

# PDXNet portal: patient-derived Xenograft model, data, workflow and tool discovery

Soner Koc<sup>1,†</sup>, Michael W. Lloyd<sup>2,†</sup>, Jeffrey W. Grover<sup>3</sup>, Nan Xiao<sup>1</sup>, Sara Seepo<sup>1</sup>, Sai Lakshmi Subramanian<sup>1</sup>, Manisha Ray<sup>1</sup>, Christian Frech<sup>1</sup>, John DiGiovanna<sup>1</sup>, Phillip Webster<sup>1</sup>, Steven Neuhauser<sup>2</sup>, Anuj Srivastava<sup>3</sup>, Xing Yi Woo<sup>3</sup>, Brian J. Sanderson<sup>3</sup>, Brian White<sup>3</sup>, Paul Lott<sup>4</sup>, Lacey E. Dobrolecki<sup>5</sup>, Heidi Dowst<sup>5</sup>, PDXNet Consortium, Yvonne A. Evrard<sup>6</sup>, Tiffany A. Wallace<sup>7</sup>, Jeffrey A. Moscow<sup>8</sup>, James H. Doroshov<sup>9</sup>, Nicholas Mitsiades<sup>5</sup>, Salma Kaochar<sup>5</sup>, Chong-xian Pan<sup>4,10</sup>, Moon S. Chen<sup>4</sup>, Luis Carvajal-Carmona<sup>4</sup>, Alana L. Welm<sup>11</sup>, Bryan E. Welm<sup>11</sup>, Michael T. Lewis<sup>5</sup>, Ramaswamy Govindan<sup>12</sup>, Li Ding<sup>12</sup>, Shunqiang Li<sup>12</sup>, Meenhard Herlyn<sup>13</sup>, Michael A. Davies<sup>14</sup>, Jack Roth<sup>14</sup>, Funda Meric-Bernstam<sup>14</sup>, Peter N. Robinson<sup>3</sup>, Carol J. Bult<sup>2</sup>, Brandi Davis-Dusenbery<sup>1</sup>, Dennis A. Dean, II<sup>1,‡</sup> and Jeffrey H. Chuang<sup>3,\*</sup>

<sup>1</sup>Seven Bridges, Charlestown, MA 02129, USA, <sup>2</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA, <sup>3</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA, <sup>4</sup>University of California - Davis, Davis, CA 95616, USA, <sup>5</sup>Baylor College of Medicine, Houston, TX 77030, USA, <sup>6</sup>Leidos Biomedical Research, Inc, Frederick National Laboratory for Cancer Research, Frederick, MD 21701, USA, <sup>7</sup>Center to Reduce Health Disparities, National Cancer Institute, Bethesda, MD 20814, USA, <sup>8</sup>Investigational Drug Branch, National Cancer Institute, Bethesda, MD 20814, USA, <sup>9</sup>Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20814, USA, <sup>10</sup>Harvard Medical School, West Roxbury, MA 02115, USA, <sup>11</sup>Huntsman Cancer Institute, Salt Lake City, UT 84112, USA, <sup>12</sup>Washington University School of Medicine, St. Louis, MO 63110, USA, <sup>13</sup>The Wistar Institute, Philadelphia, PA 19104, USA and <sup>14</sup>The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Received November 17, 2021; Revised March 04, 2022; Editorial Decision April 01, 2022; Accepted April 07, 2022

## ABSTRACT

We created the PDX Network (PDXNet) portal (<https://portal.pdxnetwork.org/>) to centralize access to the National Cancer Institute-funded PDXNet consortium resources, to facilitate collaboration among researchers and to make these data easily available for research. The portal includes sections for resources, analysis results, metrics for PDXNet activities, data processing protocols and training materials for processing PDX data. Currently, the portal contains PDXNet model information and data resources from 334 new models across 33 cancer types. Tissue samples of these models were deposited in the NCI's Patient-Derived Model Repository (PDMR) for public access. These models have 2134 associated sequencing files from 873 samples across 308 patients,

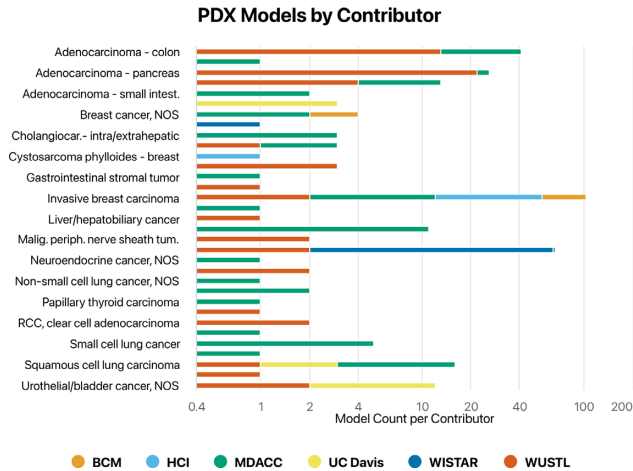
which are hosted on the Cancer Genomics Cloud powered by Seven Bridges and the NCI Cancer Data Service for long-term storage and access with dbGaP permissions. The portal includes results from freely available, robust, validated and standardized analysis workflows on PDXNet sequencing files and PDMR data (3857 samples from 629 patients across 85 disease types). The PDXNet portal is continuously updated with new data and is of significant utility to the cancer research community as it provides a centralized location for PDXNet resources, which support multi-agent treatment studies, determination of sensitivity and resistance mechanisms, and preclinical trials.

\*To whom correspondence should be addressed. Tel: +1 837 856 2473; Email: [jeff.chuang@jax.org](mailto:jeff.chuang@jax.org)

†Mr Koc and Dr Lloyd are co-first authors providing an equal contribution to the manuscript.

‡Dr Dean and Dr Chuang are co-senior authors providing equivalent oversight to work presented in this manuscript.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Patient-derived Xenograft (PDX) models are cancer models that support personalized medicine research and pre-clinical and co-clinical trials (1–5). Specific PDX research areas include the study of sensitivity and resistance mechanisms, evaluation of new treatment options, and the study of tumor heterogeneity. The PDX research community is rapidly growing, with PDX-generated data being the preferred support for proposing human clinical trials (6). In 2017, the National Cancer Institute (NCI) funded the PDX Development and Trial Centers (PDTC) research network (PDXNet, <https://www.pdxnetwork.org/>) consortium to accelerate PDX research by developing new PDX models across cancer types, identifying new multi-agent treatments to bring forward into clinical trials, generating complementary RNA-Seq and whole-exome sequencing data, and increasing the ethnic diversity of publicly available PDX models.

PDXNet was also charged with developing collaborative research projects involving the six different PDTCs to advance PDX science. Each of the PDTCs came into PDXNet with its own home-grown data standards, data analytic pipelines and workflows. To facilitate collaboration, the disparate processes and databases required harmonization at many different steps, so that data from centers could be combined and analyzed efficiently. The harmonization goal was achieved through the creation of the PDXNet portal and the analytical tools within it. The PDXNet portal resources created by this effort enabled the successful completion of several collaborative research projects (7–10) and are supporting many others.

In addition to facilitating PDXNet research, a further benefit of the PDXNet portal is to make the PDXNet-generated data and workflows of the PDXNet portal available as a public resource. These data will support cancer research by increasing the quantity and diversity of PDX data available and decreasing the effort required to analyze PDX sequencing data. We present the PDXNet portal as a utility for PDXNet data for the larger scientific community.

