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

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.

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

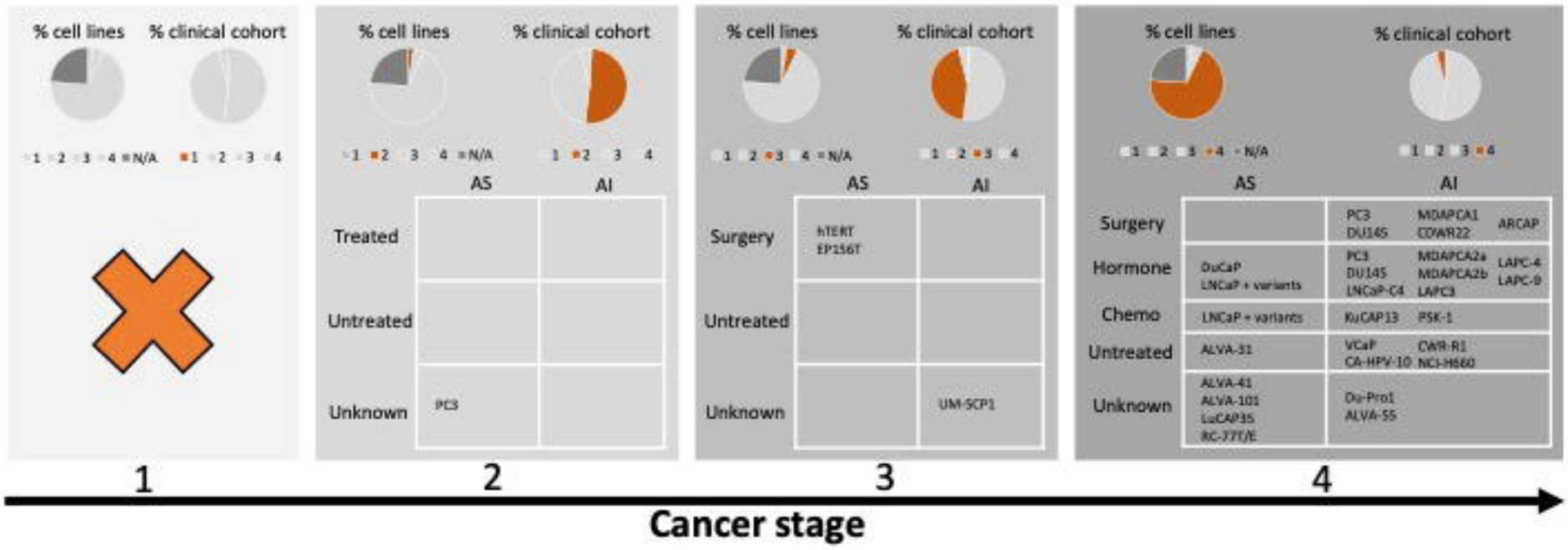
Figure 3 Cell lines and clinical cohorts have different clinical features to prostate cancer patients worldwide and the most common alterations in clinical datasets are not necessarily altered in cell lines : A Stacked bar chart showing the ethnicity distribution of original patients cell lines were 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.

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

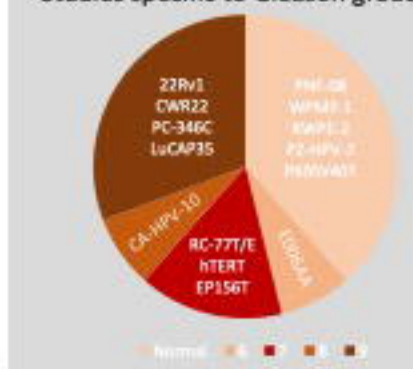
PC3

Most generalisable (validational)

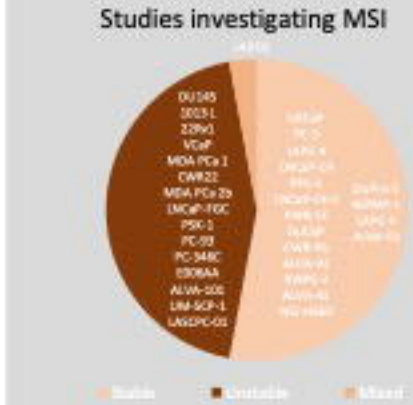
- More data exists for PC3 than any other PCa cell line, representing a wealth of knowledge for contextualisation of findings
- Most-cited PCa cell line on PubMed (15,673 citations)
- Most patented PCa cell line (37,891 patents on WIPO and 90,858 patents on Google)

