Review Article

An overview of publicly available patient-centered prostate cancer datasets

Tim Hulsen

Department of Professional Health Solutions & Services, Philips Research, Eindhoven, The Netherlands
Correspondence to: Tim Hulsen. Department of Professional Health Solutions & Services, Philips Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands. Email: tim.hulsen@philips.com.

Abstract: Prostate cancer (PCa) is the second most common cancer in men, and the second leading cause of death from cancer in men. Many studies on PCa have been carried out, each taking much time before the data is collected and ready to be analyzed. However, on the internet there is already a wide range of PCa datasets available, which could be used for data mining, predictive modelling or other purposes, reducing the need to setup new studies to collect data. In the current scientific climate, moving more and more to the analysis of "big data" and large, international, multi-site projects using a modern IT infrastructure, these datasets could be proven extremely valuable. This review presents an overview of publicly available patient-centered PCa datasets, divided into three categories (clinical, genomics and imaging) and an "overall" section to enable researchers to select a suitable dataset for analysis, without having to go through days of work to find the right data. To acquire a list of human PCa databases, scientific literature databases and academic social network sites were searched. We also used the information from other reviews. All databases in the combined list were then checked for public availability. Only databases that were either directly publicly available or available after signing a research data agreement or retrieving a free login were selected for inclusion in this review. Data should be available to commercial parties as well. This paper focuses on patient-centered data, so the genomics data section does not include gene-centered databases or pathway-centered databases. We identified 42 publicly available, patient-centered PCa datasets. Some of these consist of different smaller datasets. Some of them contain combinations of datasets from the three data domains: clinical data, imaging data and genomics data. Only one dataset contains information from all three domains. This review presents all datasets and their characteristics: number of subjects, clinical fields, imaging modalities, expression data, mutation data, biomarker measurements, etc. Despite all the attention that has been given to making this overview of publicly available databases as extensive as possible, it is very likely not complete, and will also be outdated soon. However, this review might help many PCa researchers to find suitable datasets to answer the research question with, without the need to start a new data collection project. In the coming era of big data analysis, overviews like this are becoming more and more useful.

Keywords: Prostate cancer (PCa); prostate; oncology; databases; public

Submitted Dec 03, 2018. Accepted for publication Feb 27, 2019.
doi: 10.21037/tau.2019.03.01
View this article at: http://dx.doi.org/10.21037/tau.2019.03.01

Introduction

Prostate cancer (PCa) is the second most common cancer in men, and the second leading cause of death from cancer in men (1). Many studies on PCa have been carried out, each taking much time before the data is collected and ready to be analyzed. The datasets created in these studies are usually collected by academic institutes, who are often unwilling to share the data because of concerns over ownership, publications or patient consent. Because of the new privacy regulations in the EU General Data Protection Regulation (GDPR), this data sharing is becoming increasingly more difficult (2). However, on the internet a wide range of PCa datasets are already available, ready to be used for data mining and analysis. Some of them are well-known to researchers in the field, but others remain hidden because they were published in a low-impact journal or are simply not on the first page of Google. Nevertheless, these datasets could be still used for data mining, predictive modelling or other purposes, reducing the need to setup new studies to collect data. In the current scientific climate, moving more and more to the analysis of 'big data' (3,4) and large, international, multi-site projects using a modern IT infrastructure, such as Movember GAP3 (5), ERSPC (6) and PCMM (7), these datasets could be proven extremely valuable.

This review presents an overview of publicly available patient-centered PCa datasets (8), divided into three categories (clinical, genomics and imaging) and an 'overall' section to enable researchers to select a suitable dataset for analysis, without having to go through days of work to find the right data.

The 'Clinical data' section contains datasets that have a number of clinical parameters, i.e., data that can be captured in numerical or text fields. In the area of PCa these are, for example: age, Gleason scores, TNM stages and PSA values, but also values derived from the genomics and imaging domains, such as biomarker expression values and PI-RADS scores (9).

The 'Genomics data' section describes a number of datasets resulting from genomics studies, such as microarray experiments. Websites like cBioPortal (10), GEO (11) and ArrayExpress (12) and apps like camcAPP (13) can be used to browse through genomics datasets.

In the 'Imaging data' section, a number of data sources containing Magnetic Resonance (MR), UltraSound (US), Positron-emission tomography (PET), Computed Tomography (CT) and histopathology images are listed.

The 'Overall' section brings all datasets within this review together, and shows which datasets contain information from more than one domain (clinical/genomics/imaging). It gives a complete picture of all publicly available patient-centered PCa datasets.

Methods

Scientific literature databases and academic social network sites such as PubMed/Medline, Embase, Scopus, ResearchGate, Academia.edu, Google Scholar and Microsoft Academic were searched to acquire a list of human PCa databases (*Figure 1*). We also used the information from other reviews (14) in this paper. All databases in the combined list were then checked for public availability. Only databases that were either directly publicly available or available after signing a research data agreement or retrieving a free login were selected for inclusion in this review. Data should be available to commercial parties as well. This paper focuses on patient-centered data, so the genomics data section does not include gene-centered databases, pathway-centered databases, etc. that are not linked to patients. This exclusion ensures that the genomics data can be more easily combined with the clinical and imaging data.

Results

Clinical data

The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial (15) is a randomized, controlled trial to determine whether certain screening exams reduce mortality from prostate, lung, colorectal and ovarian cancer. A total of 76,682 male participants were enrolled between November 1993 and July 2001. Data collected up to December 31, 2009 for the first 13 years of participation for each subject in the PLCO trial are available at https://biometry.nci.nih.gov/cdas/datasets/plco/20/. Six PCa screening datasets are available:

(I) The Prostate dataset is a comprehensive dataset that contains nearly all the PLCO study data available for PCa screening, incidence, and mortality analyses. The dataset contains one record for each of the 76,682 male participants in the PLCO trial.

(II) The Prostate Screening dataset (177,315 records, 35,875 subjects, one record per year of screening)

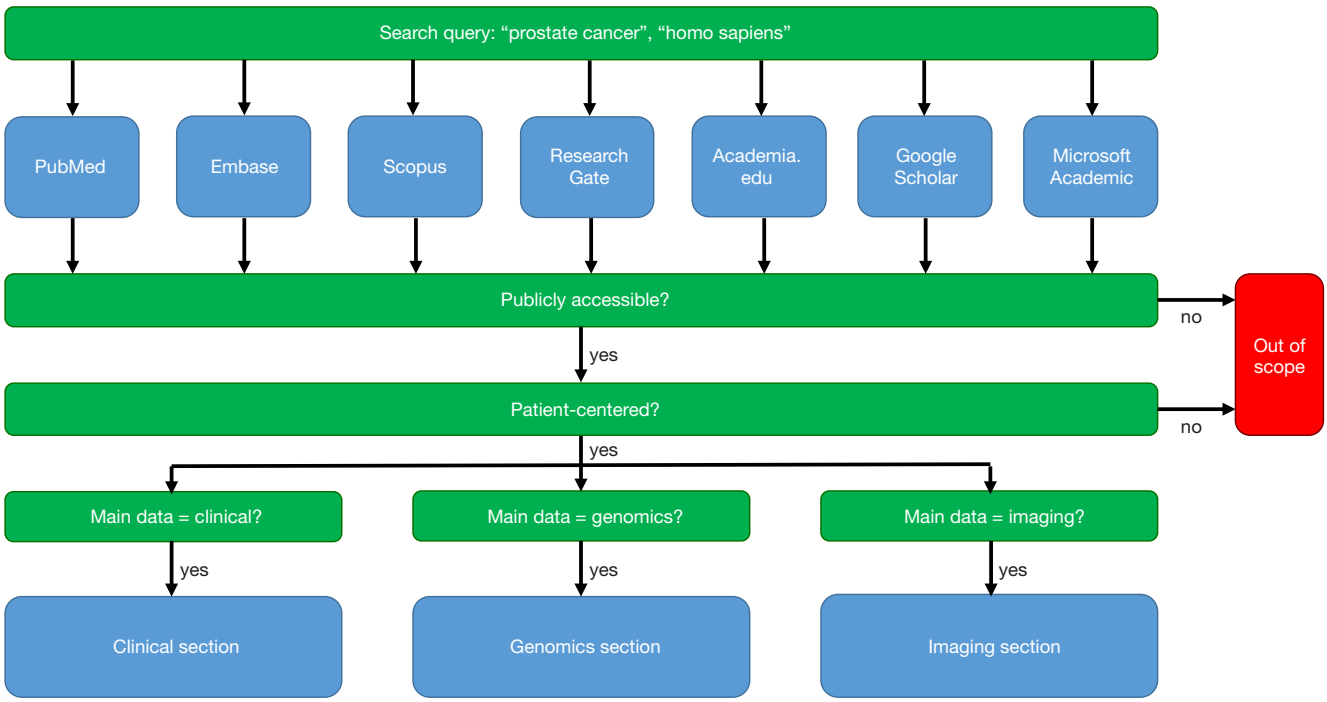

Figure 1 Workflow diagram of the evidence acquisition.

contains additional information from PSA and Digital Rectal Exam (DRE) cancer screens. This includes details of the blood draw, QA DRE results, reasons for inadequate exams, and additional findings that were not suspicious for cancer.

(III) The Prostate Screening Abnormalities dataset (10,527 records, 5,743 subjects, one record per abnormality) contains information for each induration found during the DRE screen. This includes the location, type, size, grade, and extent of each induration.

