adenocarcinoma			AI
E006AA	58	Prostate	Black	Prostate adenocarcinoma			AS
NCI-H660	63	Prostate	White	Small cell carcinoma			AI
P69SV40T	63	Prostate	Black	Normal prostate			NA
ALVA-101	62	Bone	White	Prostate adenocarcinoma			AS
DuPro-1	62	Prostate	White	Prostate adenocarcinoma			AI
WPMY-1 normal prostatic myofibroblasts	54	Prostate	White	Normal prostate			AS
UM-SCP-1	71	Prostate	White	Squamous cell carcinoma			AI
LAPC-3	NA	Prostate	NA	Prostate adenocarcinoma			AI
RC-77T/E	63	Prostate	Black	Prostate adenocarcinoma			AS
ALVA-55	62	Lymph node	White	Prostate adenocarcinoma			AI
KuCaP13	60	Penis	Asian	Neuroendocrine carcinoma			AI
LASCPC-01	75-80	Prostate	White	Neuroendocrine carcinoma			AI
hTERT EP156T	66	Prostate	NA	Prostate adenocarcinoma			AS
LNCaP-LN-3	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma			AS
PNF-08	NA	Prostate	White	Normal prostate			AS
LAPC9	NA	Bone	NA	Prostate adenocarcinoma			AI
LNCaP104S	50	Left supraclavicular lymph node	White	Prostate adenocarcinoma			AS
LuCAP35	66	Left inguinal lymph node	White	Prostate adenocarcinoma			AS
ACRJ-PC28	69	Prostate	Black	Prostate adenocarcinoma			AS

	Cell Line p value	Clinical cohort p value	Worldwide p value
D'Agostino & Pearson Test	<0.0001	0.2983 (ns)	<0.0001
Anderson-Darling test	0.0015	0.0549 (ns)	<0.0001
Shapiro-Wilkins test	0.0008	0.0384	<0.0001
Kolmogorov-Smirnov test	0.0181	>0.100 (ns)	<0.0001

Supplementary Table 8 : Key clinical variables for each cell line

Supplementary Table 9: Results of normality distribution tests for distribution of age at sample collection.

Treatment	Clinical Cohort (%)	Cell Line (%)
No treatment	8.99	30.40
ADT	1.45	30.40
Chemotherapy	0.59	13.04
Unknown	88.97	28.26

Supplementary Table 10: Table demonstrating percentage of cell lines, or patients receiving different forms of treatment for prostate cancer prior to sampling

cBioPortal Name	mRNA Expression	Protein Expression	Mutations	Copy Number Alterations	Structural Variants	Microsatellite Instability
Race Differences in Prostate Cancer (MSK, 2021) (41)	?		?	?	?	
Prostate Cancer (MSK, JCO Precis Oncol 2017) (42)			?	?	?	
Prostate Adenocarcinoma (TCGA, Firehose Legacy) (43)	?	?	?	?		
Prostate Adenocarcinoma (MSK/DFCI, Nature Genetics 2018) (44)	?	?	?	?		
Prostate Adenocarcinoma (MSK, Eur Urol 2020) (45)			?	?	?	?
Prostate Adenocarcinoma (MSK, Clin Cancer Res. 2022) (46)			?	?	?	?
Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019) (47)	?		?	?	?	

Neuroendocrine Prostate Cancer (Multi-Institute, Nat Med 2016)	?		?			
Metastatic castration-sensitive prostate cancer (MSK, Clin Cancer Res 2020) (48)			?	?	?	?
Prostate Adenocarcinoma (MSK, Cancer Cell 2010) (49)	?		?	?		
Prostate Cancer (DKFZ, Cancer Cell 2018) (50)	?		?			

Supplementary Table 11: Biological alterations screened for in clinical cohorts accessed on cBioPortal.