Specific PCa studies

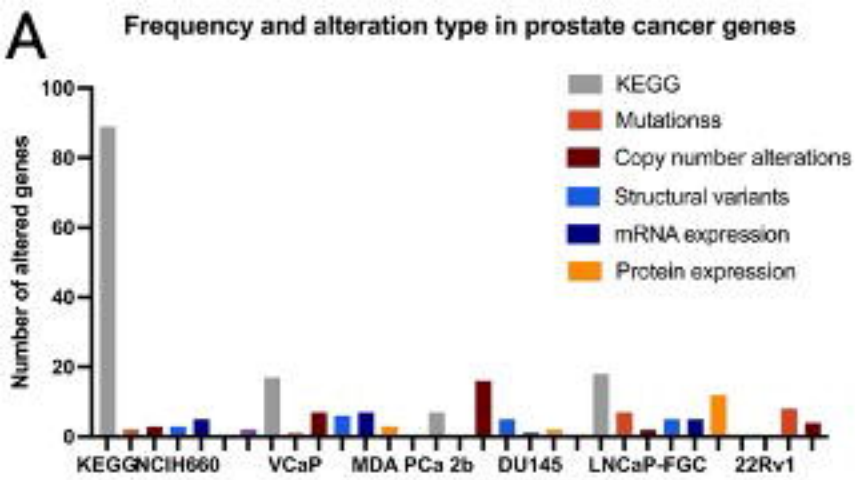
Studies specific to Gleason grade

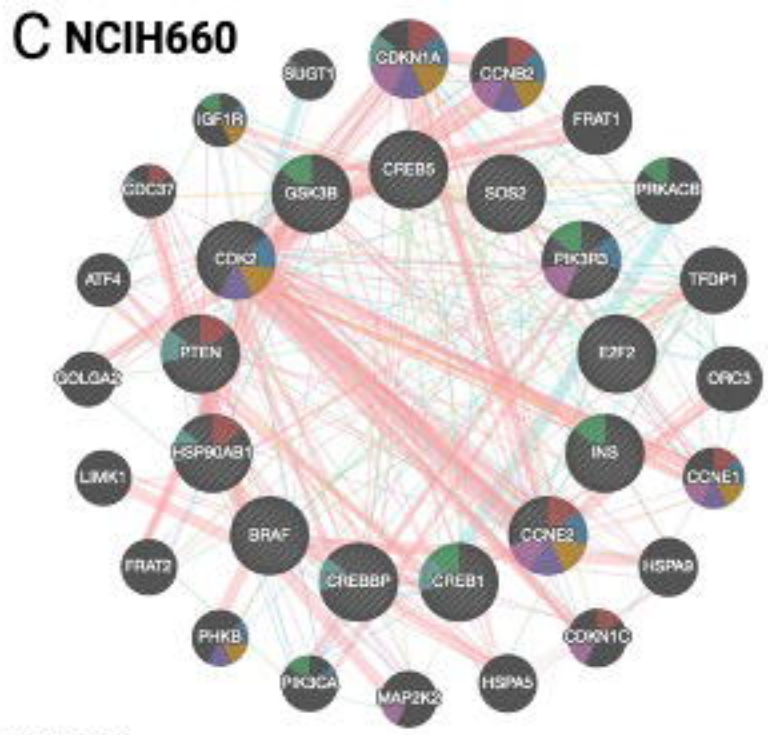


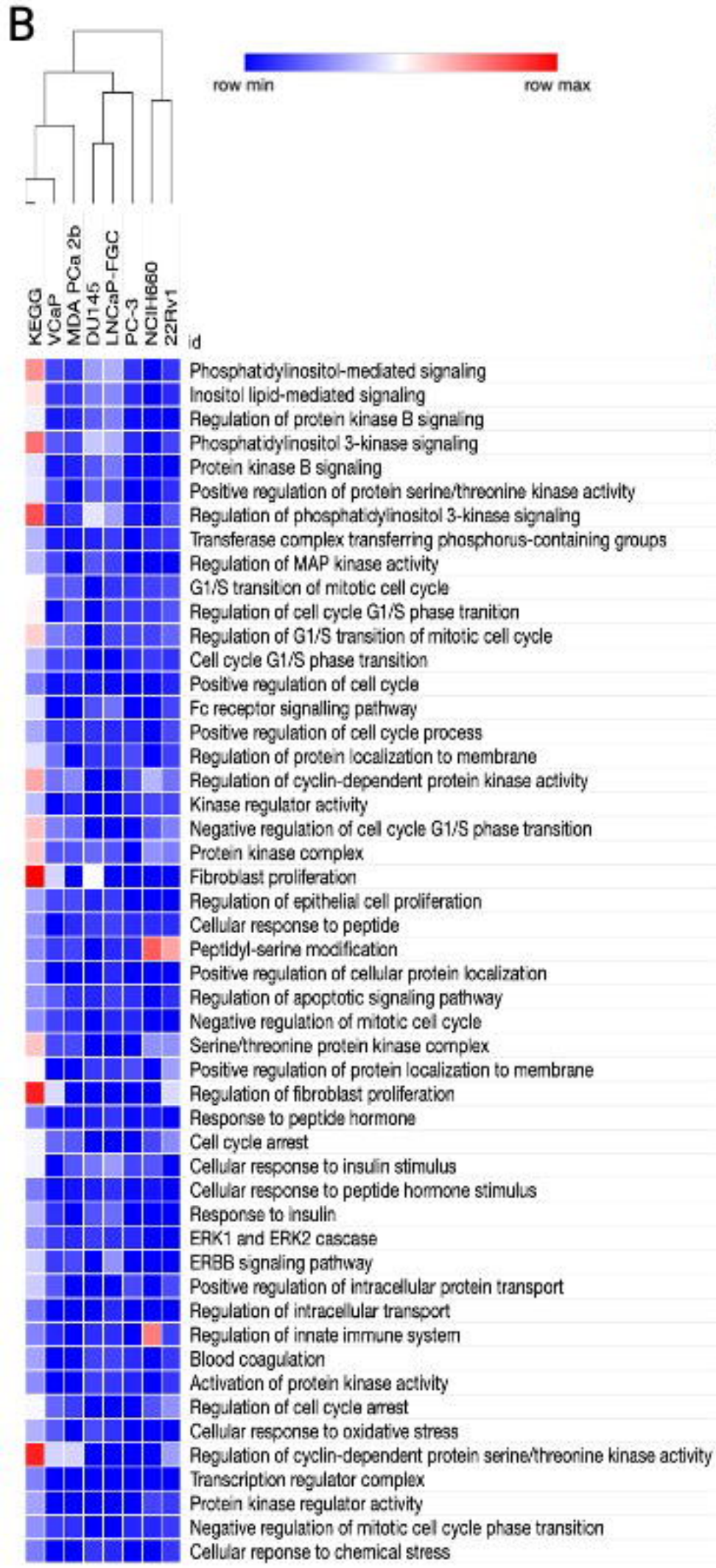
Studies investigating MSI



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





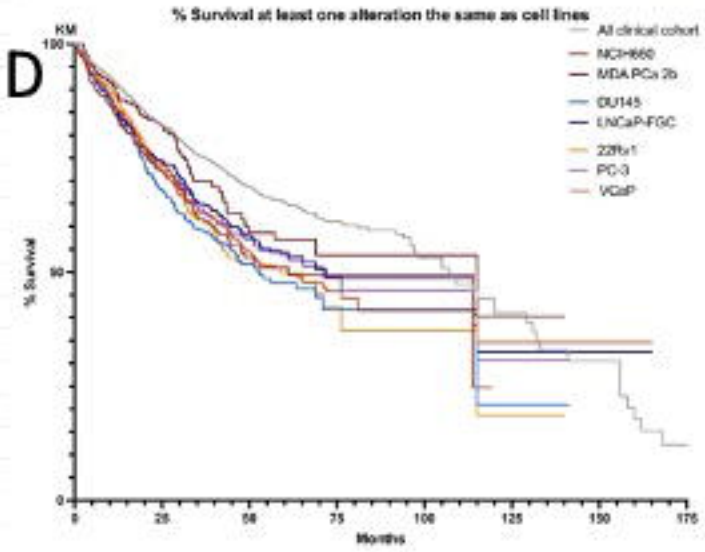


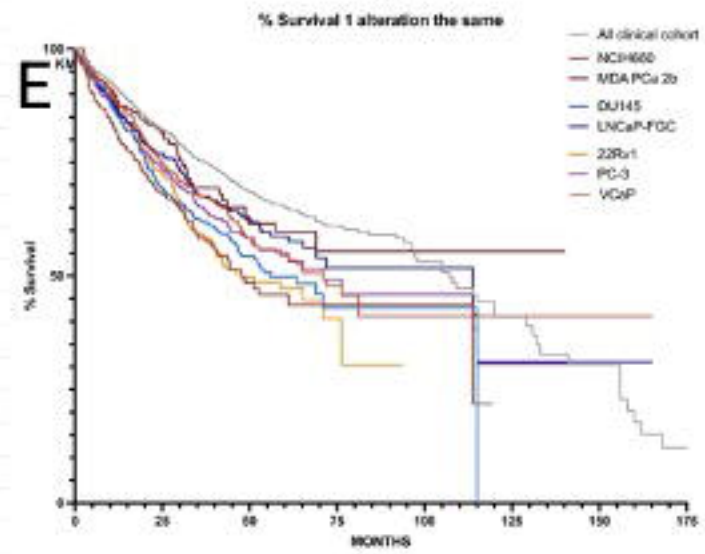