(IV) The Prostate Diagnostic Procedures dataset (95,837 records, 15,307 subjects, one record per procedure) contains information about the diagnostic procedures prompted by positive PCa screens, as well as diagnostic/staging procedures associated with any PCa diagnosed during the 13 years of follow-up.

(V) The Prostate Medical Complications dataset (3,350 records, 2,164 subjects, one record per medical complication) contains information about the medical complications caused by diagnostic workup for PCa.

(VI) The Prostate Treatments dataset (13,409 records, 7,614 subjects, one record per treatment procedure) contains specifics of the initial treatment following the diagnosis of PCa.

The Surveillance, Epidemiology, and End Results (SEER) database (16) of the National Cancer Institute at https://seer.cancer.gov/data/seerstat/nov2017/ provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program, which provides national leadership in the science of cancer surveillance as well as analytical tools and methodological expertise in collecting, analyzing, interpreting, and disseminating reliable population-based statistics. The SEER research data include SEER incidence and population data associated by age, sex, race, year of diagnosis, and geographic areas (including SEER registry and county). SEER research data are released every Spring based on the previous November's submission of data. A research data agreement needs to be signed and approved before the data can be accessed. The SEER PCa dataset consist of four parts:

(I) YR1973_2015.SEER9 contains the SEER November 2017 Research Data files from nine SEER registries (Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-

Oakland, Seattle-Puget Sound, and Utah) for 1973-2015 (n=637005).

(II) YR1992_2015.SJ_LA_RG_AK contains the SEER November 2017 Research Data files from the San Jose-Monterey, Los Angeles, Rural Georgia and Alaska Natives SEER registries for 1992-2015 (n=164576).

(III) YR2000_2015.CA_KY_LO_NJ_GA contains the SEER November 2017 Research Data files from the Greater California, Kentucky, Louisiana, New Jersey, and Greater Georgia SEER registries for 2000-2015 (n=461,552).

(IV) YR2005.LO_2ND_HALF contains the July—December 2005 diagnoses for Louisiana from their November 2017 SEER submission (n=1,352).

The National Program of Cancer Registries (NPCR) offers access to two public use databases at https://www.cdc.gov/cancer/npcr/public-use/ in collaboration with SEER. The databases include data by demographic characteristics (for example, age, sex, race, and year of diagnosis) and tumor characteristics (for example, site, histology, stage, and behavior). Hospitals, physicians, and laboratories across the nation report these data to central cancer registries supported by CDC and NCI. The two databases are:

(I) 2001–2015 database (17), which includes data for 50 states and the District of Columbia (n=3,086,534 for PCa).

(II) 2005–2015 database (18), which includes data for 50 states, the District of Columbia, and Puerto Rico (n=2,294,444 for PCa).

The popular statistical package R contains a PCa dataset from Stamey *et al.* [1989] (19), available for analysis when using the ElemStatLearn package. It contains data from 97 patients for 9 clinical variables. More information can be found at https://cran.r-project.org/web/packages/ElemStatLearn/ElemStatLearn.pdf.

Genomics data

The popular tool cBioPortal (10), a web portal for cancer genomics data, offers access to sixteen PCa datasets (including clinical and biospecimen data in some cases). cBioPortal has several built-in visualizations and analyses of the genomics data, which make it very easy to explore the data without much effort. The datasets, available at http://www.cbioportal.org/datasets, are:

(I) Genomic Hallmarks of Prostate Adenocarcinoma (CPC-GENE) (20). Comprehensive genomic profiling of 477 Prostate Adenocarcinoma samples from CPC-GENE and public data sets, including TCGA-PRAD. Data available at http://www.cbioportal.org/study?id=prad_cpcg_2017.

(II) MSK-IMPACT Clinical Sequencing Cohort (MSKCC): PCa (21). Targeted sequencing of clinical cases via MSK-IMPACT for PCa. Data available at http://www.cbioportal.org/study?id=prad_mskcc_2017.

(III) Metastatic Prostate Adenocarcinoma (MCTP) (22). Comprehensive profiling of 61 PCa samples, including 50 metastatic CRPCs and 11 high-grade localized PCa. Generated by Arul Chinnaiyan's and Scott Tomlins' labs at the University of Michigan. Data available at http://www.cbioportal.org/study?id=prad_mich.

(IV) Metastatic Prostate Cancer, SU2C/PCF Dream Team (23). Comprehensive analysis of 150 metastatic PCa samples by the SU2C/PCF Dream Team. Data available at http://www.cbioportal.org/study?id=prad_su2c_2015.

(V) Neuroendocrine Prostate Cancer (Trento/Cornell/Broad) (24). Whole exome and RNA Seq data of castration resistant adenocarcinoma and castration resistant neuroendocrine PCa (somatic mutations and copy number aberrations, 114 samples). Data available at http://www.cbioportal.org/study?id=nepc_wcm_2016.

(VI) Prostate Adenocarcinoma (Broad/Cornell 2013) (25). Comprehensive profiling of 57 PCa samples. Generated by Levi Garraway's lab at the Broad Institute and Mark Rubin's lab at Cornell. Data available at http://www.cbioportal.org/study?id=prad_broad_2013.

(VII) Prostate Adenocarcinoma (Broad/Cornell 2012) (26). Comprehensive profiling of 112 PCa samples. Generated by Levi Garraway's lab at the Broad Institute and Mark Rubin's lab at Cornell. Data available at http://www.cbioportal.org/study?id=prad_broad.

(VIII) Prostate Adenocarcinoma (Sun Lab) (27). Whole-genome and Transcriptome Sequencing of 65 Prostate Adenocarcinoma Patients. Generated by the Sun Lab 2017. Data available at http://www.cbioportal.org/study?id=prad_eururol_2017.

(IX) Prostate Adenocarcinoma (Fred Hutchinson CRC) (28). Comprehensive profiling of PCa samples. Generated by Peter Nelson's lab at the Fred Hutchinson Cancer Research Center. Data available at http://www.cbioportal.org/study?id=prad_fhcrc.

(X) Prostate Adenocarcinoma (MSKCC) (29). MSKCC Prostate Oncogenome Project. 181 primary, 37 metastatic PCa samples, 12 PCa cell lines and xenografts. Data available at http://www.cbioportal.org/study?id=prad_mskcc.

(XI) Prostate Adenocarcinoma (MSKCC/DFCI) (30). Whole Exome Sequencing of 1013 PCa samples. Data available at http://www.cbioportal.org/study?id=prad_p1000.

(XII) Prostate Adenocarcinoma (TCGA) (31). Integrated profiling of 333 primary prostate adenocarcinoma samples. Data available at http://www.cbioportal.org/study?id=prad_tcga_pub.

(XIII) Prostate Adenocarcinoma (TCGA, PanCancer Atlas) (32). Comprehensive TCGA PanCanAtlas data from 11k cases and all TCGA tumor types (33). Data available at http://www.cbioportal.org/study?id=prad_tcga_pan_can_atlas_2018.

(XIV) Prostate Adenocarcinoma (TCGA, Provisional). TCGA Prostate Adenocarcinoma (499 samples). Data available at http://www.cbioportal.org/study?id=prad_tcga.

(XV) Prostate Adenocarcinoma CNA study (MSKCC) (33). Copy-number profiling of 103 primary PCa samples from MSKCC. Data available at http://www.cbioportal.org/study?id=prad_mskcc_2014.

(XVI) Prostate Adenocarcinoma Organoids (MSKCC) (34). Exome profiling of PCa samples and matched organoids (12 samples). Data available at http://www.cbioportal.org/study?id=prad_mskcc_cheny1_organoids_2014.

A subsite of cBioPortal, http://www.cbioportal.org/genie, contains data from the Genomics Evidence Neoplasia Information Exchange (GENIE) project (35) of the American Association for Cancer Research (AACR). The GENIE project seeks to identify and validate genomic biomarkers relevant to cancer treatment by linking tumor genomic data from clinical sequencing efforts with longitudinal clinical outcomes. The project includes data from eleven cancer centers from the USA (7×), Canada, the Netherlands, France and the United Kingdom. GENIE version 5.0 contains data from 2,214 PCa samples (from 2,008 patients): 2,172× Prostate Adenocarcinoma, 28× Prostate Neuroendocrine Carcinoma, 13× Prostate Small Cell Carcinoma and 1× Prostate Squamous Cell Carcinoma. The data is also accessible through https://www.synapse.org/genie.

The International Cancer Genome Consortium (ICGC) Data Portal (36) currently contains six PCa datasets, which can be found at https://dcc.icgc.org/q?q=prostate&type=project:

(I) PRAD-CA (37) (125 subjects). Prostate Adenocarcinoma—Canada. Collected by the CPC-GENE network and connected to the 1st dataset in cBioportal mentioned above. Data available at https://dcc.icgc.org/projects/PRAD-CA.

(II) PRAD-FR (38) (25 subjects). Prostate Adenocarcinoma—France. Collected by ten French and one Spanish research organization. Data available at https://dcc.icgc.org/projects/PRAD-FR.

(III) EOPC-DE (39) (211 subjects), Early Onset Prostate Cancer—Germany. Collected by six German research organizations. Data available at https://dcc.icgc.org/projects/EOPC-DE.

(IV) PRAD-UK (40) (216 subjects). Prostate Adenocarcinoma—United Kingdom. Collected by the international Cancer Research UK funded Prostate Cancer Network (CR-UKPCN). Data available at https://dcc.icgc.org/projects/PRAD-UK.