The PDXNet portal is already supporting the goals of data harmonization, discovery, generation and analysis. To

demonstrate the utility of the portal resource, we highlight an in progress PDXNet pilot project. Initial review of model metadata identified the need for standardized ancestry information across the PDTCs. For this project, the analyst accessed the metadata (downloadable from the PDXNet portal and described below) to identify related sequencing data from PDX model IDs. For each available model, the analyst then generated standardized ancestry estimates with a published ancestry inference workflow (see result section for details). These ancestry estimates were then added to the portal metadata in a searchable interface. Moreover, we created a portable and reproducible version of the publish ancestry inference workflow, allowing other researchers to use and independently validate once linked to the PDXNet Portal.

There are several existing public PDX resources that complement the PDXNet portal. Launched in 2012, the NCI Patient-Derived Model Patient Repository (PDMR; 11) collects and develops PDX models and associated standardized sequencing data (RNA-Seq and whole-exome), with the goal of supporting academic and industry research. The PDMR maintains a publicly available database of models and a File Transfer Protocol (FTP) site for accessing sequencing data (<https://pdmr.cancer.gov/database/default.htm>). Another resource is PDXFinder (<https://www.pdxfinder.org>, 12). PDXFinder is an online resource that aims to harmonize internationally generated PDX models and their associated metadata. PDXFinder is a collaboration between the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and the Jackson Laboratory. A key component of data harmonization in PDXFinder is the PDX minimal information standard (PDX-MI; 13), which allows for standardized PDX information exchange. PDXFinder employs PDX-MI to support complex model searches that enable users to identify model descriptions and subsequently link to model information and a request form. EuroPDX is a consortium of eighteen non-profit cancer institutes that collaborate and coordinate PDX model development and access to improve cancer patients' standard of care. Next, the EuroPDX Data Portal (<https://dataportal.europdx.eu/>) is a resource that provides information about PDX models generated by EuroPDX researchers and clinicians. Lastly, the Baylor College of Medicine PDX portal (<https://pdxportal.research.bcm.edu/>) provides access to breast cancer, leukemia, pediatric liver cancer, pancreatic cancer and sarcoma model collections.

The primary aim of the PDXNet portal is to support the Cancer Moonshot (14) model and data sharing goals (15). The PDXNet portal facilitates the distribution of resources, complementary data analyses and developed tools. The resources generated include data collections and standardized bioinformatics workflows. Complementary data analyses such as data quality control analyses that support data-use are also available from the portal. Tools developed to support the use of the data (e.g. workflow cost estimation) are also accessible. Integration with the NCI Cloud Resource, the Cancer Genomics Cloud powered by Seven Bridges (CGC; 16), allows approved researchers to directly analyze PDXNet data or use developed workflows on private data.

This manuscript details the PDXNet portal features that provide a gateway for identifying and accessing resources generated by the PDXNet community.

## MATERIALS AND METHODS

### Portal features

The PDXNet portal is a publicly accessible website (<https://portal.pdxnetwork.org/>) with the primary function of providing access to the PDXNet models and information on how to obtain sequencing data. We extended the portal's primary mission beyond PDXnet to incorporate additional resources, including supporting access to the PDMR sequencing file data set, a PDXNet hematoxylin and eosin stain image data set, and tumor volume data. Below, we describe the features and sections of the Portal in detail.

### PDXNet portal landing page

The PDXNet portal landing page includes an overview and summary panel of contents (Figure 1). The portal overview identifies the primary PDXNet funding sources and participants. In summary, the NCI Cancer Therapy Evaluation Program (CTEP) funds four PDTCs and the PDX Data Commons and Coordinating Center (PDCCC), whereas the NCI Center to Reduce Cancer Health Disparities funds two PDTCs (see Table 1 for additional details). The portal directs questions and requests for additional information to the PDXNet website at <https://pdxnetwork.org>.

The data summary panel on the right side of the screen allows the reader to review model and data summaries and the portal update timeline. The data summary panel lists the number of PDTCs contributing data, the number of files uploaded by the PDTCs and available from the PDMR, the total number of models, and the number of cancer types represented in those models. Dropdown menus on the top bar of the landing page include 'Resources', 'Analysis', 'Tools', 'Help', 'About' and 'Contact', which are further summarized below. Buttons at the bottom left provided quick access to PDXNet model information, data resources, and data search. Tabs associated with the right-hand data summaries allows the reader to review summary figures for PDX Models by cancer type and contributing PDTC, sequencing files by experimental strategy, ancestry and the Portal Update Timeline.

### Resources tab

The PDXNet Portal Resources tab includes the following links to separate pages: PDXnet models, Data Collection, Interactive Exploration, and Workflows, and subpages are described in subsequent sections.

**PDXNet models.** The PDXNet portal models tab summarizes verified model submissions to the PDMR made by each PDTC (Supplementary Figure S1). The PDXNet models are a primary consortium deliverable. Each model submitted by a PDTC to the PDMR includes a completed model submission form that details the general PDX information, model-specific information, and tissue implantation details. Metadata are consistent with the PDX-MI

and the PDMR data format. The metadata includes model id information to facilitate search and cross referencing to related PDMR models.

To date, PDXNet researchers have submitted 334 models to the PDMR across 33 cancer types. The most prevalent model cancer types include invasive breast carcinoma (30.8%, 103), melanoma (20.1%, 67), adenocarcinoma—colon (12.3%, 41), and adenocarcinoma—pancreas (7.8%, 26). See Table 2 for additional details.

**Data search.** The PDXNet Portal data search page allows users to query across data types, and harmonizes metadata across model information, sequencing files and image files. Users query from data source, samples and disease types. Query results are shown in Venn diagrams and users can download data tables organized by sample, model or patient at the bottom of the page. The data search feature specific currently queries across with corresponding PDMR image, RNA-seq and WES data.

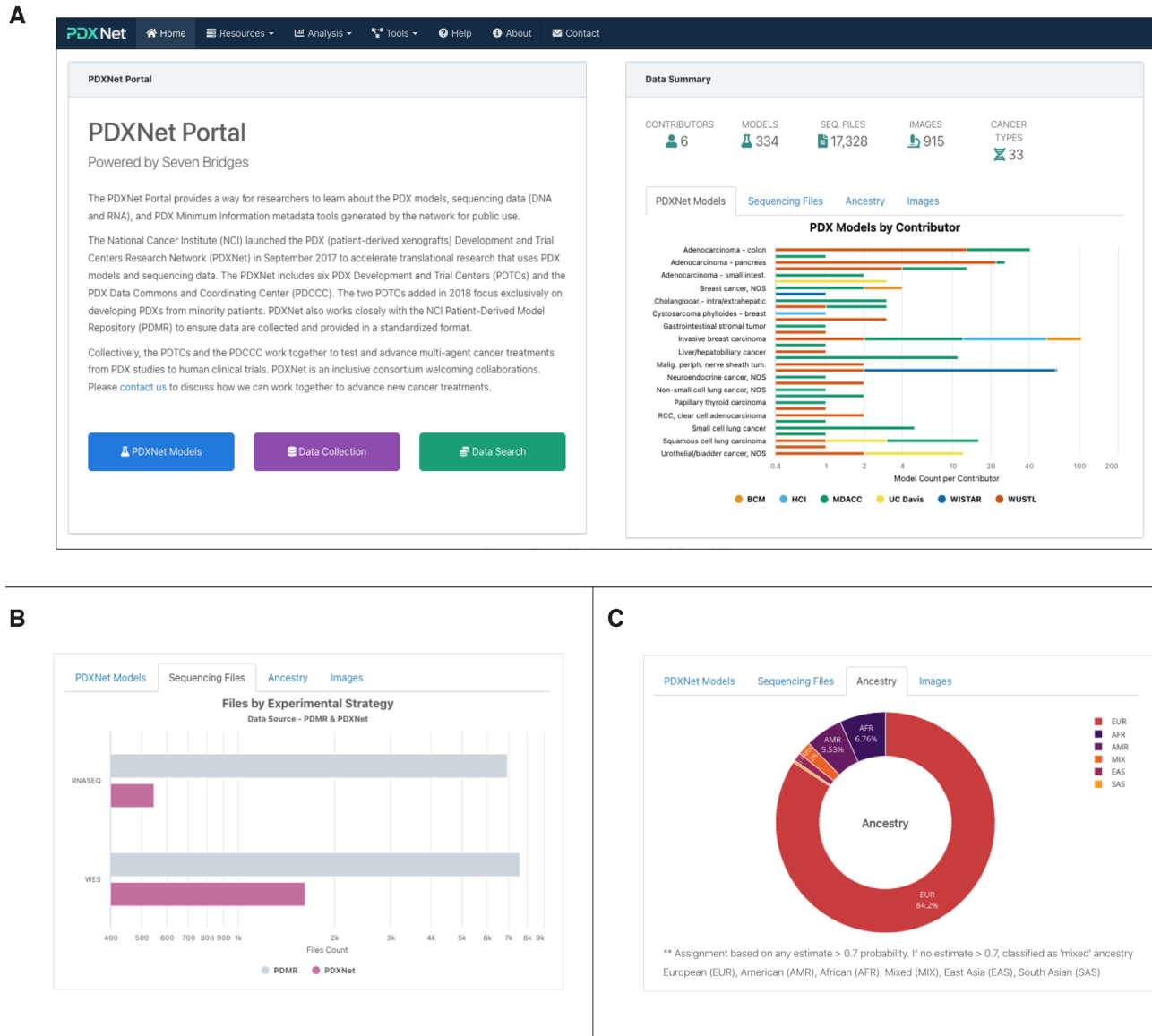
**Data collection.** The PDXNet Portal Data Collection tab summarizes sequencing data from the PDMR and PDTCs, and image files.