Functions

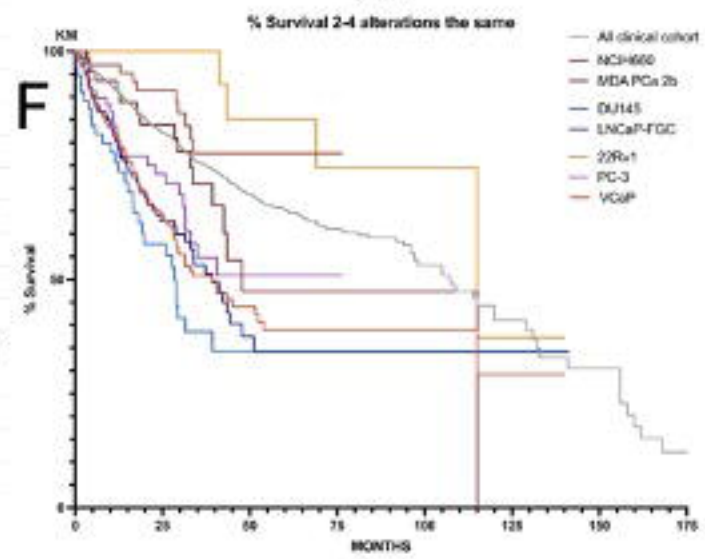
- regulation of cyclin-dependent protein kinase activity
- transferase complex, transferring phosphorus-containing groups
- protein kinase complex
- serine/threonine protein kinase complex
- kinase regulator activity
- regulation of protein stability
- response to peptide hormone