1. Horoszewicz JS, Leong SS, Chu TM, Wajsman ZL, Friedman M, Papsidero L, et al. The LNCaP cell line--a new model for studies on human prostatic carcinoma. Prog Clin Biol Res. 1980;37:115-32.
2. Kaighn ME, Lechner JF, Narayan KS, Jones LW. Prostate carcinoma: tissue culture cell lines. Natl Cancer Inst Monogr. 1978(49):17-21.
3. Stone KR, Mickey DD, Wunderli H, Mickey GH, Paulson DF. Isolation of a human prostate carcinoma cell line (DU 145). Int J Cancer. 1978;21(3):274-81.
4. Williams RD. Human urologic cancer cell lines. Invest Urol. 1980;17(5):359-63.
5. Sramkoski RM, Pretlow TG, 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS, et al. A new human prostate carcinoma cell line, 22Rv1. In Vitro Cell Dev Biol Anim. 1999;35(7):403-9.
6. Korenchuk S, Lehr JE, L MC, Lee YG, Whitney S, Vessella R, et al. VCaP, a cell-based model system of human prostate cancer. In Vivo. 2001;15(2):163-8.
7. Zhau HY, Chang SM, Chen BQ, Wang Y, Zhang H, Kao C, et al. Androgen-repressed phenotype in human prostate cancer. Proc Natl Acad Sci U S A. 1996;93(26):15152-7.
8. Pretlow TG, Wolman SR, Micale MA, Pelley RJ, Kursh ED, Resnick MI, et al. Xenografts of primary human prostatic carcinoma. J Natl Cancer Inst. 1993;85(5):394-8.
9. Klein KA, Reiter RE, Redula J, Moradi H, Zhu XL, Brothman AR, et al. Progression of metastatic human prostate cancer to androgen independence in immunodeficient SCID mice. Nat Med. 1997;3(4):402-8.
10. Wu HC, Hsieh JT, Gleave ME, Brown NM, Pathak S, Chung LW. Derivation of androgen-independent human LNCaP prostatic cancer cell sublines: role of bone stromal cells. Int J Cancer. 1994;57(3):406-12.
11. Navone NM, Olive M, Ozen M, Davis R, Troncoso P, Tu SM, et al. Establishment of two human prostate cancer cell lines derived from a single bone metastasis. Clin Cancer Res. 1997;3(12 Pt 1):2493-500.
12. Brothman AR, Lesho LJ, Somers KD, Wright GL, Jr., Merchant DJ. Phenotypic and cytogenetic characterization of a cell line derived from primary prostatic carcinoma. Int J Cancer. 1989;44(5):898-903.
13. Carroll AG, Voeller HJ, Sugars L, Gelmann EP. p53 oncogene mutations in three human prostate cancer cell lines. Prostate. 1993;23(2):123-34.
14. Weijerman PC, König JJ, Wong ST, Niesters HG, Peehl DM. Lipofection-mediated immortalization of human prostatic epithelial cells of normal and malignant origin using human papillomavirus type 18 DNA. Cancer Res. 1994;54(21):5579-83.
15. Webber MM, Bello D, Kleinman HK, Wartinger DD, Williams DE, Rhim JS. Prostate specific antigen and androgen receptor induction and characterization of an immortalized adult human prostatic epithelial cell line. Carcinogenesis. 1996;17(8):1641-6.
16. Lee YG, Korenchuk S, Lehr J, Whitney S, Vessela R, Pienta KJ. Establishment and characterization of a new human prostatic cancer cell line: DuCaP. In Vivo. 2001;15(2):157-62.
17. Gregory CW, Johnson RT, Jr., Mohler JL, French FS, Wilson EM. Androgen receptor stabilization in recurrent prostate cancer is associated with hypersensitivity to low androgen. Cancer Res. 2001;61(7):2892-8.
18. Loop SM, Rozanski TA, Ostenson RC. Human primary prostate tumor cell line, ALVA-31: a new model for studying the hormonal regulation of prostate tumor cell growth. Prostate. 1993;22(2):93-108.