**Data collection, PDMR.** This sub-tab summarizes sequencing data transferred from the PDMR FTP server to the CGC as of December 2021 (Supplementary Figure S2). The PDMR generates whole-exome and transcriptome sequencing data from models submitted according to tissue collection best practices and model quality control practices (11). Molecular characterizations include whole-exome sequencing and gene expression profiling. We processed the PDMR sequencing data with standardized workflows as for the PDTC data.

**Data collection, PDTC.** This sub-tab summarizes sequencing data submitted by the PDTCs for intra-consortium sharing and for public sharing (Figure 2). The PDXNet Data Collection, PDTC tab presents the core PDXNet sequencing data set. We processed submitted sequencing data with standardized workflows (e.g. whole exome capture) according to a written standard operating procedure provided on the CGC. See the workflow section for description of the workflows used for standardized processing.

PDXNet researchers contributed 2134 total sequencing files that include both whole-exome (74.5%, 11 590) and RNA-Seq (25.5%, 544) data. Six institutional contributors submitted 873 samples from 308 patients. The sequencing sample types include PDX (56.0%, 1194), tumor (25.7%, 548) and normal (18.4%, 392). The most prevalent diseases represented among the samples include breast (27.5%, 586), skin (18.8%, 402), lung (17.1%, 364) and colon (8.4%, 180). Metadata provided by centers did not include disease information for 0.9% (20) samples (see Table 3 for additional information).

The PDMR sequencing dataset on the CGC includes 14 506 paired-end sequencing files that include both whole-exome (52.2%, 7576) and RNA-Seq (47.8%, 6926) data. The data set includes 3857 samples from 629 patients covering



**Figure 1.** PDXNet Portal Landing Page Views from the PDXNet Portal Landing Page. (A) The initial PDXNet Portal Landing Page. (B) Experimental strategies (whole-exome and RNA-Seq) for the PDXNet (magenta) and PDMR (gray) sequencing datasets. (C) Computed ancestry in a pie chart. Ancestry is classified in the following categories: European (EUR-Red), African (AFR-Blue), American (AMR-Purple), Mixed (MIX-Orange), East Asian (EAS-Light Purple), South Asian (SAS-Yellow).

85 disease types. The sequencing sample types include PDX (81.4%, 11 808), primary tumor (5.6%, 812), pdc: mixed tumor culture (4.7%, 684) and normal germline (4.4%, 644). The most prevalent diseases represented among the samples include colon (19.2%, 2,780), pancreas (9.2%, 1,332), melanoma and (6.5%, 938). Due to the size and cost associated with data transfers, synchronization between the PDMR sequencing database and the CGC dataset is done periodically. The PDMR data webpage has the most updated list of available PDMR sequencing data processed with standardized PDXNet workflows.

**Data collection, Image.** This sub tab summarizes hematoxylin-eosin stain (H&E) image data provided

by the PDMR (Supplementary Figure S3). The PDMR image data on the CGC includes 194 images scanned from PDX (96.4%, 187) and primary tumors (3.6%, 7). The images correspond to 194 samples taken from 27 patients across 4 disease types. The PDX passages ranged from P0 to P6 with the top four passages corresponding to P1(40.6%, 73), P0 (25.0%, 45), P2 (22.8%, 41) and P3 (7.2%, 13). The PDXNet Portal currently supports 43 metadata fields that data submitters can populate upon submission (See Supplementary Table S1 for the complete list).

**Interactive exploration.** The PDXNet Portal data explorer allows users to interactively create summary tables from



**Table 1.** PDXNet Development and Trial Centers (PDTC) and the PDX Data Commons and Coordinating Center (PDCCC)

PDXNet	
<b>PDX Development and Trials Centers (PDTC)</b>	
HCI-BCM*	Huntsman Cancer Institute and Baylor College of Medicine
MDACC	MD Anderson Cancer Center
WUSTL*	Washington University at St. Louis
WISTAR*	Wistar Institute and MD Anderson Cancer Center
BCM& UCDAVIS&	Baylor College of Medicine The University of California at Davis
<b>PDX Data Commons and Coordinating Center</b>	
JAX-SB*&	Jackson Laboratory and Seven Bridges

\*NCI Cancer Therapy Evaluation Program Funding

&amp;NCI Center to Reduce Cancer Health Disparities Funding

PDMMR and PDTC sequencing file metadata (Supplementary Figure S4). The interactive table supports 10 table and chart types including simple tables, bar charts, line charts, and heat maps (see Supplementary Table S2 for full list). Interactive tables also support 22 data summary options including count, sum, average and variance (see Supplementary Table S3 for the full list). The user drags and drops from 20 metadata fields to the table type area to construct the table. Metadata field examples include contributor, sample type, experimental strategy and passage (see Supplementary Table S4 for complete list).

**Workflows.** The PDXNet Portal Workflows page summarizes analysis workflows developed by the PDXNet community (Supplementary Figure S5). The CGC based analysis workflows are a significant resource developed by the PDXNet community, allowing for reproducible and standardized analysis of PDX data (8). We selected workflows for standard consortium-wide data processing and public release from those submitted by each PDTC after benchmarking with simulated and experimentally derived PDX data (8). Since the initial public release, we have restructured the workflows to efficiently process normal (tissue), tumor-only and tumor-normal data. These workflows are implemented on the CGC using the Common Workflow Language (CWL; 17) with Docker containerized tools, which allows for easy sharing and analysis reproducibility. The PDXNet consortium developed a set of 15 workflows validated for processing of both whole-exome and RNA-Seq data (See Supplementary Table S5). We are sharing these validated and tested workflows with the broader community via the CGC Public Apps Gallery (<https://cgc.sbggenomics.com/public/apps#q?search=pdx>). The use of CWL allows these workflows to be portable to any CWL-compliant execution environment. The workflows collectively facilitate the analysis of whole-exome or RNA-Seq data via mouse read disambiguation, read alignment, variant calling or transcript quantification, and sample and cohort level quality control. For whole-exome data, we also compute copy number variation (CNV), tumor mutational burden (TMB), microsatellite instability (MSI) and homologous recombination deficiency (HRD) during standardized processing. A full explanation of inputs, outputs and

data processing steps for each workflow is provided on the CGC in the respective description section.