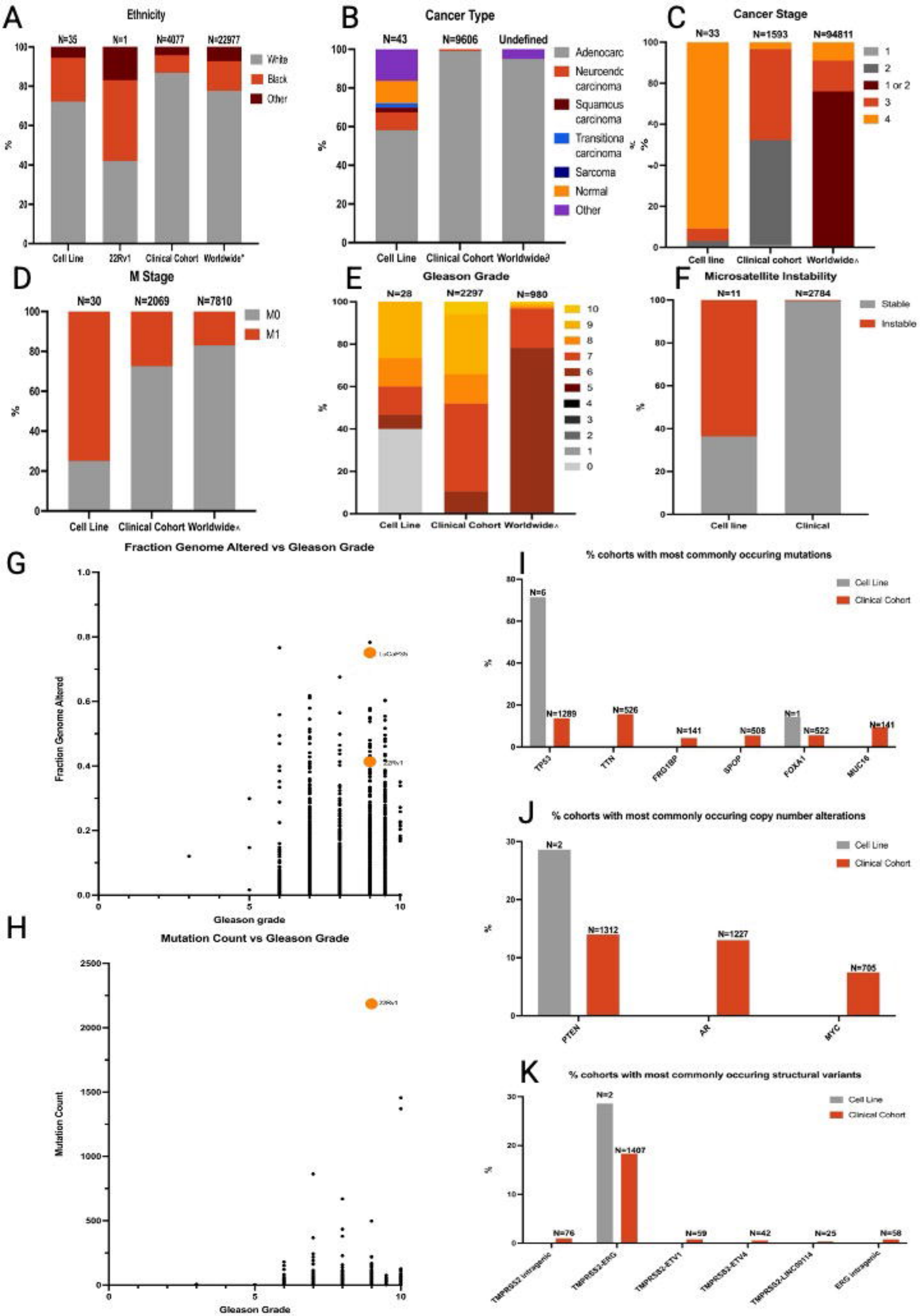
Networks

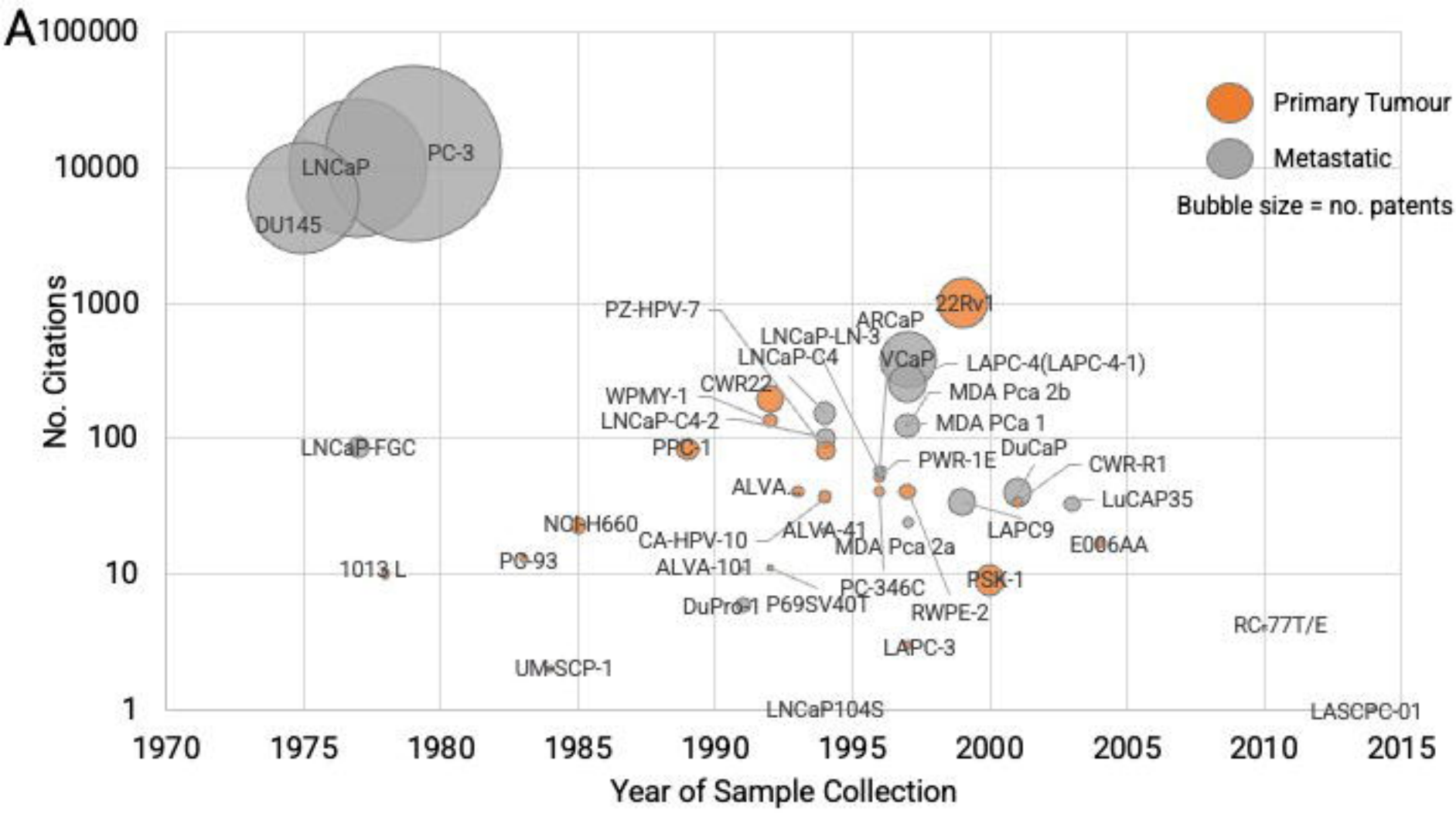
- Physical Interactions
- Co-expression
- Pathway
- Genetic Interactions
- Co-localization
- Predicted
- Shared protein domains