(V) PRAD-US (41) (500 subjects). Prostate Adenocarcinoma TCGA—United States. Collected by sixteen American and one Canadian research organization. Data available at https://dcc.icgc.org/projects/PRAD-US.

(VI) PRAD-CN (27,42) (65 subjects). Prostate Cancer—China. Collected by the Sun Lab. The same as the 8th dataset in cBioportal mentioned above. Data available at https://dcc.icgc.org/projects/PRAD-CN.

The Genomics Data Commons (GDC) (43) gives access to The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) dataset, which is the same as the 14th dataset in cBioPortal mentioned above, and the 5th dataset in the ICGC Data Portal. It can be accessed at https://portal.gdc.cancer.gov/projects/TCGA-PRAD.

The Gene Expression Omnibus (GEO) database (11) from the National Center for Biotechnology Information (NCBI) contains 51 curated human PCa datasets, which

can be retrieved at https://www.ncbi.nlm.nih.gov/gds/?term=%E2%80%9Cprostate+cancer%E2%80%9D%5BTitle%5D+AND+%22Homo+sapiens%22%5Bporgn%3A__txid9606%5D+AND+gds%5BFilter%5D (*Table S1*). These are only the curated datasets; there are in total 834 human PCa series available in GEO.

The ArrayExpress database (12) from the European Bioinformatics Institute (EBI) contains 126 human PCa datasets. We used the "ArrayExpress data only" checkbox to avoid datasets that are in GEO as well. The list of datasets can be retrieved at https://www.ebi.ac.uk/arrayexpress/browse.html?keywords=prostate+cancer&organism=Homo+sapiens&directsub=on (*Table S2*).

Imaging data

The Cancer Imaging Archive (TCIA) (44) has nine PCa datasets available, which can be found at http://www.cancerimagingarchive.net/:

(I) The Prostate-MRI collection (26 subjects) (45) of prostate Magnetic Resonance Images (MRIs) was obtained with an endorectal and phased array surface coil at 3T (Philips Achieva). Each patient had biopsy confirmation of cancer and underwent a robotic-assisted radical prostatectomy. A mold was generated from each MRI, and the prostatectomy specimen was first placed in the mold, then cut in the same plane as the MRI. The data was generated at the National Cancer Institute, Bethesda, Maryland, USA between 2008 and 2010, and can be downloaded from https://wiki.cancerimagingarchive.net/display/Public/PROSTATE-MRI (limited access).

(II) In the Prostate-Diagnosis project (92 subjects) (46), PCa T1- and T2-weighted magnetic resonance images (MRIs) were acquired on a 1.5 T Philips Achieva by combined surface and endorectal coil, including dynamic contrast-enhanced images obtained prior to, during and after I.V. administration of 0.1 mmol/kg body weight of Gadolinium-DTPA (pentetic acid). Data is available at https://wiki.cancerimagingarchive.net/display/Public/PROSTATE-DIAGNOSIS.

(III) NaF Prostate (9 subjects) (47,48) is a collection of F-18 NaF positron emission tomography/computed tomography (PET/CT) images in patients with PCa, with suspected or known bone involvement. This dataset is available for download at https://wiki.cancerimagingarchive.net/display/Public/NaF+Prostate.

(IV) The Prostate-3T project (64 subjects) (49) provided imaging data to TCIA as part of an ISBI challenge competition in 2013. Prostate transversal T2-weighted magnetic resonance images (MRIs) acquired on a 3.0T Siemens TrioTim using only a pelvic phased-array coil were acquired for PCa detection. Data can be downloaded from https://wiki.cancerimagingarchive.net/display/Public/Prostate-3T.

(V) The QIN PROSTATE collection (22 subjects) (50,51) of the Quantitative Imaging Network (QIN) contains multiparametric MRI images collected for the purposes of detection and/or staging of PCa . The MRI parameters include T1- and T2-weighted sequences as well as Diffusion Weighted and Dynamic Contrast-Enhanced MRI. The images were obtained using endorectal and phased array surface coils at 3.0T (GE Signa HDx 15.0) The value of this collection is to provide clinical image data for the development and evaluation of quantitative methods for PCa characterization using multiparametric MRI. Data can be accessed, after a request, through https://wiki.cancerimagingarchive.net/display/Public/QIN+PROSTATE (limited access).

(VI) The TCGA-PRAD project (14 subjects) (52), also mentioned in the Genomics section of this review, also has imaging data (CT, PT, MR and pathology images) available, which can be accessed through https://wiki.cancerimagingarchive.net/display/Public/TCGA-PRAD. It also contains a link to the clinical data belonging to this study.

(VII) The Prostate Fused-MRI-Pathology collection (28 subjects) (53) is a combination of MRI images and histopathology slides. It comprises a set of 3 Tesla T1-weighted, T2-weighted, Diffusion weighted and Dynamic Contrast Enhanced prostate MRI along with accompanying digitized histopathology (H&E stained) images of corresponding radical prostatectomy specimens. The MRI scans also have a mapping of extent of PCa on them. The dataset is accessible at https://wiki.cancerimagingarchive.net/display/Public/Prostate+Fused-MRI-Pathology.

(VIII) The PROSTATEx Challenge dataset (346 subjects) (54,55) is a retrospective set of prostate MR studies. All studies included T2-weighted (T2W), proton density-weighted (PD-W), dynamic contrast enhanced (DCE), and diffusion-weighted (DW) imaging. Data can be downloaded at https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM-NCI+PROSTATEx+Challenges.

(IX) The QIN-PROSTATE-Repeatability dataset (15 subjects) (56-58) is a dataset with multiparametric prostate MRI applied in a test-retest setting, allowing to evaluate repeatability of the MRI-based measurements in the prostate. The imaging data is accompanied by two types of derived data: (I) manual segmentations of the total prostate gland, peripheral zone of the prostate gland, suspected tumor and normal regions (where applicable) and (II) volume measurements (for axial T2w images and ADC images) and mean ADC (for ADC images) corresponding to the segmented regions. Data can be accessed, after a request, through https://wiki.cancerimagingarchive.net/display/Public/QIN-PROSTATE-Repeatability.

Overall

The above sections show all clinical datasets, genomics datasets and imaging datasets. The most valuable datasets however are those that consist of a combination of these three domains, because it enables researchers to study connections, determine correlations, etc. *Table 1* shows a combined overview of the clinical, genomics and imaging datasets. There is only one dataset that has data from all three domains: the TCGA dataset (52) [also known as PRAD-US (41)]. Furthermore, there are 20 clinical + genomics datasets and 1 clinical + imaging dataset. The full list of URLs from which each dataset can be downloaded, has been submitted to the Awesome Public Datasets list at https://github.com/awesomedata/awesome-public-datasets#prostatecancer.

Discussion

Despite all the attention that has been given to making this overview of publicly available databases as extensive as possible, it is very likely not complete, and will also be outdated soon. However, this review might help many PCa researchers to find suitable datasets to answer the research question with, without the need to start a new data collection project. In the coming era of big data analysis and precision medicine (4), overviews like this are becoming more and more useful, and even necessary because of stricter privacy regulations (2). In the shift to data-driven research, the focus should be on data quality, as researchers depend more and more on the data not only for analysis, but also to generate hypotheses. The large amounts of data make it more difficult to do manual quality control, increasing the need for data quality control software. The datasets discussed within this overview seem to be of high quality, although it should be noted that some non-PCa-specific datasets such as the SEER and NPCR database, needed quite a lot of decoding work (i.e., translating codes to their PCa-specific description), increasing the risk of human errors. The SEER database, which started in 1973, also has some legacy issues (e.g., containing different versions of cancer staging scores). It should be noted as well that most datasets do not adhere to the FAIR (Findability, Accessibility, Interoperability, Reusability) guiding principles for scientific data management and stewardship (59), but this could be expected since almost all datasets were generated before these principles were published. Hopefully they will be FAIRified in the near future. Some datasets in this overview contain only a small number of patients, such as the NaF Prostate study and the Prostate Adenocarcinoma Organoids (MSKCC) study. In these cases it might be useful to combine datasets, to get to a higher sample size (and statistical power) by manual or automated data model mapping (60). It might also be useful for scientists to have access to the original biomaterial from which the data was derived. Therefore, the Prostate Cancer Biorepository Network (61) is an interesting initiative: its goal is to develop a biorepository with high quality, well-annotated specimens obtained in a systematic, reproducible fashion using optimized and standardized protocols. It is a collaboration between six U.S. academic institutes and the U.S. department of defense. Finally, the success of big data analysis does not only depend on access to data and/or biospecimens, but also on the collaboration between field experts (urologists, but also imaging and genomics experts) and IT experts (62). There are very little people that have an in-depth knowledge about the disease area, the used techniques, data integration and data analysis, which is why multi-disciplinary research teams are a must in this 'big data' age.