19. Webber MM, Bello D, Quader S. Immortalized and tumorigenic adult human prostatic epithelial cell lines: characteristics and applications Part 2. Tumorigenic cell lines. Prostate. 1997;30(1):58-64.
20. Kim CJ, Kushima R, Okada Y, Seto A. Establishment and characterization of a prostatic small-cell carcinoma cell line (PSK-1) derived from a patient with Klinefelter syndrome. Prostate. 2000;42(4):287-94.
21. Claas FH, van Steenbrugge GJ. Expression of HLA-like structures on a permanent human tumor line PC-93. Tissue Antigens. 1983;21(3):227-32.
22. Plymate SR, Loop SM, Hoop RC, Wiren KM, Ostenson R, Hryb DJ, et al. Effects of sex hormone binding globulin (SHBG) on human prostatic carcinoma. J Steroid Biochem Mol Biol. 1991;40(4-6):833-9.
23. Romijn J.C. E-SS, van Weerden W.M., Teubel W., Trapman J., Schroder F.H. Development and characterization of a new in vitro model to study the mechanisms of prostate cancer progression. Urol Res. 1995;23:P81-265.
24. Koochekpour S, Maresh GA, Katner A, Parker-Johnson K, Lee TJ, Hebert FE, et al. Establishment and characterization of a primary androgen-responsive African-American prostate cancer cell line, E006AA. Prostate. 2004;60(2):141-52.
25. Carney DN, Gazdar AF, Bepler G, Guccion JG, Marangos PJ, Moody TW, et al. Establishment and identification of small cell lung cancer cell lines having classic and variant features. Cancer Res. 1985;45(6):2913-23.
26. Bae VL, Jackson-Cook CK, Brothman AR, Maygarden SJ, Ware JL. Tumorigenicity of SV40 T antigen immortalized human prostate epithelial cells: association with decreased epidermal growth factor receptor (EGFR) expression. Int J Cancer. 1994;58(5):721-9.
27. Gingrich JR, Tucker JA, Walther PJ, Day JW, Poulton SH, Webb KS. Establishment and characterization of a new human prostatic carcinoma cell line (DuPro-1). J Urol. 1991;146(3):915-9.
28. Webber MM, Trakul N, Thraves PS, Bello-DeOcampo D, Chu WW, Storto PD, et al. A human prostatic stromal myofibroblast cell line WPMY-1: a model for stromal-epithelial interactions in prostatic neoplasia. Carcinogenesis. 1999;20(7):1185-92.
29. Grossman HB, Wedemeyer G, Ren L, Carey TE. UM-SCP-1, a new human cell line derived from a prostatic squamous cell carcinoma. Cancer Res. 1984;44(9):4111-7.
30. Theodore S, Sharp S, Zhou J, Turner T, Li H, Miki J, et al. Establishment and characterization of a pair of non-malignant and malignant tumor derived cell lines from an African American prostate cancer patient. Int J Oncol. 2010;37(6):1477-82.
31. Mehta PP, Lokeshwar BL, Schiller PC, Bendix MV, Ostenson RC, Howard GA, et al. Gap-junctional communication in normal and neoplastic prostate epithelial cells and its regulation by cAMP. Mol Carcinog. 1996;15(1):18-32.
32. Okasho K, Mizuno K, Fukui T, Lin YY, Kamiyama Y, Sunada T, et al. Establishment and characterization of a novel treatment-related neuroendocrine prostate cancer cell line KUCaP13. Cancer Sci. 2021;112(7):2781-91.
33. Lee JK, Phillips JW, Smith BA, Park JW, Stoyanova T, McCaffrey EF, et al. N-Myc Drives Neuroendocrine Prostate Cancer Initiated from Human Prostate Epithelial Cells. Cancer Cell. 2016;29(4):536-47.
34. Kogan I, Goldfinger N, Milyavsky M, Cohen M, Shats I, Dobler G, et al. hTERT-immortalized prostate epithelial and stromal-derived cells: an authentic in vitro model for differentiation and carcinogenesis. Cancer Res. 2006;66(7):3531-40.