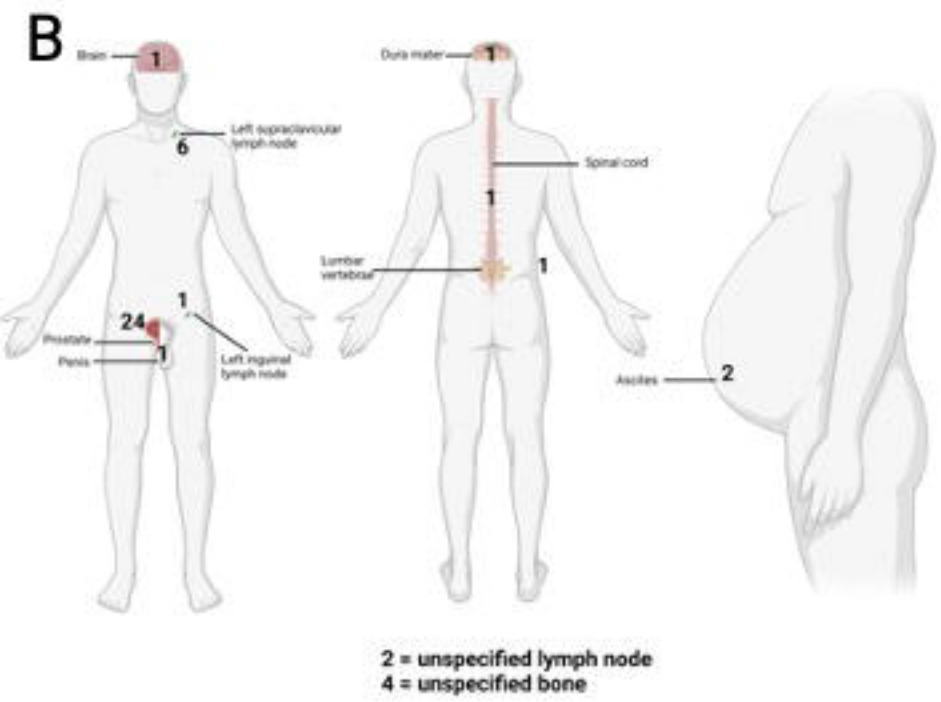


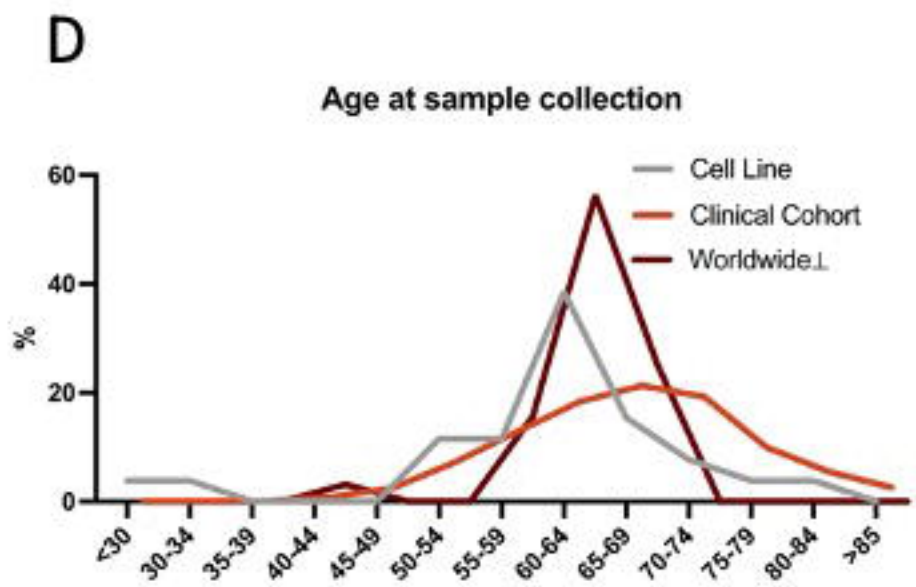


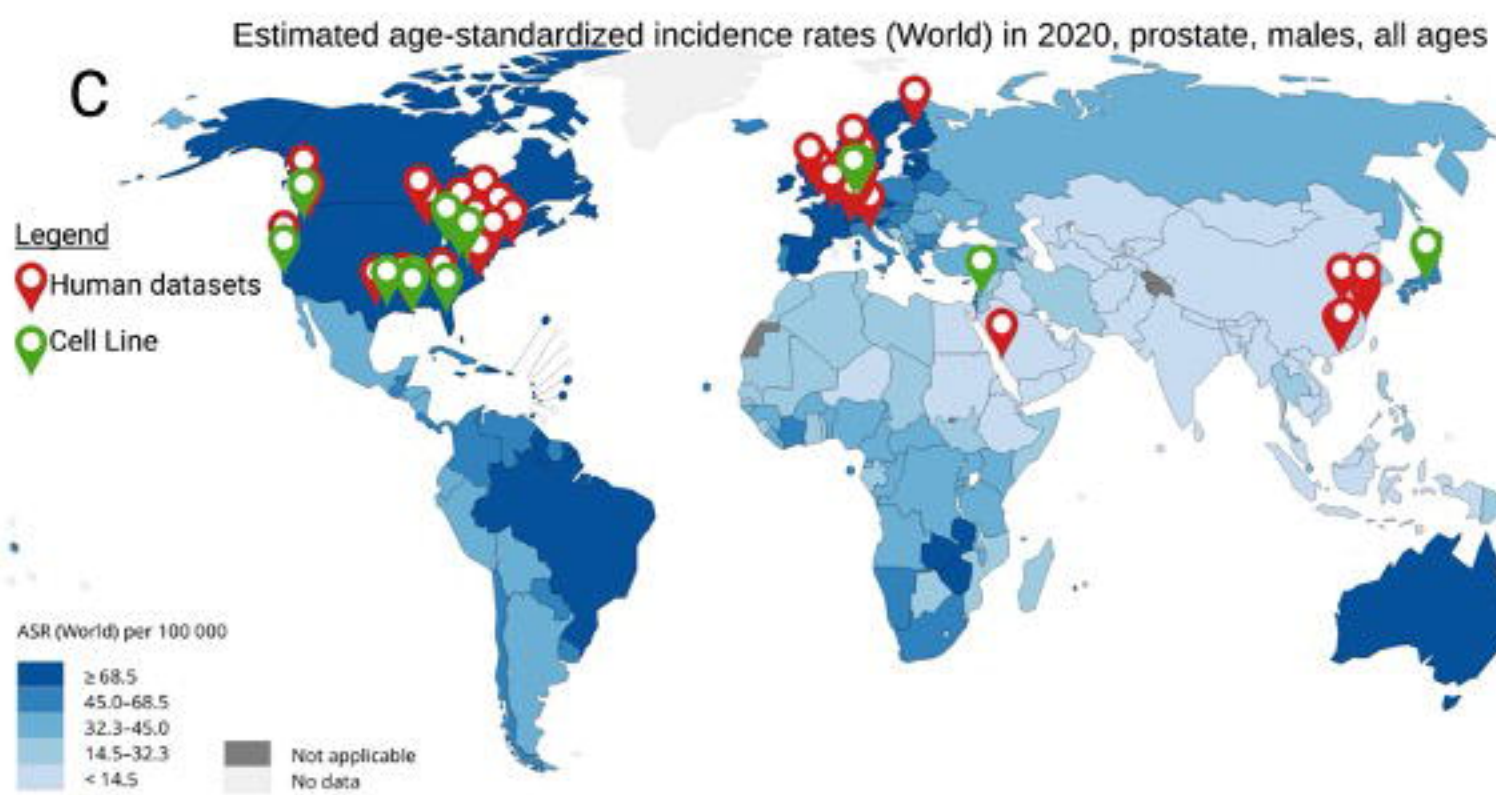



A
Ethnicity
N=35
N=1
N=4077
N=22977
White
Black
Other
%
Cell Line
22Rv1
Clinical Cohort
Worldwide*
B
Cancer Type
N=43
N=9606
Undefined
Adenocarc
Neuroendc carcinoma
Squamous carcinoma
Transitiona carcinoma
Sarcoma
Normal
Other
Cell Line
Clinical Cohort
Worldwide
C
Cancer Stage
N=33
N=1593
N=94811
1
2
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