### Analysis tab

The PDXNet Portal Analysis tab includes the following menu options: Ancestry Analysis, HRD-MSI-TMB Analysis, Tumor Volume Analysis, QC Analysis and Variant Search. These sections include metrics derived from primary data sources and are described in the results section below in more detail. These results were generated from standard processing analysis workflows or through PDXNet research activities (8), and we provide these analyses to support independent research by the broader research community.

### Tools tab

The PDXNet Portal Tools tab includes the following choices: Workflow Cost Prediction, Metadata -Create, Metadata -Validate. These tools, described in the following sections, support present and future PDXNet and other independent general research activities.

**Workflow cost prediction.** The Workflow cost prediction tool allows users to estimate the cost of processing their samples on the CGC with the PDXNet workflows. This tool uses prediction models (gradient boosting trees; 18) generated from 7000 workflow runs. The user can select either whole-exome or RNA-Seq workflows and provide the number and optionally size of files to process. The calculator computes the storage and computation cost for processing the user defined dataset. The estimated costs assume the workflows were run on the CGC using spot instances. We expect the tool to allow users to estimate data storage and computational cost for their own analyses allowing for estimating grant budgets and budgeting lab expenses.

**Create.** The generate metadata tab allows users to interactively generate a PDX minimum information metadata sheet (PDX-MI). As described above, the PDXFinder working group developed the PDX-MI as a standard for exchanging PDX information among institutions. The generate metadata tab allows the user to create a PDX-MI spreadsheet by stepping through data entry dialog boxes. The interface supports entry of patient information, treatment information, tumor information, model and sequencing metadata. The user downloads the spreadsheet upon data entry completion, and no information is stored permanently on the PDXNet Portal site.

**Metadata, validate.** The validate metadata tab allows users to upload and validate a PDX-MI metadata spreadsheet. The user can review uploaded contents at the bottom of the page. The validation verifies that required metadata fields are present and that entries are valid. The validation feature generates a summary of required fields that includes percent completed and most common data entry per field. The validation feature reduces the amount of time necessary to review and check submitted PDX-MI spreadsheets.

**Table 2.** PDX models generated by PDX Development and Trials Centers

	HCI-BCM	MDACC	WUSTL	Wistar	UC Davis	BCM	Totals
Breast	78	12	2	0	0	16	108
Head and Neck	0	0	2	0	0	0	2
Digestive/Gastrointestinal	0	50	44	0	3	0	97
Endocrine and Neuroendocrine	0	2	0	0	0	0	2
Musculoskeletal	0	1	4	0	0	0	5
Respiratory/Thoracic	0	30	1	0	2	0	33
Skin	0	3	2	63	0	0	68
Genitourinary	0	0	5	0	10	0	15
Gynecologic	0	3	0	0	0	0	3
Unknown Primary	0	0	0	1	0	0	1
<b>Total</b>	<b>78</b>	<b>101</b>	<b>60</b>	<b>64</b>	<b>15</b>	<b>16</b>	<b>334</b>

*HCI-BCM*: Huntsman Cancer Institute and Baylor College of Medicine, *MDACC*: MD Anderson Cancer Center, *WUSTL*: Washington University at St. Louis, *Wistar*: The Wistar Institute, *UC Davis*: University of California Davis, *BCM*: Baylor College of Medicine

### Portal design and organization

The PDXNet Portal is designed to support coordination with other resources, including the PDMR, the CGC and the NCI Cancer Data Service (CDS; <https://datacommons.cancer.gov/>). Currently, the PDXNet portal references PDMR model information, genomic, transcriptomic and tumor volume response data used in PDXNet research activities. The CGC serves as a PDXNet data staging area supporting data harmonization, standardized data processing and research activities. The PDXNet Portal supports submission of sequencing data to the CDS to provide long-term research access to PDXNet data resources. The CDS is part of the NCI Cancer Research Data Commons which aims to store data resources generated by NCI-funded research. The CDS is available from across the NCI data infrastructure through a dbGaP access mechanism. The PDXNet portal augments the dbGaP submission process, through an administrative feature for generating data reports and through a dbGaP submission tool written to support CDS submissions. The PDXNet Portal aims to use data standards when they exist, supporting both the PDX-MI standard (13) and the PDMR data structures (11). These existing data structures allow for collaboration and information sharing with existing PDX resources.

### Implementation

The backend of the PDXNet Portal is an R-Shiny app hosted on a cloud-based server. The portal uses the PDMR and PDX-MI metadata standards. The PDMR and PDX-MI standards allow us to harmonize data across sources, quickly import data from the PDMR and other data sources, and exchange information with other PDX related portals. We also collect additional metadata required to facilitate computation on the CGC, including omics-related information. We use the Cancer Therapy Evaluation Program (CTEP) disease classification to standardize disease entries although we initially accepted institutionally defined disease classification. In the cases where a standard does not exist; we collect sufficient metadata required to display and process the data source. For example, we take a minimalistic approach to managing tumor volume and H&E image data. Several PDXNet teams are working toward the development of best practices for these data types. Until

these best practices are published, we will evolve these operational standards to support harmonization and analysis.

The PDXNet Portal team updates information on the portal semi-automatically using the same data model as the PDMR, allowing PDXNet to sync with PDMR model information. The PDMR provides regular updates to PDXNet on PDTC model submissions to update the PDX Model's page. We receive sequencing data upload updates from the PDTCs and the PDMR, and we have developed scripts for extracting PDXNet standardized processing results, allowing for quality control information and computational metrics to be tabulated for semi-automated PDXNet Portal updates. The PDXNet Portal source code will not be made publicly available for security reasons. Future PDXNet Portal versions will support controlled access sign-in to provide links to controlled files.

## RESULTS

All analysis subsections described below can be found as selections under the 'Analysis' tab in the menu bar of the portal page.

### Ancestry analysis

The PDXNet Portal Ancestry Analysis page summarizes genetic ancestry analysis for datasets on the portal (Supplementary Figure S6). We compute ancestry with SNPweight (19) using a reference dataset generated from the 1000 Genomes Project Phase III (20). We classify each sample into one of five categories, which correspond roughly to the concept of 'continental ancestry' (21). These categories include European (EUR), African (AFR), American (AMR), East Asia (EAS) and South Asian (SAS). Samples that could not be confidently assigned to one of these categories are labeled Mixed (MIX). On the left side of the page, ancestry data filters allow the user to select the data contributors, ancestry and disease type. Applying the selected filter to the data regenerates the two summary figures. The first summary figure is a bar chart that shows the ancestry distribution for the selected disease types. The second summary figure shows a pie chart with each slice corresponding to the ancestry types chosen. Supplementary Table S6 shows the summary of ancestry estimation from PDX Models submitted to the PDMR.