Table 1 A combined overview of the clinical, genomics and imaging datasets, ordered by number of patients included

Data source	Dataset name	Clinical	Genomics	Imaging	No. of patients
NPCR/SEER	2001–2015 Database (PCa)	31 clinical parameters, such as age, race, grade, diagnostic confirmation and laterality	–	–	3,086,534
NPCR/SEER	2005–2015 Database (PCa)	25 clinical parameters, such as age, race, grade, diagnostic confirmation and laterality	–	–	2,294,444
SEER	YR1973_2015.SEER9 (PCa)	133 clinical parameters, such as age, race, Gleason scores, TNM stages, PSA values, survival data and therapy data	–	–	637,005
SEER	YR2000_2015.CA_KY_LO_NJ_GA (PCa)	133 clinical parameters, such as age, race, Gleason scores, TNM stages, PSA values, survival data and therapy data	–	–	461,552
SEER	YR1992_2015.SJ_LA_RG_AK (PCa)	133 clinical parameters, such as age, race, Gleason scores, TNM stages, PSA values, survival data and therapy data	–	–	164,576
PLCO	Prostate	Data for PCa screening, incidence, and mortality analyses	–	–	76,682
PLCO	Prostate Screening	Additional information from PSA and DRE cancer screens	–	–	35,875
PLCO	Prostate Diagnostic Procedures	Information about the diagnostic procedures prompted by positive PCa screens	–	–	15,307
PLCO	Prostate Treatments	Specifics of the initial treatment following the diagnosis of PCa	–	–	7,614
PLCO	Prostate Screening Abnormalities	Information for each induration found during the DRE screen	–	–	5,743
PLCO	Prostate Medical Complications	Information about the medical complications caused by diagnostic workup for PCa	–	–	2,164
cBioPortal/ Synapse	GENIE	13 clinical parameters, such as age, race and ethnicity	Mutation data	–	2,008
SEER	YR2005.LO_2ND_HALF (PCa)	133 clinical parameters, such as age, race, Gleason scores, TNM stages, PSA values, survival data and therapy data	–	–	1,352
cBioPortal	Prostate Adenocarcinoma (MSKCC/DFCI)	19 clinical parameters, such as cancer type, diagnosis age and Gleason scores	Mutation data and copy number alteration data	–	1,013
cBioPortal/ ICGC/GDC/ TCIA	Prostate Adenocarcinoma (TCGA, Provisional), aka PRAD-US	100 clinical parameters, such as Gleason scores, TNM values, survival data, age, weight, ethnicity, PSA values and MRI results	Mutation data and copy number alteration data	16,790 CT, PT, MR images in 207 series from 14 patients. 3.74 GB of data. Tissue slide images included	498
cBioPortal	Prostate adenocarcinoma (TCGA, PanCancer Atlas)	83 clinical parameters, such as diagnosis age, cancer type, ethnicity category, patient weight and race category	Mutation data and copy number alteration data	–	494

Table 1 *(continued)*

Table 1 *(continued)*

Data source	Dataset name	Clinical	Genomics	Imaging	No. of patients
cBioPortal	Genomic Hallmarks of Prostate Adenocarcinoma (CPC-GENE)	89 clinical parameters, such as Gleason scores, PSA values, weight, survival data, TNM stages and MRI results	Comprehensive genomic profiling of 477 Prostate Adenocarcinoma samples from CPC-GENE and public data sets, including TCGA-PRAD	–	477
cBioPortal	MSK-IMPACT Clinical Sequencing Cohort (MSKCC): Prostate Cancer	17 clinical parameters, such as clinical Gleason, age and mutation data	Targeted sequencing of clinical cases via MSK-IMPACT for PCa	–	451
TCIA	PROSTATEx Challenge	–	–	309,251 MR (T2W, PD-W, DCE and DW) images, 15.1 GB of data	346
cBioPortal	Prostate Adenocarcinoma (TCGA)	89 clinical parameters, such as clinical and reviewed Gleason scores, age and gene mutation data	Integrated profiling of 333 primary prostate adenocarcinoma samples	–	333
cBioPortal	Prostate Adenocarcinoma (MSKCC)	25 clinical parameters, such as radical prostatectomy Gleason scores, survival data, tumor stages and ERG Fusion data	181 primary, 37 metastatic PCa samples, 12 PCa cell lines and xenografts	–	216
ICGC	PRAD-UK: Prostate Adenocarcinoma - United Kingdom	6 files with clinical data: donor, donor exposure, donor family, donor therapy, sample and specimen	Simple Somatic Mutations (SSM) for 215 patients. Copy Number Somatic Mutations (CNSM) for 13 patients. Structural Somatic Mutations (StSM) for 13 patients	–	216
ICGC	EOPC-DE: Early Onset Prostate Cancer - Germany	6 files with clinical data: donor, donor exposure, donor family, donor therapy, sample and specimen	Simple Somatic Mutations (SSM) for 202 patients. Copy Number Somatic Mutations (CNSM) for 11 patients. Structural Somatic Mutations (StSM) for 11 patients	–	211
cBioPortal	Metastatic Prostate Cancer, SU2C/PCF Dream Team	20 clinical parameters, such as age and prior medications	Comprehensive analysis of 150 metastatic PCa samples	–	150
ICGC	PRAD-CA: Prostate Adenocarcinoma - Canada	6 files with clinical data: donor, donor exposure, donor family, donor therapy, sample and specimen	SSM data for 124 patients. CNSM data for 125 patients. StSM data for 123 patients. SGV data for 123 patients. METH-A data for 102 patients	–	125

Table 1 *(continued)*

Table 1 *(continued)*

Data source	Dataset name	Clinical	Genomics	Imaging	No. of patients
cBioPortal	Prostate Adenocarcinoma (Broad/ Cornell 2012)	15 clinical parameters, such as Gleason score 4–5%, age, PSA values, radical prostatectomy Gleason scores and modified Capra S Scores	Comprehensive profiling of 112 PCa samples	–	112
cBioPortal	Prostate Adenocarcinoma CNA study (MSKCC)	37 clinical parameters, such as biopsy and pathology Gleason scores, survival data, PSA values, age, extracapsular extension and treatment data	Copy-number profiling of 103 primary PCa samples from MSKCC	–	104
R ElemStatLearn package	Prostate (R)	9 clinical parameters: cancer volume, prostate weight, age, amount of benign prostatic hyperplasia, seminal vesicle invasion, capsular penetration, Gleason scores, percent of Gleason score 4 or 5 and PSA values	–	–	97
TCIA	Prostate-Diagnosis	4 clinical text fields: path report biopsy, path prostate specimen, MRI report, treatment	–	32,537 MR images (T1, T2, and DCE sequences) in 368 series, 5.6 GB of data. 3D segmentation files included	92
cBioPortal	Neuroendocrine Prostate Cancer (Trento/Cornell/ Broad)	16 clinical parameters, such as genomic burden, pathology classification and ploidy	Whole exome and RNA Seq data of castration resistant adenocarcinoma and castration resistant neuroendocrine PCa (somatic mutations and copy number aberrations)	–	81
cBioPortal/ ICGC	Prostate Adenocarcinoma (Sun Lab), aka PRAD-CN	20 clinical parameters, such as cancer type, diagnosis age, PSA values, Gleason scores and TNM stage	Mutation data and copy number alteration data	–	65
TCIA	Prostate-3T	–	–	1,258 MR (T2W) images in 64 series, 284 MB of data. Files with segmentation data included	64
cBioPortal	Prostate Adenocarcinoma (Fred Hutchinson CRC)	26 clinical parameters, such as chemotherapy data, EXOME data, number of tumors and PSA values	Comprehensive profiling of 176 PCa samples	–	63
cBioPortal	Metastatic Prostate Adenocarcinoma (MCTP)	26 clinical parameters, such as therapy info, PSA values, Gleason scores and survival data	Comprehensive profiling of 50 metastatic CRPCs and 11 high-grade localized PCa	–	59

Table 1 *(continued)*

Table 1 *(continued)*

Data source	Dataset name	Clinical	Genomics	Imaging	No. of patients
cBioPortal	Prostate Adenocarcinoma (Broad/ Cornell 2013)	20 clinical parameters, such as Gleason score 4–5%, age, PSA values, radical prostatectomy Gleason scores and tumor stages	Comprehensive profiling of 57 PCa samples	–	57
TCIA	Prostate Fused-MRI-Pathology	–	–	32,508 MR images in 325 series, 4.4 GB of data. Annotated whole slide pathology images and fused Rad-Path Matlab files included	28
TCIA	Prostate-MRI	–	–	22,036 MR (with some PET/CT) images in 182 series, 3.2 GB of data. Pathology images included	26
ICGC	PRAD-FR: Prostate Adenocarcinoma-France	6 files with clinical data: donor, donor family, donor surgery, sample and specimen	SSM data, CNSM data, StSM data, SGV data	–	25
TCIA	QIN PROSTATE	–	–	25,981 MR images in 319 series, 4.4 GB of data	22
TCIA	QIN-PROSTATE-Repeatability	–	–	2,504 MR images in 270 series, 1.1 GB of data. Manual segmentations and volume measurements included	15
TCIA	NaF Prostate	–	–	64,535 PET/CT images, 12.9 GB of data. DICOM metadata digest included	9
cBioPortal	Prostate Adenocarcinoma Organoids (MSKCC)	18 clinical parameters, such as PSA values, HGB values, ALP values, LDH values and therapy info	Exome profiling of PCa samples and matched organoids	–	7
GEO	51 datasets, see *Table S1*	–	see *Table S1*	–	see *Table S1*
ArrayExpress	126 datasets, see *Table S2*	–	see *Table S2*	–	see *Table S2*

Acknowledgements

The author thanks Chris Bangma for presenting a poster version of this manuscript at the AUA 2018 (8). The NPCR/SEER data were provided by central cancer registries participating in CDC's National Program of Cancer Registries (NPCR) and/or NCI's Surveillance, Epidemiology, and End Results (SEER) Program and submitted to CDC and NCI in November, 2017. The author thanks the National Cancer Institute for access to NCI's data collected by the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. The statements contained herein are solely those of the author and do not represent or imply concurrence or endorsement by NCI. The author would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of study author. The results on the TCGA-PRAD dataset shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. The author would like to acknowledge the U01 CA151261 award that supported collection and sharing of the QIN PROSTATE and QIN-PROSTATE-Repeatability datasets.