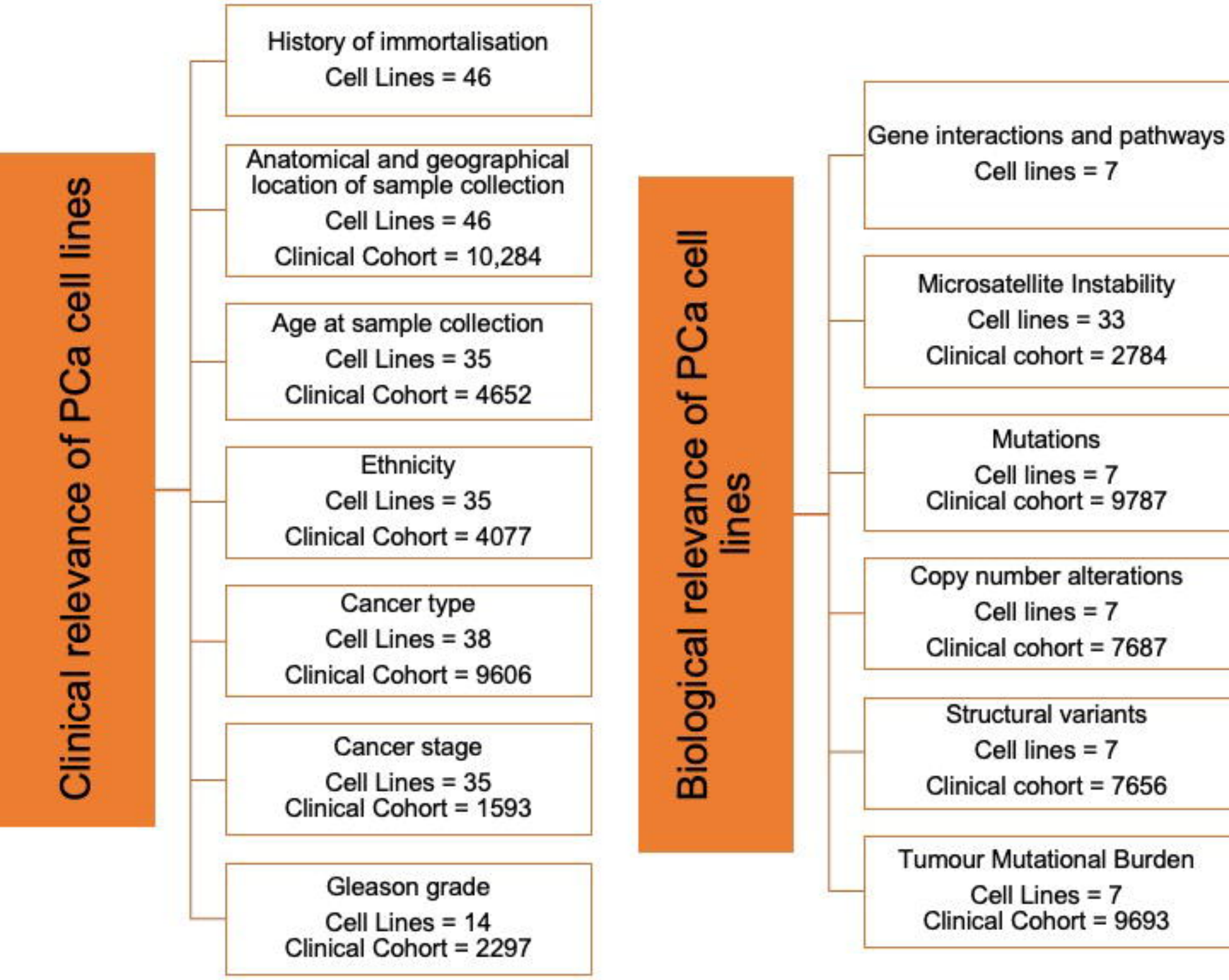




All rights reserved. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization / International Agency for Research on Cancer concerning the legal status of any country, territory, 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

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