A



B

**Metadata**

Column visibility Copy Print Download

Show 5 entries

Search:

contributor	file_name	patient_id	model_id	tumor_id	case_id	sample_id	sample_type	experimental_strategy	gender	paired_end	capture_kit	capture_assembly	availability	public	acc
1	BCM LZML_37246_Noindex_L00X_R1_001.fastq.gz	24452	BCM-0002	2011012-2		BCM-0002	PDX	RNA-Seq	Female	1	Nugen Ovation v2		PDXNet	Yes	
2	BCM LZML_37246_Noindex_L00X_R2_001.fastq.gz	24452	BCM-0002	2011012-2		BCM-0002	PDX	RNA-Seq	Female	2	Nugen Ovation v2		PDXNet	Yes	
3	BCM ML6065_GATCAGCG_L00X_R1_001.fastq.gz	2219	BCM-0046	2011012-46		BCM-0046	PDX	RNA-Seq	Female	1	Nugen Ovation v2		PDXNet	Yes	
4	BCM ML6065_GATCAGCG_L00X_R2_001.fastq.gz	2219	BCM-0046	2011012-46		BCM-0046	PDX	RNA-Seq	Female	2	Nugen Ovation v2		PDXNet	Yes	
5	BCM ML6061_CAGATCTG_L00X_R1_001.fastq.gz	24561	BCM-0104	2011012-104		BCM-0104	PDX	RNA-Seq	Female	1	Nugen Ovation v2		PDXNet	Yes	

Showing 1 to 5 of 2,166 entries

Previous 1 2 3 4 5 ... 434 Next

**Figure 2.** PDXNet sequencing data page on the PDXNet Portal Components of the PDXNet sequencing data page. (A) Panel shows summary statistics including number of sequencing files (green), contributors (yellow), total samples (orange) and total patients (blue). Also, shown are donut plots for contributors, sample types, disease type, experimental strategy, WES contributors and RNA-Seq contributors. (B) Panel shows metadata for the PDXNet sequencing data in a spreadsheet format. The interface supports searching and sorting metadata. Users can copy, print, and download metadata into accessible formats.

**Table 3.** Sequencing data files generated by PDX Development and Trial Centers

	Overall N = 2134	BCM-HCI N = 408	MDACC N = 418	UC Davis N = 48	WISTAR N = 382	WUSTL N = 878
<b>Experimental strategy n (%)</b>						
RNA-Seq	544 (25%)	102 (25%)	6 (1%)	32 (67%)	144(38%)	260(30%)
WES	1590 (75%)	306 (75%)	412(99%)	16 (33%)	238(62%)	618(70%)
<b>Disease type, n (%)</b>						
Bladder	114(5.0%)	2(<0.5%)	4(1.0%)	48(100%)	4(1%)	56(6.4%)
Blood	84(4.0%)	0(0%)	84(20.0%)	0(0%)	0(0%)	0(0%)
Bone	8(0.4%)	0(0%)	0(0%)	0(0%)	0(0%)	8(0.9%)
Breast	586(27%)	382(93%)	0(0%)	0(0%)	0(0%)	204(23%)
Colon	180(8.4%)	4(1.0%)	8(2.0%)	0(0%)	0(0%)	160(18%)
Gastrointestinal	34(1.6%)	0(0%)	0(0%)	0(0%)	0(0%)	34(3.9%)
Head and Neck	24(1.1%)	0(0%)	0(0%)	0(0%)	0(0%)	24(2.7%)
Kidney	58(2.7%)	0(0%)	0(0%)	0(0%)	0(0%)	58(6.6%)
Lung	364(17%)	0(0%)	322(77%)	0(0%)	0(0%)	42(4.8%)
Ovarian	8(0.4%)	0(0%)	0(0%)	0(0%)	0(0%)	8(0.9%)
Pancreas	174(8.1%)	0(0%)	0(0%)	0(0%)	0(0%)	174(20%)
Rectum	22(1.0%)	0(0%)	0(0%)	0(0%)	0(0%)	22(2.5%)
Skin	402(19.0%)	0(0%)	0(0%)	0(0%)	370(97%)	32(3.7%)
Soft Tissue Neoplasm	56(2.6%)	0(0%)	0(0%)	0(0%)	0(0%)	56(6.4%)
Unknown	20(0.9%)	20(5.0%)	0(0%)	0(0%)	0(0%)	0(1.7%)
<b>Sample type, n (%)</b>						
Normal	392(18%)	32(8%)	142(34%)	0(0%)	0(0%)	218(25%)
PDX	1,194(56%)	328(80%)	146(35%)	0(0%)	202(53%)	515(59%)
Tumor	548(26%)	48(12%)	130(31%)	48(100)	180(47%)	142(16%)

*HCI-BCM*: Huntsman Cancer Institute and Baylor College of Medicine, *MDACC*: MD Anderson Cancer Center, *WUSTL*: Washington University at St. Louis, *Wistar*: The Wistar Institute, *UC Davis*: University of California Davis, *BCM*: Baylor College of Medicine

### HRD-TMB-MSI analysis

The PDXNet Portal HRD-HSI-TMB analysis page allows the user to filter and summarize three computational metrics generated from whole-exome sequencing data by PDXNet standardized processing (Figure 3). The three computed metrics are homologous recombination deficiency (HRD), tumor mutational burden (TMB) and microsatellite instability (MSI). HRD is computed with ScarHRD (22) for matched normal data. TMB is calculated as the number of coding mutations that meet all quality criteria per Mb of the genome. Quality criteria are assessed using coverage, allele frequency, mapping quality and strand bias. Variants included in the calculation are somatic and non-polymorphic, and are defined in SnpEff (23) as ‘high’ or ‘moderate’ functional impact. As only a portion of the genome was sequenced, genome coverage (Mb) is calculated from the input target coverage BED file. MSI is calculated with MANTIS (24) for samples with matched normal data and calculated with MSIsensor2 (25) for tumor-only samples. For each metric, users can set data filters for the visualizations. The data filters, on the left side of the page, allow the user to select data contributor, sample type, experimental strategy and disease type. Applying the selected filter to the data generates a boxplot chart displaying the selected metrics for each disease type chosen (Figure 3).

### Tumor volume analysis

The PDXNet Portal Tumor Volume Analysis Page allows the user to visualize raw tumor volume growth data provided by the PDTCs (Figure 4). The filters enable the user to select contributor, treatment and disease type on the page’s left side. Applying the selected filter to the data regener-

ates the Tumor Volume and the Tumor Disease Types figure tabs. The currently available volume data is from 20 models representing 1 disease type, which were treated with 17 possible agents. The dataset has 10 432 volume measures from 20 treatment studies. The Tumor Volume tab allows the user to choose plot level (Animal and Treatment Arm) and plot pattern (multiple and combined), reorganizing the plots to correspond to select values. Additionally, under the ‘Tumor Volume – Study’ subtab, users can dynamically generate RECIST response classifications (Figure 4). The RECIST classification calculation is based on combination of the best response and average response adapted from Gao *et al.* (26).

### Quality control analysis

The PDXNet portal QC analysis page provides plots and tabular results for selected QC metrics (Supplementary Figure S7). The page displays QC metrics generated during the standardized data processing procedure for each relevant data type. The page provides sub-tabs showing whole-exome and RNA-Seq quality control metrics for a selected dataset PDXNet or PDMR. The whole-exome tabs plot mean target coverage, percent target bases with >20% coverage, and percent duplication by data contributor. The RNA-Seq tabs plot percent usable bases, percent ribosomal bases and percent correct strand reads. The plotted metrics, along with additional QC metrics are available in a table at the bottom of the page.