Footnote

Conflicts of Interest: Dr. Hulsen is employed by Philips Research. This manuscript assumes that the datasets listed here were collected in a GDPR compliant manner.

References

1. American Cancer Society. Key Statistics for Prostate Cancer. Available online: https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html
2. Simell BA, Tornwall OM, Hamalainen I, et al. Transnational access to large prospective cohorts in Europe: Current trends and unmet needs. N Biotechnol 2019;49:98-103.
3. New PhD researchers will crunch big data to help fight against prostate cancer. Available online: https://prostatecanceruk.org/about-us/news-and-views/2016/11/new-phd-researchers-will-crunch-big-data-to-help-fight-against-prostate-cancer
4. Hulsen T, Jamuar SS, Moody AR, et al. From Big Data to Precision Medicine. Front Med (Lausanne) 2019;6:34.
5. Hulsen T, Obbink JH, Van der Linden W, et al. 958 Integrating large datasets for the Movember Global Action Plan on active surveillance for low risk prostate cancer. Eur Urol Suppl 2016;15:e958.
6. Hulsen T, Van der Linden W, De Jonge C, et al. PT-073 Developing a future-proof database for the European Randomized study of Screening for Prostate Cancer (ERSPC). Eur Urol Suppl 2019;18;e1766.
7. Hulsen T, Obbink H, Schenk E, et al. PCMM Biobank, IT-infrastructure and decision support. CTMM meeting 2013. Available online: http://tim.hulsen.net/documents/pcmm_wp3_130912.pdf
8. Hulsen T, Bangma CH. MP70-02 An Overview of Publicly Available Patient-centered Prostate Cancer Datasets. J Urol 2018;199:e934.
9. Röthke M, Blondin D, Schlemmer HP, et al. PI-RADS classification: structured reporting for MRI of the prostate. Rofo 2013;185:253-61.
10. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-10.
12. Kolesnikov N, Hastings E, Keays M, et al. ArrayExpress update--simplifying data submissions. Nucleic Acids Res 2015;43:D1113-6.
13. Dunning MJ, Vowler SL, Lalonde E, et al. Mining Human Prostate Cancer Datasets: The "camcAPP" Shiny App. EBioMedicine 2017;17:5-6.
14. Gandaglia G, Bray F, Cooperberg MR, et al. Prostate Cancer Registries: Current Status and Future Directions. Eur Urol 2016;69:998-1012.
15. Gohagan JK, Prorok PC, Kramer BS, et al. Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute. J Urol 1994;152:1905-9.
16. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.
17. 2001–2015 Database: National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence – USCS 2001–2015 Public Use Research Database,

United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2018, based on the November 2017 submission. Available online: www.cdc.gov/cancer/uscs/public-use
18. 2005–2015 Database: National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database: NPCR and SEER Incidence – USCS 2005–2015 Public Use Research Database, United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2018, based on the November 2017 submission. Available online: www.cdc.gov/cancer/uscs/public-use
19. Stamey TA, Kabalin JN, McNeal JE, et al. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. J Urol 1989;141:1076-83.
20. Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature 2017;541:359-64.
21. Cheng DT, Mitchell TN, Zehir A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 2015;17:251-64.
22. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012;487:239-43.
23. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. Cell 2015;161:1215-28.
24. Beltran H, Prandi D, Mosquera JM, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med 2016;22:298-305.
25. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. Cell 2013;153:666-77.
26. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685-9.
27. Ren S, Wei GH, Liu D, et al. Whole-genome and Transcriptome Sequencing of Prostate Cancer Identify New Genetic Alterations Driving Disease Progression. Eur Urol 2017. [Epub ahead of print].
28. Kumar A, Coleman I, Morrissey C, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med 2016;22:369-78.
29. Taylor BS, Schultz N, Hieronymus H, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010;18:11-22.
30. Armenia J, Wankowicz SAM, Liu D, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet 2018;50:645-51.
31. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell 2015;163:1011-25.
32. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 2018;173:291-304.e6.
33. Hieronymus H, Schultz N, Gopalan A, et al. Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci U S A 2014;111:11139-44.
34. Gao D, Vela I, Sboner A, et al. Organoid cultures derived from patients with advanced prostate cancer. Cell 2014;159:176-87.
35. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov 2017;7:818-31.
36. Zhang J, Baran J, Cros A, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database (Oxford) 2011;2011:bar026.
37. Bristow R, Boutros P, Hudson T, et al. Prostate Adenocarcinoma - Canada. Available online: https://icgc.org/icgc/cgp/70/392/70542
38. Cussenot O. Prostate Adenocarcinoma - France. Available online: https://icgc.org/icgc/cgp/70/355/1002116
39. Sültmann H, Sauter G. Early Onset Prostate Cancer - Germany. Available online: https://icgc.org/icgc/cgp/70/345/53039
40. Cooper C, Eeles R, Stratton M, et al. Prostate Adenocarcinoma - United Kingdom. Available online: https://icgc.org/icgc/cgp/70/508/71331
41. Consortium T. Prostate Adenocarcinoma TCGA - United States. Available online: https://icgc.org/icgc/cgp/70/509/70272
42. Sun Y. Prostate Cancer - China. Available online: https://icgc.org/icgc/cgp/70/371/1003238
43. Grossman RL, Heath AP, Ferretti V, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med 2016;375:1109-12.
44. Clark K, Vendt B, Smith K, et al. The Cancer Imaging

Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26:1045-57.
45. Choyke P, Turkbey B, Pinto P, et al. (2016). Data From PROSTATE-MRI. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2016.6046GUDv
46. Bloch BN, Jain A, Jaffe CC (2015). Data From PROSTATE-DIAGNOSIS. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2015.FOQEUJVT
47. Kurdziel KA, Shih JH, Apolo AB, et al. The kinetics and reproducibility of 18F-sodium fluoride for oncology using current PET camera technology. J Nucl Med 2012;53:1175-84.
48. Kurdziel KA, Shih JH, Apolo AB, et al. (2015). Data From NaF_PROSTATE. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2015.ISOQTHKO
49. Litjens G, Futterer J, Huisman H (2015). Data From Prostate-3T. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2015.QJTV5IL5
50. Fedorov A, Fluckiger J, Ayers GD, et al. A comparison of two methods for estimating DCE-MRI parameters via individual and cohort based AIFs in prostate cancer: a step towards practical implementation. Magn Reson Imaging 2014;32:321-9.
51. Fedorov A, Tempany C, Mulkern R, et al. (2016). Data From QIN PROSTATE. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2016.fADs26kG
52. Zuley ML, Jarosz R, Drake BF, et al. (2016). Radiology Data from The Cancer Genome Atlas Prostate Adenocarcinoma [TCGA-PRAD] collection. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2016.YXOGLM4Y
53. Madabhushi A, Feldman M (2016). Fused Radiology-Pathology Prostate Dataset. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2016.TLPMR1AM
54. Litjens G, Debats O, Barentsz J, et al. Computer-aided detection of prostate cancer in MRI. IEEE Trans Med Imaging 2014;33:1083-92.
55. Litjens G, Debats O, Barentsz J, et al. (2017). ProstateX Challenge data. The Cancer Imaging Archive. Available online: https://doi.org/10.7937/K9TCIA.2017.MURS5CL
56. Fedorov A, Schwier M, Clunie D, et al. An annotated test-retest collection of prostate multiparametric MRI. Sci Data 2018;5:180281.
57. Fedorov A, Schwier M, Clunie D, et al. (2018). Data From QIN-PROSTATE-Repeatability. The Cancer Imaging Archive. Available online: http://doi.org/10.7937/K9/TCIA.2018.MR1CKGND
58. Fedorov A, Vangel MG, Tempany CM, et al. Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. Invest Radiol 2017;52:538-46.
59. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
60. Hulsen T, Van der Linden W, Pletea D, et al. Data Model Mapping. (2017). Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2017167628
61. Darshan M, Zheng Q, Fedor HL, et al. Biobanking of derivatives from radical retropubic and robot-assisted laparoscopic prostatectomy tissues as part of the prostate cancer biorepository network. Prostate 2014;74:61-9.
62. Bangma C, Obbink, H. The future of prostate cancer research: bringing data together, looking back and forward. Transl Androl Urol 2018;7:188-94.