35. Pettaway CA, Pathak S, Greene G, Ramirez E, Wilson MR, Killion JJ, et al. Selection of highly metastatic variants of different human prostatic carcinomas using orthotopic implantation in nude mice. Clin Cancer Res. 1996;2(9):1627-36.
36. Szczyrba J, Löprich E, Wach S, Jung V, Unteregger G, Barth S, et al. The microRNA profile of prostate carcinoma obtained by deep sequencing. Mol Cancer Res. 2010;8(4):529-38.
37. Craft N, Chhor C, Tran C, Belldegrun A, DeKernion J, Witte ON, et al. Evidence for clonal outgrowth of androgen-independent prostate cancer cells from androgen-dependent tumors through a two-step process. Cancer Res. 1999;59(19):5030-6.
38. Kokontis J, Takakura K, Hay N, Liao S. Increased androgen receptor activity and altered c-myc expression in prostate cancer cells after long-term androgen deprivation. Cancer Res. 1994;54(6):1566-73.
39. Corey E, Quinn JE, Buhler KR, Nelson PS, Macoska JA, True LD, et al. LuCaP 35: a new model of prostate cancer progression to androgen independence. Prostate. 2003;55(4):239-46.
40. Valentine H, Aiken W, Morrison B, Zhao Z, Fowle H, Wasserman JS, et al. Expanding the Prostate Cancer Cell Line Repertoire with ACRJ-PC28, an AR-negative Neuroendocrine Cell Line Derived From an African-Caribbean Patient. Cancer Research Communications. 2022;2(11):1355-71.
41. Stopsack KH, Nandakumar S, Arora K, Nguyen B, Vasselman SE, Nweji B, et al. Differences in Prostate Cancer Genomes by Self-reported Race: Contributions of Genetic Ancestry, Modifiable Cancer Risk Factors, and Clinical Factors. Clinical Cancer Research. 2022;28(2):318-26.
42. Abida W, Armenia J, Gopalan A, Brennan R, Walsh M, Barron D, et al. Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making. JCO Precision Oncology. 2017(1):1-16.
43. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry Christopher D, et al. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163(4):1011-25.
44. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nature Genetics. 2018;50(5):645-51.
45. Nguyen B, Mota JM, Nandakumar S, Stopsack KH, Weg E, Rathkopf D, et al. Pan-cancer Analysis of CDK12 Alterations Identifies a Subset of Prostate Cancers with Distinct Genomic and Clinical Characteristics. European Urology. 2020;78(5):671-9.
46. Chakraborty G, Nandakumar S, Hirani R, Nguyen B, Stopsack KH, Kreitzer C, et al. The Impact of PIK3R1 Mutations and Insulin–PI3K–Glycolytic Pathway Regulation in Prostate Cancer. Clinical Cancer Research. 2022;28(16):3603-17.
47. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical outcome in advanced prostate cancer. Proceedings of the National Academy of Sciences. 2019;116(23):11428-36.
48. Stopsack KH, Nandakumar S, Wibmer AG, Haywood S, Weg ES, Barnett ES, et al. Oncogenic Genomic Alterations, Clinical Phenotypes, and Outcomes in Metastatic Castration-Sensitive Prostate Cancer. Clinical Cancer Research. 2020;26(13):3230-8.
49. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative Genomic Profiling of Human Prostate Cancer. Cancer Cell. 2010;18(1):11-22.
50. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell. 2018;34(6):996-1011.e8.

51. Du Y, Long Q, Guan B, Mu L, Tian J, Jiang Y, et al. Robot-Assisted Radical Prostatectomy Is More Beneficial for Prostate Cancer Patients: A System Review and Meta-Analysis. Medical science monitor : international medical journal of experimental and clinical research. 2018;24:272-87.
52. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F Global Cancer Observatory: Cancer Today. : Lyon, France: International Agency for Research on Cancer; 2020 [Available from: https://gco.iarc.fr/today,.
53. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.
54. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941-53.
55. Fletcher SA, Marchese M, Cole AP, Mahal BA, Friedlander DF, Krimphove M, et al. Geographic Distribution of Racial Differences in Prostate Cancer Mortality. JAMA Network Open. 2020;3(3):e201839-e.
56. Weerakoon M, Papa N, Lawrentschuk N, Evans S, Millar J, Frydenberg M, et al. The current use of active surveillance in an Australian cohort of men: a pattern of care analysis from the Victorian Prostate Cancer Registry. BJU International. 2015;115(S5):50-6.
57. Ackerman LA, Rosai J. Surgical pathology. Surgical pathology1974. p. 1394-.