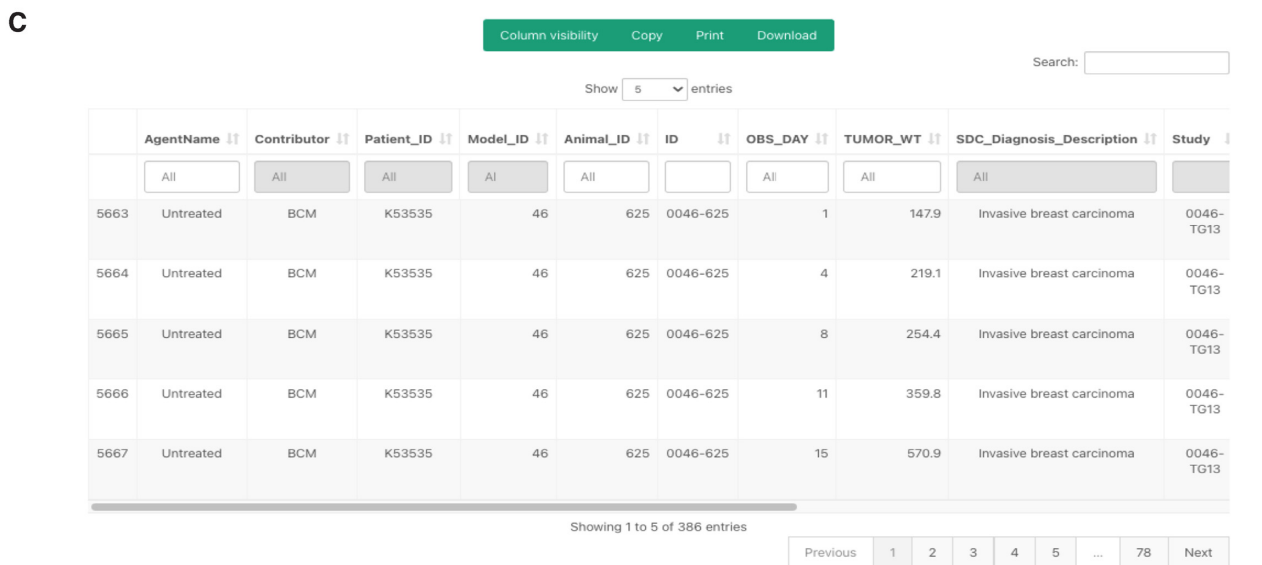
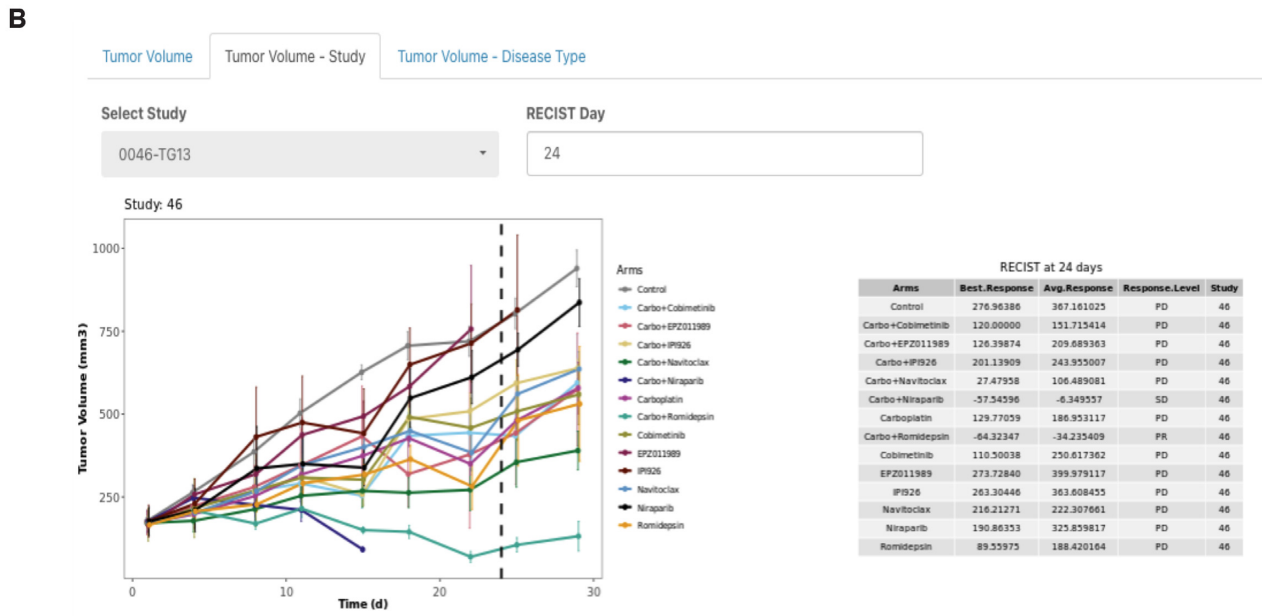
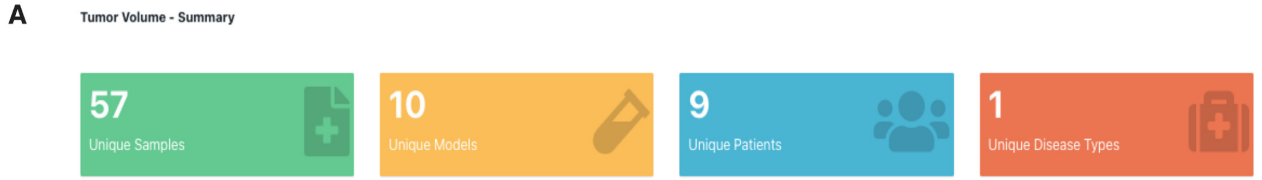
### Variant search

The PDXNet Portal Variant Search page allows the user to search for variants derived from whole exome data pro-





**Figure 3.** Examples figures generated from the HRD-TMB-MSI page on the PDXNet Portal Plots generated on the PDXNet Portal HRD-TMB-MSI page. **(A)** Plot of Homologous Recombination Deficiency (HRD) computed from sequencing data provided by PDXNet researchers. The plot shows HRD by disease type. **(B)** Plot of Tumor Mutational Burden (TMB) computed from sequencing data provided by PDXNet researchers. The plot shows TMB by disease type. **(C)** Plot of TMB computed from sequencing data provided by PDXNet researchers, by disease type.



**Figure 4.** Tumor volume data page on the PDXNet Portal Components of the PDXNet tumor volume page. The figure shows a filtered dataset. (A) Panel shows summary statistics including number tumor volume datasets (green), number of models in the selected dataset (yellow), total number of patients (blue), and total number of diseases represented (red). (B) Panel shows the tumor volume data organized by study and automatically computed resist category for each treatment arm. (C) Panel shows the tumor volume response and metadata in a spreadsheet format. The interface supports searching and sorting metadata. Users can copy, print, and download metadata into accessible formats

cessed using the PDXNet whole exome pipelines on the CGC. The page allows users to query based on either gene symbol or chromosome position. For each query, a summary of variant types, variants by gene and SNP types is shown. The variants returned in the query are available at the bottom of the page. We currently present all called variants for 491 samples as a beta feature release, with the ability for the user to further filter results as needed.

## DISCUSSION

The PDXNet portal is a vital component of the PDXNet consortium. The portal establishes a mechanism for public discovery of consortium-generated resources including models, data and workflows. The portal allows researchers to examine the data, models and metadata using integrated query features. These portal capabilities facilitate cancer data discovery, a core goal of the NCI Cancer Moonshot program. Additionally, the PDXNet portal allows the consortium to manage data analysis projects by clarifying which data are available, their source, their quality and their suitability for research projects and scientific questions. Within the PDXNet consortium, the PDXNet portal functions as a centralized source of information for the status of the available models. The standardized sequence processing, quality control, and computation of common metrics (e.g. MSI, TMB, HRD and genetic ancestry) further enhance data analysis, model use, prioritization of future models and data collection.

Enhancement of collaboration between researchers is a main objective of the PDXNet portal. To accomplish this, we are integrating the PDXNet portal and PDXNet data with existing NCI, NIH, and NCBI infrastructure. All data visible on the PDXNet Portal will also be available through the Cancer Data Service (CDS) through dbGaP (27) access. Accessing PDXNet data on the Cancer Genomics Cloud (16) allows users to perform sophisticated analyses utilizing cloud computing within an integrated bioinformatics ecosystem. By co-locating data and analysis, as well as integrating data management, this infrastructure can decrease the time required for researchers to perform analyses.

For large consortia such as PDXNet, metadata and secondary data types are often just as important as sequencing data for supporting impactful research. Examples of these additional data types include high-resolution images and tumor volume/drug response data. These data extend the types of problems researchers can address. We expect that the portal's image and tumor volume functionality will expand as these datasets grow. Future iterations of the portal will include interactive exploration across data types allowing users to address complex research problems.