Cite this article as: Hulsen T. An overview of publicly available patient-centered prostate cancer datasets. Transl Androl Urol 2019;8(Suppl 1):S64-S77. doi: 10.21037/tau.2019.03.01

Supplementary

Table S1 An overview of the prostate cancer datasets in GEO, ordered by number of samples

DataSet	Title	Type	Platform	Series	No. of samples
GDS2545	Metastatic prostate cancer (HG-U95A)	Expression profiling by array, count, 4 tissue sets	GPL8300	GSE6919	171
GDS2546	Metastatic prostate cancer (HG-U95B)	Expression profiling by array, count, 4 tissue sets	GPL92	GSE6919	167
GDS2547	Metastatic prostate cancer (HG-U95C)	Expression profiling by array, count, 4 tissue sets	GPL93	GSE6919	164
GDS3289	Prostate cancer progression at the cellular level	Expression profiling by array, log2 ratio, 2 cell type, 6 disease state, 14 other sets	GPL2013	GSE6099	104
GDS4395	External beam radiation therapy effect on prostate cancer patients: peripheral white blood cells	Expression profiling by array, count, 20 individual, 2 protocol, 8 time sets	GPL570	GSE30174	80
GDS4109	Recurrent and non-recurrent prostate cancer primary tumors	Expression profiling by array, count, 2 disease state sets	GPL96	GSE25136	79
GDS2384	Xenograft model of prostate carcinoma progression	Expression profiling by array, log2 ratio, 4 disease state, 3 other, 10 protocol, 14 specimen sets	GPL3349	GSE4084	52
GDS5267	Cyclin-dependent kinase inhibitor R547 effect on prostate cancer cell line: dose response and time course	Expression profiling by array, count, 3 agent, 4 dose, 4 time sets	GPL570	GSE15392	45
GDS1746	Primary epithelial cell cultures from prostate tumors	Expression profiling by array, count, 7 disease state, 2 protocol sets	GPL96	GSE3868	30
GDS4952	BET bromodomain inhibitor I-BET762 effect on prostate cancer cell lines: dose response	Expression profiling by array, count, 4 cell line, 3 dose sets	GPL570	GSE56352	24
GDS4114	Reactive stroma of breast and prostate cancer	Expression profiling by array, transformed count, 2 disease state, 2 tissue sets	GPL570	GSE26910	24
GDS4824	Prostate cancer	Expression profiling by array, count, 2 disease state, 3 genotype/variation sets	GPL570	GSE55945	21
GDS1390	Prostate cancer progression after androgen ablation	Expression profiling by array, count, 2 disease state sets	GPL96	GSE2443	20
GDS1439	Prostate cancer progression	Expression profiling by array, count, 3 disease state sets	GPL570	GSE3325	19
GDS4964	Telomere-elongated, prostate cancer cells	Expression profiling by array, transformed count, 2 genotype/variation, 2 protocol sets	GPL570	GSE41559	16
GDS4158	LNCap prostate cancer cell line response to loss of COnstitutive Photomorphogenic-1 and ETV1	Expression profiling by array, transformed count, 3 genotype/variation sets	GPL570	GSE27914	16
GDS4159	LNCap prostate cancer cell line response to loss of COnstitutive Photomorphogenic-1, ETV1 and c-JUN	Expression profiling by array, transformed count, 3 genotype/variation sets	GPL570	GSE27914	15
GDS3358	Androgen deprivation effect on LNCaP prostate cancer cells: time course	Expression profiling by array, count, 2 growth protocol, 6 time sets	GPL570	GSE8702	15
GDS535	Prostate cancer antiandrogen resistance	Expression profiling by array, count, 2 cell type sets	GPL91	GSE847	14
GDS6100	MicroRNA-135b overexpression effect on prostate cancer cell line: time course	Expression profiling by array, transformed count, 2 protocol, 3 time sets	GPL10558	GSE57820	12
GDS4957	FOXA1 overexpression effect on prostate cancer cell line	Expression profiling by array, transformed count, 2 protocol sets	GPL10558	GSE49153	12
GDS4951	Lysophosphatidic acid effect on breast and prostate cancer cell lines	Expression profiling by array, count, 2 agent, 3 cell line sets	GPL570	GSE56265	12
GDS4107	KUCaP-2 xenograft model of castration-resistant prostate cancer: various stages	Expression profiling by array, transformed count, 3 development stage sets	GPL570	GSE21887	12
GDS3973	Docetaxel resistant prostate cancer cell line	Expression profiling by array, transformed count, 4 cell line sets	GPL570	GSE33455	12
GDS3861	Synthetic androgen R1881 effect on transcription factor SRF-deficient prostate cancer cells	Expression profiling by array, transformed count, 2 agent, 2 protocol sets	GPL570	GSE22606	12
GDS2971	Hemiasterlin analog HTI-286 effect on docetaxel-resistant prostate cancer cell line	Expression profiling by array, log2 ratio, 2 agent sets	GPL3877	GSE8325	12
GDS5072	High grade prostate cancer	Expression profiling by array, count, 2 disease state, 3 other sets	GPL570	GSE45016	11
GDS5440	Androgen effect on carboxyl terminal-binding protein 2-deficient prostate cancer cell line	Expression profiling by array, transformed count, 2 agent, 4 genotype/variation sets	GPL6244	GSE58309	10
GDS3111	Prostate cancer cell line response to dihydrotestosterone: time course	Expression profiling by array, count, 2 agent, 3 time sets	GPL570	GSE7868	9
GDS3634	miR-205 expression effect on prostate cancer cell line	Expression profiling by array, count, 2 protocol sets	GPL6104	GSE11701	8
GDS3095	Zinc effect on malignant and non-malignant prostate cell lines: time course	Expression profiling by array, count, 2 agent, 2 cell line, 4 time sets	GPL2986	GSE5590	8
GDS2034	Prostate cancer cell line LNCaP response to synthetic androgen R1881: time course	Expression profiling by array, log2 ratio, 4 time sets	GPL3349	GSE4027	8
GDS1736	Arachidonic acid effect on prostate cancer cells	Expression profiling by array, count, 2 agent sets	GPL96	GSE3737	8
GDS1699	Androgen sensitive and insensitive prostate cancer cell lines: expression profiles	Expression profiling by array, log2 ratio, 8 cell line, 2 cell type sets	GPL3341	GSE4016	8
GDS5805	Peptidyl-prolyl cis/trans isomerase Pin1 deficiency effect on prostate cancer cells	Expression profiling by array, transformed count, 2 cell line, 3 protocol sets	GPL6244	GSE67457	6
GDS5222	U2OS osteosarcoma cell line response to strigolactone analogs ST362 and MEB55: 24 hours	Expression profiling by array, count, 3 agent sets	GPL10558	GSE54820	6
GDS5221	U2OS osteosarcoma cell line response to strigolactone analogs ST362 and MEB55: 6 hours	Expression profiling by array, count, 3 agent sets	GPL10558	GSE54820	6
GDS5173	G-protein coupled receptor kinase 3 expression effect on prostate cancer cell line	Expression profiling by array, count, 2 agent sets	GPL6883	GSE36022	6
GDS4124	Genetic reprogramming of prostate cancer-associated stromal cells	Expression profiling by array, transformed count, 5 cell type, 2 protocol sets	GPL570	GSE35373	6
GDS4121	Hepatocyte growth factor treatment of prostate cancer DU145 cell line: time course	Expression profiling by array, count, 2 agent, 3 time sets	GPL570	GSE16659	6
GDS4113	Late passage LNCaP prostate tumor cells treated with androgen receptor shRNA or androgen R1881	Expression profiling by array, count, 3 genotype/variation sets	GPL570	GSE22483	6
GDS2865	Metastatic prostate tumor model	Expression profiling by array, count, 2 disease state sets	GPL96	GSE7930	6
GDS4123	Isoflavone and 3,3'-diindolylmethane effect on C4-2B prostate cancer cells	Expression profiling by array, count, 3 agent, 4 time sets	GPL570	GSE35324	5
GDS5804	PI3K/mTOR Inhibitor NVP-BEZ235 and taxotere effects on prostate cancer xenograft tumors	Expression profiling by array, count, 4 agent sets	GPL570	GSE49232	4
GDS5606	Androgen effect on runt-related transcription factor 1-deficient prostate cancer cell line	Expression profiling by array, transformed count, 2 agent, 2 genotype/variation sets	GPL6244	GSE62454	4
GDS5373	miR-221 expression effect on prostate cancer cell line	Expression profiling by array, count, 2 protocol sets	GPL570	GSE45627	4
GDS4846	MED1 overexpression effect on prostate cancer cell line	Expression profiling by array, count, 2 protocol sets	GPL571	GSE41150	4
GDS4829	VprBP depletion effect on prostate cancer cell line	Expression profiling by array, count, 2 genotype/variation sets	GPL10558	GSE50414	4
GDS3797	beta-TrCP inhibition and androgen ablation effects on prostate cancer cell line LAPC4	Expression profiling by array, transformed count, 2 genotype/variation, 2 growth protocol sets	GPL571	GSE19141	4
GDS1697	DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine effect on prostate cancer cell lines	Expression profiling by array, log2 ratio, 4 cell line sets	GPL3295	GSE4089	4
GDS1423	Lunasin effect on prostate epithelial cells	Expression profiling by array, count, 2 agent, 2 disease state sets	GPL96	GSE2992	4

Table S2 An overview of the prostate cancer datasets in ArrayExpress, ordered by number of assays