Although PDXs have caveats as a model system, e.g. the fact that only tumors that successfully engraft can be studied, they are widely used in cancer research. To ensure their optimal use, continuing attention to standards for data submission, access, and quality is essential. The PDXNet Portal reflects PDXNet activities to implement such standards not only for sequencing data but also metadata and sec-

ondary data types. Another consideration requiring careful implementation is to balance data security versus ease of use. The PDXNet portal will grow with new data and features as the PDXNet consortium continues to generate new models. Consequently, standardized processing and batch effects are of increasing concern for downstream analyses. To ensure that researchers have confidence in the data quality, we will continue to share the informative metrics computed by the standardized PDXNet quality control workflows.

Several key features are the focus for the next iteration of the PDXNet portal. These include tools to expand our search features for commonly found genomic variants (e.g. INDELs and copy number variations) within models, diseases and genes of interest, and interactive exploration of gene expression data. These tools would enable researchers to perform meaningful analyses directly from the portal and more rapidly realize value from PDXNet data. Visualization and analysis of associated data, including imaging and tumor volume/drug response data, will also be a focus. These data types have the potential for high impact, particularly given the innovation in large scale data visualization techniques in many fields.

Further development of links between the PDXNet portal and NCI computational infrastructure will benefit researchers as well. Moving large quantities of data is time-consuming and can be expensive. Enabling researchers to perform their analysis where the data is already present lowers the entry barrier into the computational analysis of PDX model data. To facilitate these links, we envision users will be able to create cohorts for analysis using the PDXNet portal and transferring that selection to the Cancer Genomics Cloud or other computational platform, where they will be able to easily take advantage of well-developed computational infrastructure.

We will extend PDXNet portal capabilities as the size and complexity of PDXNet datasets grow. These enhancements will allow the research community to quickly find and evaluate PDXNet resources to supplement their research studies. We will continue to improve the PDXNet portal value by collaborating with related PDX initiatives, including the PDMR, PDXFinder and EuroPDX. Such collaborations will demonstrate how to effectively conduct studies across institutions, providing examples for the broader research community in how to optimize their PDX studies with respect to the public PDX models and datasets that are becoming increasingly available.

## DATA AVAILABILITY

Each PDXNet portal data tab allows users to download metadata. Data for smaller data types such as tumor volume data and computed metrics (e.g. HRD and TMB) can be downloaded directly from the portal. For larger data types, please request data from the PDXNet Portal's Contact page. We will coordinate with PDTs to make data available either directly or through dbGaP as required by the PDXNet data sharing agreement.

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Cancer Online.

## ACKNOWLEDGEMENTS

We would like to thank the patients who provided the tissues that support the PDXNet model and sequencing data generation.

## FUNDING

National Institutes of Health [NCI U24-CA224067, NCI U54-CA224083, NCI U54-CA224070, NCI U54-CA224065, NCI U54-CA224076, NCI U54-CA233223, NCI U54-CA233306]; National Cancer Institute; National Institutes of Health [HHSN261201400008C, HHSN261201500003I].

*Conflict of interest statement.* The University of Utah may choose to license PDX models developed in the Welm labs, which may result in tangible property royalties to Dr Welm and members of their lab who developed the models. Michael T. Lewis is a founder and limited partner in StemMed Ltd. and a manager in StemMed Holdings, it's general partner. He is a founder and equity stake holder in Tvardi Therapeutics Inc. Some PDX are exclusively licensed to StemMed Ltd. resulting in royalty income to MTL. Lacey Dobrolecki is a compensated employee of StemMed Ltd. The other authors declare no competing interest.

## REFERENCES

- Koga, Y. and Ochiai, A. (2019) Systematic review of patient-derived xenograft models for preclinical studies of anti-cancer drugs in solid tumors. *Cells*, **8**, 418.
- Dobrolecki, L.E., Airhart, S.D., Alferes, D.G., Aparicio, S., Behbod, F., Bentires-Alj, M., Briskin, C., Bult, C.J., Cai, S., Clarke, R.B. *et al.* (2016) Patient-derived xenograft (PDX) models in basic and translational breast cancer research. *Cancer Metastasis Rev.*, **35**, 547–573.
- Collins, A.T. and Lang, S.H. (2018) A systematic review of the validity of patient derived xenograft (PDX) models: the implications for translational research and personalised medicine. *PeerJ*, **6**, e5981.
- Brown, K.M., Xue, A., Mittal, A., Samra, J.S., Smith, R. and Hugh, T.J. (2016) Patient-derived xenograft models of colorectal cancer in pre-clinical research: a systematic review. *Oncotarget*, **7**, 66212–66225.
- Grandori, C. and Kemp, C.J. (2018) Personalized cancer models for target discovery and precision medicine. *Trends Cancer*, **4**, 634–642.
- Hidalgo, M., Amant, F., Biankin, A.V., Budinska, E., Byrne, A.T., Caldas, C., Clarke, R.B., de Jong, S., Jonkers, J., Maelandsmo, G.M. *et al.* (2014) Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.*, **4**, 998–1013.
- Woo, X.Y., Srivastava, A., Graber, J.H., Yadav, V., Sarsani, V.K., Simons, A., Beane, G., Grubb, S., Ananda, G., Liu, R. *et al.* (2019) Genomic data analysis workflows for tumors from patient-derived xenografts (PDXs): challenges and guidelines. *BMC Med. Genomics*, **12**, 92.
- Evrard, Y.A., Srivastava, A., Randjelovic, J., Doroshow, J.H., Dean, D.A. 2nd, Morris, J.S., Chuang, J.H. and NCI PDXNet Consortium. (2020) Systematic establishment of robustness and standards in patient-derived xenograft experiments and analysis. *Cancer Res.*, **80**, 2286–2297.
- Sun, H., Cao, S., Mashl, R.J., Mo, C.K., Zaccaria, S., Wendl, M.C., Davies, S.R., Bailey, M.H., Primeau, T.M., Hoog, J. *et al.* (2021) Comprehensive characterization of 536 patient-derived xenograft models prioritizes candidates for targeted treatment. *Nat. Commun.*, **12**, 5086.
- Woo, X.Y., Giordano, J., Srivastava, A., Zhao, Z.M., Lloyd, M.W., de Bruijn, R., Suh, Y.S., Patidar, R., Chen, L., Scherer, S. *et al.* (2021) Conservation of copy number profiles during engraftment and passaging of patient-derived cancer xenografts. *Nat. Genet.*, **53**, 86–99.
- NCI Patient-Derived Models Repository (2022) *Frederick National Laboratory for Cancer Research, Frederick, MD.* <https://pdmr.cancer.gov/>, (14 March 2022, date last accessed).
- Conte, N., Mason, J.C., Halmagyi, C., Neuhauser, S., Mosaku, A., Yordanova, G., Chatzipli, A., Begley, D.A., Krupke, D.M., Parkinson, H. *et al.* (2019) PDX Finder: A portal for patient-derived tumor xenograft model discovery. *Nucleic Acids Res.*, **47**, D1073–D1079.
- Meehan, T.F., Conte, N., Goldstein, T., Inghirami, G., Murakami, M.A., Brabetz, S., Gu, Z., Wiser, J.A., Dunn, P., Begley, D.A. *et al.* (2017) PDX-MI: Minimal information for patient-derived tumor xenograft models. *Cancer Res.*, **77**, e62–e66.
- Lowy, D.R. and Collins, F.S. (2016) Aiming High—changing the trajectory for cancer. *N. Engl. J. Med.*, **374**, 1901–1904.
- Sharpless, N.E. and Singer, D.S. (2021) Progress and potential: the cancer moonshot. *Cancer Cell*, **39**, 889–894.
- Lau, J.W., Lehnert, E., Sethi, A., Malhotra, R., Kaushik, G., Onder, Z., Groves-Kirkby, N., Mihajlovic, A., DiGiovanna, J., Srdic, M. *et al.* (2017) The cancer genomics cloud: Collaborative, reproducible, and democratized—a new paradigm in large-scale computational research. *Cancer Res.*, **77**, e3–e6.
- Amstutz, P., Crusoe, M.R., Tijanić, N., Chapman, B., Chilton, J., Heuer, M., Kartashov, A., Kern, J., Lehr, D., Ménager, H. *et al.* (2016) In: *Common Workflow Language, v1.0. Specification*. Common Workflow Language Working Group.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) In: *The Elements of Statistical Learning*. Springer, NY.
- Chen, C., Pollack, S., Hunter, D.J., Hirschhorn, J.N., Kraft, P. and Price, A.L. (2013) Improved ancestry inference using weights from external reference panels. *Bioinformatics*, **29**, 1399–1406.
- Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
- Weiss, K.M. and Long, J.C. (2009) Non-Darwinian estimation: my ancestors, my genes' ancestors. *Genome Res.*, **19**, 703–710.
- Sztupinski, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J., Eklund, A.C., Syd, A. and Szallasi, Z. (2018) Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer*, **4**, 16.
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, **6**, 80–92.
- Kautto, E.A., Bonneville, R., Miya, J., Yu, L., Krook, M.A., Reeser, J.W. and Roychowdhury, S. (2017) Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget*, **8**, 7452–7463.
- Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M.D., Wendl, M.C. and Ding, L. (2014) MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*, **30**, 1015–1016.
- Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y. *et al.* (2015) High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.*, **21**, 1318–1325.
- Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.