Accession	Title	Type	No. of assays
E-MTAB-3732	A comprehensive human expression map	transcription profiling by array	27,871
E-MTAB-5214	RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project	RNA-seq of coding RNA	18,879
E-TABM-185	Transcription profiling by array of integrated human experiments involving the hgu133a platform to investigate a global map of human gene expression	transcription profiling by array	5,896
E-MTAB-62	Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines	transcription profiling by array	5,372
E-MTAB-2919	RNA-seq from 53 human tissue samples from the Genotype-Tissue Expression (GTEx) Project	RNA-seq of coding RNA, RNA-seq of non coding RNA	3,282
E-MTAB-2914	Cross-laboratory validation of the OncoScan FFPE Assay, a multiplex tool for whole genome tumour profiling	genotyping by array	972
E-MTAB-37	Transcriptomics for Cancer Cell Line Project	transcription profiling by array	950
E-MTAB-2770	RNA-seq of 934 human cancer cell lines from the Cancer Cell Line Encyclopedia	RNA-seq of coding RNA	934
E-MTAB-38	Genotyping of human cancer cell lines	genotyping by array	676
E-MTAB-2706	RNA-seq of 675 commonly used human cancer cell lines	RNA-seq of coding RNA	675
E-MTAB-3983	Sanger Genomics of Drug Sensitivity in Cancer Project (GDSC) RNA-seq cancer cell line gene expression data	RNA-seq of coding RNA	462
E-MTAB-6131	Methylation array for Multi-omics molecular profiling of primary prostate adenocarcinoma	methylation profiling by array	390
E-AFMX-5	Transcription profiling of human cell lines and tissues (GNF/Novartis)	transcription profiling by array	316
E-TABM-970	Transcription profiling by array of human normal tissues	microRNA profiling by array	274
E-TABM-969	Transcription profiling by array of human normal tissues	microRNA profiling by array	255
E-TABM-47	MicroRNA profiling of human normal lung and lung cancer samples to investigate the role of miRNA involvement in lung carcinogenesis	microRNA profiling by array	246
E-MEXP-113	Transcription profiling of multiple human tumour specimens of different anatomical origin arrayed against a common reference	transcription profiling by array	242
E-MTAB-2980	RNA-seq of 39 human cancer cell lines that are in the NCI-60 set from the Cancer Cell Line Encyclopedia	RNA-seq of coding RNA	217
E-MTAB-6411	Short Tandem Repeats - Targeted-Sequencing of human cells for Lineage tracing	genotyping by high throughput sequencing	210
E-MTAB-3397	MiRNA profiles in Lymphoblastoid Cell Lines of Finnish Prostate Cancer Families	microRNA profiling by array	193
E-TABM-184	MicroRNA profiling of human cancer samples identifies ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas	microRNA profiling by array	193
E-MTAB-1041	Transcription profiling by array of human prostate cancer samples in order to examine the changes in gene transcription underlying the aberrant citrate and choline metabolism	transcription profiling by array	168
E-TABM-145	Transcription profiling of human cell lines and tissues - Luscombe re-analysis of GNF/Novartis data		158
E-MTAB-6128	Expression array for Multi-omics molecular profiling of primary prostate adenocarcinoma	transcription profiling by array	141
E-MTAB-6126	SNP array for Multi-omics molecular profiling of primary prostate adenocarcinoma	genotyping by array	132
E-MTAB-3222	Cancers of unknown primary (CUP) are characterized by chromosomal instability (CIN) compared to metastasis of know origin	transcription profiling by array	129
E-TABM-26	Transcription profiling of human prostate tissues obtained from multiple Institutions	transcription profiling by array	114
E-SMDB-2486	Transcription profiling of 2 primary human prostate tumors, 41 normal prostate specimens and nine lymph node metastases,	transcription profiling by array	112
E-TABM-90	Transcription profiling by array of human lymphocytes from prostate carcinoma patients after X-radiation treatment	transcription profiling by array	108
E-TABM-794	Transcription profiling of human prostate cancer	transcription profiling by array	102
E-TABM-1202	Transcriptional profiling by array of primary rhabdomyosarcoma samples with different PAX3/FOXO1 fusion gene status	transcription profiling by array	101
E-TABM-1135	MicroRNA profiling by array of human cancers to identify cancers with unknown primary tissue-of-origin	microRNA profiling by array	101
E-MTAB-2523	Next-Generation Sequencing of RNA Isolated from Paired Fresh-Frozen and Formalin-Fixed Paraffin-Embedded Samples of Human Cancer and Normal Tissue	DNA-seq, RNA-seq of coding RNA	86
E-MEXP-1327	Transcription profiling of human prostate cancer cells, normal epithelial prostatic cells and stroma cells from patients in placebo, selenium, vitamin E or selenium and vitamin E treatment groups	transcription profiling by array	85
E-MEXP-1243	Transcription profiling by array of human prostate from patients with a previous diagnosis of Prostatic Intraepithelial Neoplasia and following consumption of high glucosinolate broccoli or peas to investigate interactions with the GSTM1 genotype	transcription profiling by array	81
E-TABM-948	Transcription profiling of human hypoxia-stimulated prostate tumor cell lines and primary prostate epithelial cells	transcription profiling by array	73
E-MTAB-2968	Androgen stimulation time-course of TMPRSS2-ERG fusion positive VCaP cells	transcription profiling by array	72
E-MTAB-327	MicroRNA profiling by array of NCI-60 human cancer cell-lines	microRNA profiling by array	72
E-MEXP-1029	MicroRNA profiling of the NCI-60 panel of human cancer cell lines	microRNA profiling by array	72
E-MTAB-6127	SNP array Multi-omics molecular profiling of primary prostate adenocarcinoma	genotyping by array	66
E-TABM-49	MicroRNA profiling of human normal prostate and prostate cancer samples to investigate the role of miRNA involvement in prostate carcinogenesis	microRNA profiling by array	63
E-PROT-2	Proteomic profiling of NCI60 cell lines from Cancer Cell Line Encyclopedia	proteomic profiling by mass spectrometer	60
E-TABM-65	Comparative genomic hybridization of cell lines from 9 different cancer tissue of origin types (Breast, Central Nervous System, Colon, Leukemia, Melanoma, Non-Small Cell Lung, Ovarian, Prostate, Renal) from NCI-60 panel	comparative genomic hybridization by array	60
E-MTAB-567	RNA-seq of prostate cancer and adjacent normal tissues from 14 patients	RNA-seq of coding RNA	56
E-MTAB-408	miRNA expression profiling of prostate cancer	microRNA profiling by array	54
E-MTAB-2964	Methylation profiling blood, adjacent benign and multiple discrete tumour samples from locally advanced prostate cancers	methylation profiling by array	48
E-MTAB-513	RNA-Seq of human individual tissues and mixture of 16 tissues (Illumina Body Map)	RNA-seq of coding RNA	48
E-MEXP-2906	Transcription profiling by array of human prostate cells treated with sodium selenite or 5-2-deoxycytidine	transcription profiling by array	48
E-MTAB-4519	Analysis of transcriptomes from 21 tissues, 13 melanoma samples and 7 breast cancer cell lines, enriched for transcripts from haploblocks with intronic and intergenic GWAS SNPs	RNA-seq of coding RNA	41
E-TABM-50	MicroRNA profiling of human normal stomach and gastric cancer samples to investigate the role of miRNA involvement in stomach carcinogenesis	microRNA profiling by array	41
E-MEXP-3005	metastatic signature is present in primary prostate tumor	transcription profiling by array	40
E-MEXP-2602	MicroRNA profiling by array of mouse prostate cancer cell lines treated with dihydrotestosterone and prostate xenografts in intact or castrated mice	transcription profiling by array	40
E-MEXP-2034	Transcription profiling by array of human primary prostate epithelial and stromal cells after treatment with 4-methylsulphinylbutyl and 3-methylsulphinypropyl isothiocyanates	transcription profiling by array	40
E-MTAB-3715	Context dependent regulatory patterns of the androgen receptor and androgen receptor target genes	transcription profiling by array	39
E-TABM-626	Kinase activity profiling shows osteoblast-induced EGFR/ERBB2 signaling in human androgen-sensitive prostate carcinoma cells	transcription profiling by array	39
E-MTAB-4858	Microarray analysis of Du145, PC3 and LNCaP human prostate cancer cell lines	transcription profiling by tiling array	36
E-MEXP-2966	The purpose of the experiment was to study miRNA expression in prostate cancer cell lines and xenografts and to combine it with miRNA gene copy number data that we already had to identify miRNAs that could be overexpressed or underexpressed as a consequence of amplification or deletion of the miRNA gene, respectively	transcription profiling by array	36
E-MEXP-993	Transcription profiling by array of human prostate cancer stem cells	transcription profiling by array	36
E-MEXP-3640	Transcription profiling by array of cancerous and non-cancerous human prostate cell lines treated with PY-ITC or sulforaphane in the presence and absence of the PI3K inhibitor LY294002	transcription profiling by array	35
E-MEXP-1331	Transcription profiling of normal, tumor and pure stromal tissue samples from patients with prostate adenocarcinoma, together with 4 cell lines	transcription profiling by array	35
E-MEXP-3020	Low Dose PDT - Human Cells	transcription profiling by array	32
E-MEXP-2286	Transcription profiling of human prostate cancer cells over-expressing androgen receptor following dihydrotestosterone treament		32
E-MEXP-3530	MicroRNA profiling by array of prostate after goserelin and bicalutamide treatments	microRNA profiling by array	28
E-MEXP-3081	Transcription profiling by array of human prostate cancer samples after treatment with bicalutamide (antiandrogen) or goserelin (GnRH agonist)	transcription profiling by array	28
E-MEXP-1058	MicroRNA profiling of human prostate cancer cell lines, xenografts and tumor samples	microRNA profiling by array	28