## APPENDIX

### PDXNet Consortium Members:

Matthew Bailey<sup>1</sup>, Emilio Cortes-Sanchez<sup>1</sup>, Sandra Scherer<sup>1</sup>, Chieh-Hsiang Yang<sup>1</sup>, Maihi Fujita<sup>1</sup>, Zhengtao Chu<sup>1</sup>, Ling Zhao<sup>1</sup>, Andrew Butterfield<sup>1</sup>, Argun Akcakanat<sup>2</sup>, Gao Boning<sup>2</sup>, Kurt Evans<sup>2</sup>, Bingliang Fang<sup>2</sup>, Don Gibbons<sup>2</sup>, Vanessa Jensen<sup>2</sup>, Dara Keener<sup>2</sup>, Michael Kim<sup>2</sup>, Scott Kopetz<sup>2</sup>, Mourad Majidi<sup>2</sup>, David Menter<sup>2</sup>, John



Minna <sup>2</sup>, Hyunsil Park <sup>2</sup>, Fei Yang <sup>2</sup>, Brenda Timmons <sup>2</sup>, Jing Wang <sup>2</sup>, Shannon Westin <sup>2</sup>, Timothy Yap <sup>2</sup>, Jianhua Zhang <sup>2</sup>, Ran Zhang <sup>2</sup>, Min Jin Ha <sup>2</sup>, Huiqin Chen <sup>2</sup>, Yuanxin Xi <sup>2</sup>, Luc Girard <sup>2</sup>, Erkan Yucan <sup>2</sup>, Bryce P Kirby <sup>2</sup>, Bingbing Dai <sup>2</sup>, Yi Xu <sup>2</sup>, Alexey Sorokin <sup>2</sup>, Kelly Gale <sup>2</sup>, Jithesh Augustine <sup>2</sup>, Stephen Scott <sup>2</sup>, Ismail Meraz <sup>2</sup>, Dylan Fingerman <sup>3</sup>, Andrew Kossenkov <sup>3</sup>, Qin Liu <sup>3</sup>, Min Xiao <sup>3</sup>, Jayamanna Wickramasinghe <sup>3</sup>, Haiyin Lin <sup>3</sup>, Eric Ramirez-Salazar <sup>3</sup>, Kate Nathanson <sup>4</sup>, Mike Tetzlaff <sup>2</sup>, George Xu <sup>5</sup>, Vashisht G. Yennu-Nanda <sup>6</sup>, Rebecca Aft <sup>7</sup>, Jessica Andrews <sup>7</sup>, Alicia Asaro <sup>7</sup>, Song Cao <sup>7</sup>, Feng Chen <sup>7</sup>, Sherri Davies <sup>7</sup>, John DiPersio <sup>7</sup>, Ryan Fields <sup>7</sup>, Steven Foltz <sup>7</sup>, Katherine Fuh <sup>7</sup>, Kian Lim <sup>7</sup>, Jason Held <sup>7</sup>, Jeremy Hoog <sup>7</sup>, Reyka G. Jayasinghe <sup>7</sup>, Yize Li <sup>7</sup>, Jinqin Luo <sup>7</sup>, Cynthia Ma <sup>7</sup>, Jay Mashl <sup>7</sup>, Chia-Kuei Mo <sup>7</sup>, Fernanda Rodriguez <sup>7</sup>, Hua Sun <sup>7</sup>, Nadezhda V. Terekhanova <sup>7</sup>, Rose Tipton <sup>7</sup>, Brian VanTine <sup>7</sup>, Andrea Wang-Gillam <sup>7</sup>, Mike Wendl <sup>7</sup>, Yige Wu <sup>7</sup>, Matt Wyczalkowski <sup>7</sup>, Lijun Yao <sup>7</sup>, Daniel Cui Zhou <sup>7</sup>, Matthew Ellis <sup>8</sup>, Michael Ittmann <sup>8</sup>, Susan Hilsenbeck <sup>8</sup>, Bert O'Malley <sup>8</sup>, Amanda Kirane <sup>9</sup>, May Cho <sup>9</sup>, David Gandara <sup>9</sup>, Jonathan Reiss <sup>9</sup>, Tiffany Le <sup>9</sup>, Ralph De Vere White <sup>9</sup>, Cliff Tepper <sup>9</sup>, David Cooke <sup>9</sup>, Luis Godoy <sup>9</sup>, Lisa Brown <sup>9</sup>, Marc Dall'Era <sup>9</sup>, Christopher Evans <sup>9</sup>, Rashmi Verma <sup>9</sup>, Sepideh Gholami <sup>9</sup>, David J Segal <sup>9</sup>, John Albeck <sup>9</sup>, Edward Pugh <sup>9</sup>, Susan Stewart <sup>9</sup>, David Roche <sup>9</sup>, Hongyong Zhang <sup>10</sup>, Nicole Coggins <sup>10</sup>, Ana Estrada <sup>10</sup>, Ted Toal <sup>10</sup>, Alexa Morales <sup>10</sup>, Guadalupe Polanco Echeverry <sup>10</sup>, Sienna Rocha <sup>10</sup>, Ai-Hong Ma <sup>9,11</sup>

- 1 Huntsman Cancer Institute, Salt Lake City, UT, USA
- 2 The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA
- 3 The Wistar Institute, Philadelphia, PA, USA
- 4 Abramson Cancer Center, The University of Pennsylvania, Philadelphia, PA, USA
- 5 Hospital of the University of Pennsylvania, Philadelphia, PA, USA
- 6 Department of Melanoma Medical Oncology, MD Anderson Cancer Center, Houston, TX, USA
- 7 Washington University School of Medicine, St. Louis, MO, USA
- 8 Baylor College of Medicine, Houston, TX, USA
- 9 UC Davis Comprehensive Cancer Center, University of California – Davis, Davis, CA, USA
- 10 UC Davis Comprehensive Genomics Center, University of California – Davis, Davis, CA, USA
- 11 Jamaica Plains VA Medical Center, Veteran Administration, Jamaica Plains, MA, USA