E-MTAB-6062	Transcription profiling of irradiated non-adherent anoikis-resistant DU145 and MCF-7 cells and 5-azacytidine-treated non-adherent anoikis-resistant HeLa cells in contrast to control (non-irradiated, non-treated) cells	transcription profiling by array	26
E-SMDB-3636	Transcription profiling of human androgen receptor expressing prostate carcinoma cell line LNCaP and normal human foreskin fibroblasts expressing the androgen receptor treated with dihydrotestosterone (DHT), ethanol, or untreated *vs.* a common reference, see E-SMDB-3637	transcription profiling by array	26
E-MTAB-4869	Transcription profiling of IGR-CaP1 prostate cancer cells resistant to docetaxel compared to non-resistant cells	transcription profiling by array	24
E-MTAB-4753	DNA methylation variations are required for reversible EMT induced by cancer-associated fibroblasts in PCa cells	methylation profiling by array	24
E-SMDB-3637	Transcription profiling of human androgen receptor expressing prostate carcinoma cell line LNCaP and normal human foreskin fibroblasts expressing the androgen receptor treated with dihydrotestosterone (DHT), ethanol, or untreated *vs.* a common reference, see E-SMDB-3636	transcription profiling by array	24
E-MTAB-5121	The role of the transcription factors GATA2 and FOXA1 in immortalized basal-like prostate epithelial cells	transcription profiling by array	23
E-MTAB-3438	Transcription profiling of effected genes by compound BIO,3G4 and knockdown of MED23	transcription profiling by array	23
E-MEXP-2313	miRNA profiling of human prostate cancer cell lines treated with 5azadC and TSA to investigate epigenetic modifications	microRNA profiling by array	22
E-MTAB-1572	Proteomic profiling by array of prostate cancer tumor samples with different sensitivities to androgen deprivation and under different severities of hypoxia	proteomic profiling by array	21
E-SMDB-3867	Transcription profiling of human prostatic stromal cells cultured from diseased *vs.* normal tissues	transcription profiling by array	19
E-MEXP-335	Comparative genomic hybridization of 5 human prostate cancer cell lines and 13 prostate cancer xenografts to create genomic profiles of copy number alterations	comparative genomic hybridization by array	19
E-MTAB-5021	Transcriptional differences between the peripheral and the transcription zone of the prostate	RNA-seq of coding RNA	18
E-MTAB-3691	Differential Ago-RIP-Seq for the identification of miR-375 targets in prostate cancer cells	RIP-seq	18
E-MTAB-3421	Knockdown of DHRS7 in the human prostate cancer cell line LNCaP	transcription profiling by array	18
E-MTAB-1521	Transcription profiling by array of human prostate cancer cell lines to investigate drug targeting of the IL6/STAT3 pathway	transcription profiling by array	16
E-MTAB-986	ChIP-seq study using a cell line model of ER_ AR+ molecular apocrine tumours (AR_FoxA1_molecular_apocrine)	ChIP-seq	16
E-MTAB-4966	Expression profiling of a prostate cancer cell line(OPCT1) and its clonal progenies with different functional characteristics	transcription profiling by array	15
E-TABM-532	Transcription profiling of human prostate carcinoma cell line PC3 treated with reverse transcriptase inhibitor abacavir	transcription profiling by array	15
E-TABM-1049	Transcription profiling by array of human prostate cancer cells treated with monensin to investigate the effect on apoptosis induction and oxidative stress		14
E-MEXP-2319	MicroRNA profiling of human prostate cancer	microRNA profiling by array	13
E-MEXP-520	Methylation profiling in various human cell lines and tissues by mDIP - methylated DNA precipitation with antibodies against methylated cytosine	methylation profiling by array	13
E-MTAB-5102	Development of a small molecule for treatment of castration resistant prostate cancer via androgen receptor and IL6/STAT3 pathways	transcription profiling by array	12
E-MTAB-3730	Transcriptome profiling after CBX7 knockdown in LNCaP cells	transcription profiling by array	12
E-MTAB-4752	DNA methylation variations are required for reversible EMT induced by cancer-associated fibroblasts in PCa cells	RNA-seq of coding RNA	12
E-MTAB-2838	IGR_JUNCONCO_STUDY_LM	transcription profiling by array	12
E-MTAB-1749	ChIP-seq of human LNCaP prostate cancer cell line and MDA-MB-453 molecular apocrine breast cancer cell line with antibodies against androgen receptor (AR) with or without overexpression of FoxA1	ChIP-seq	12
E-MTAB-1221	Transcription profiling by array of Docetaxel resistant human prostate cancer cell lines established by exposure to different doses of Docetaxel	transcription profiling by array	12
E-SMDB-6	Transcription profiling of HPEC senescent *vs.* immortalized cells	transcription profiling by array	12
E-MTAB-773	Transcriptional profiling of PC-3 human prostate cancer cells in response to caffeic acid phenethyl ester treatment	transcription profiling by array	12
E-SMDB-3259	Transcription profiling of human prostate cancer cells treated with resveratrol	transcription profiling by array	12
E-MEXP-336	Transcription profiling of four human prostate cancer cell lines and seven prostate cancer xenografts	transcription profiling by array	11
E-MTAB-108	Transcription profiling by array of human LNCaP cells transfected with GFP-FOXP3 cDNA	transcription profiling by array	10
E-MEXP-461	Transcription profiling of human Ki-ras transformed embryo prostate epithelial cells (267B1) to identify mRNAs under differential translational control	transcription profiling by array	10
E-SMDB-3938	Transcription profiling of human prostate cancer cells (LNCaP) treated with selenomethionine or methylselenic acid	transcription profiling by array	10
E-MEXP-476	Transcription profiling of CD146 immunomagnetically enriched circulating endothelial cells (CECs) from healthy donors and patients with metastatic breast, colorectal, prostate, lung and renal cancer	transcription profiling by array	10
E-SMDB-2030	Transcription profiling of prostate cancer cells with Akt activation	transcription profiling by array	9
E-MTAB-2142	Transcription profiling by array of the compound U0126 in PC3 prostate cancer cells	transcription profiling by array	8
E-MTAB-1204	ChIP-seq of human cells from a primary prostate cancer with poor outcome and metastatic LNCaP cells in basal condition and after 17b-Estradiol (E2) treatment	ChIP-seq	8
E-TABM-1172	Transcription profiling by array of human VCaP prostate cancer cell line after PLA2G7 siRNA treatment	transcription profiling by array	8
E-TABM-949	Transcription profiling by array of human prostate carcinoma cells during a stepwise epithelial to mesenchymal transition	transcription profiling by array	8
E-TABM-635	Chromatin immunoprecipitation of human prostate cell lines indicates an H3K4me3/H3K27me3 epigenetic signature of prostate carcinogenesis	ChIP-chip by array	8
E-MEXP-803	Comparative genomic hybridization of benign epithelial and prostate cancer cell lines derived from the same patient	comparative genomic hybridization by array	8
E-SMDB-2973	Transcription profiling and comparative genomic hybridization of prostate cancer cell lines	transcription profiling by array	8
E-SMDB-2972	Transcription profiling and comparative genomic hybridization of prostate cancer cell lines	transcription profiling by array	8
E-MTAB-5150	3prime RNA-seq of human prostate cancer cell line DU-145 treated with Senexin A	RNA-seq of coding RNA	6
E-MTAB-845	Transcription profiling by array of human DU145 cells treated with small molecule MS0019266	transcription profiling by array	6
E-SMDB-4028	Transcription profiling of human prostate cancer cell lines after androgen depletion and AR knock-down	transcription profiling by array	6
E-MEXP-136	Transcription profiling of circulating tumor cells (CTC) from peripheral blood from patients with breast and prostate cancer	transcription profiling by array	6
E-MTAB-4118	Controls and CNTN1 overexpression in DU145 cells and CNTN1 knockdown in DU145 cell-derived prostate cancer stem-like cells	transcription profiling by array	4
E-MTAB-1786	Transcription profiling by array of castration-resistant prostate cancer PC-3 cells treated with Hsp27-siRNA or control siRNA to study the role of Heat shock protein (Hsp) 27 in splicing	transcription profiling by array	4
E-MEXP-2943	Searching targets for miR-32 and miR-148a	microRNA profiling by array	4
E-MEXP-581	Transcription profiling of human PC3 prostate cells transfected with FGF-8b *vs.* control vector		4
E-MEXP-2172	Transcription profiling by array of human DU-145 and PC-3MM2 cells after gamma irradiation	transcription profiling by array	4
E-TABM-78	Transcription profiling of neuroendocrine-like LNCaP-cells	transcription profiling by array	4
E-SMDB-3416	Transcription profiling of 4 prostate cancer cell lines treated with the DNA methyltransferase inhibitor 5-aza-dC	transcription profiling by array	4
E-MTAB-3504	Integrated and functional genomics analysis validates the relevance of the nuclear variant ErbB380kDa in prostate cancer progression	ChIP-chip by array	3
E-MTAB-3499	Integrated and functional genomics analysis validates the relevance of the nuclear variant ErbB380kDa in prostate cancer progression	ChIP-chip by array	3
E-MTAB-3087	Comparative MicroRNA Expression Profiles of Penile Cancer Revealed by Next-Generation Small RNA Deep Sequencing	microRNA profiling by high-throughput sequencing	2
E-MEXP-1585	Chromatin immunoprecipitation of trimethylated histone H3-K27 in human prostate cancer cell line PC3	ChIP-chip by array	2
E-MEXP-1581	RNAi knock-down of EZH2 in mouse prostate cancer cell line PC3	RNAi profiling by array	2
E-MEXP-1627	Transcription profiling of human PC-3 prostate cancer cells expressing shTCEB1 leading to TCEB1 silencing	transcription profiling by array	1