PDX Finder: A Portal for Patient-Derived tumor Xenograft Model Discovery

Authors: Nathalie Conte[1], Jeremy Mason[1], Csaba Halmagyi[1], Steven Neuhauser[2], Abayomi Mosaku[1], D.A. Begley[2], D.M. Krupke[2], Helen Parkinson[1], Terrence F. Meehan[1], Carol Bult[2]

1 European Molecular Biology Laboratory- European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

2 The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

ABSTRACT

Patient-derived tumor xenograft (PDX) mouse models are a versatile oncology research platform for studying tumor biology and for testing chemotherapeutic approaches tailored to genomic characteristics of individual patient's tumors. PDX models are generated and distributed by a diverse group of academic labs, research organizations, multi-institution consortia, and contract research organizations. The distributed nature of PDX repositories and the use of different standards in the associated metadata presents a significant challenge to finding PDX models relevant to specific cancer research questions. The Jackson Laboratory and EMBL-EBI are addressing these challenges by co-developing PDX Finder, a comprehensive open global catalog of PDX models and their associated datasets. Within PDX Finder, model attributes are harmonized and integrated using a previously developed community minimal information standard to support consistent searching across the originating resources. Links to repositories are provided from the PDX Finder search results to facilitate model acquisition and/or collaboration. The PDX Finder resource currently contains information for more than 1900 PDX models of diverse cancers including those from large resources such as the Patient-Derived Models Repository, PDXNet and EurOPDX. Individuals or organizations that generate and distribute PDXs are invited to increase the "findability" of their models by participating in the PDX Finder initiative at www.pdxfinder.org

INTRODUCTION

PDX models recapitulate many of the disease hallmarks of cancer patients and are increasingly being used to study therapeutics, tumor evolution and drug resistance mechanisms. PDX

models are typically generated by the implantation of human tumor tissues or cells into severely immunodeficient mouse hosts. Tumors that engraft successfully are passaged further to generate cohorts of tumor-bearing mice for experimental studies. PDX models are generated and used by researchers in university, clinical, and pharmaceutical industry settings as well as specialized commercial organizations. International consortia focusing on the use of PDX models including PDX-Net and EurOPDX have recently been funded ensuring PDX models will continue to be an important contributor to understanding and treating cancer.

The distributed and diverse nature of PDX repositories as well as differences in associated metadata describing the models presents a significant challenge to researchers seeking to find PDXs that are relevant to specific cancer research questions. To address this issue, The Jackson Laboratory and the European Molecular Biological Laboratory-European Bioinformatics Institute (EMBL-EBI) have implemented PDX Finder (http://www.pdxfinder.org), a freely available and searchable catalog of global PDX repositories. The data model for PDX Finder is based on the minimal information standard for PDX models developed in collaboration with a broad range of stakeholders who create and/or use PDX models in basic and pre-clinical cancer research (1). PDX Finder currently provides access to information about more than 1900 PDX models in 7 repositories around the world, including NCI's Patient Derived Model Repository, The Jackson Laboratory's PDX Resource, members of the EurOPDX Consortium, and members of NCI's recently launched PDXNet. Here we describe the implementation of PDX Finder and illustrate how it can be used to locate relevant PDX models. We also provide information on how investigators with small or large repositories of PDX models can have their resource indexed in PDX Finder to improve the visibility of their models.

PDX FINDER FEATURES AND FUNCTIONALITY

PDX Finder was designed and implemented following a series of workshops and on-line surveys aimed to identify the diverse needs of researchers who use PDX models. The three top-rated functions/interfaces to emerge from the requirements gathering process were:

1. Availability of high level summaries (graphical and tabular) that provides an overview of all the models and repositories included in PDX Finder.
2. Search forms that allow researchers to find PDX models based on cancer type, diagnosis, availability of specific datasets, molecular markers and results from drug resistance/sensitivity dosing studies.

3. Summary pages for Individual models that includes salient details about how the model was generated and links to available data sets and model acquisition information.

PDX Finder home page: The PDX Finder home page (Figure 1) prominently features search capabilities allowing users to perform a lexical search over a number of categories and summary graphics of the resource. The number of PDX models are represented in a list of models by cancer types as well as a pie chart grouping models by anatomical systems. A list of all model providers is also featured. The top menu bar provides details about the PDX Finder project, including our information gathering process, contact information, and instructions on how to submit PDX model information to the resource. A "News and Events" section on the bottom of page provides headline summary and links describing news about with PDX resource as well as a Twitter feed where users can follow to be updated on the latest developments.

Search functions: Users navigate to the "Search Results" page by searching for a term or choosing a category from one of the graphical visualizations. Results are presented in a tabular format with each row depicting a model (Figure 2). Key features of the models are presented in columns and include model ID, critical patient clinical data, tumor histological classification, links to available datasets and the original provider. Multiple filters provided on the left margin of the page allow users to exactly specify the models they would like to appear in the results table. Filters are selected by expanding on a category and selecting the relevant sub-categories with the ability to choose multiple filters. The search functionality allows users to narrow model availability by specific criteria such as "Find all colorectal cancer models with BRAF V600E mutation" will retrieve 31 models over 2 different sources. For searches that return multiple models, the results can be exported in a tabular format. Users can navigate to a model page or to the data of interest by clicking on the unique PDX Model ID or data links in each row.

Summary pages: The PDX Model Detail page (Figure 3) presents key features about the model. The top of the page displays the model ID and tumor histological classification as well as prominent links to the originating resource where users can find more information and contact the relevant institution for further collaboration. A tabulated section beneath the overview provides summary views of the models with clinical, model, and validation information as well as views of differing types of data submitted with from the PDX model such as genomic or dosing studies when provided by the source.

NOMENCLATURE AND METADATA STANDARDS

PDX Finder facilitates discovery of similar models across resources by the use and enforcement of nomenclature and metadata standards. Cancer type, diagnosis, and other cancer attributes are represented by the NCI Thesaurus, a community resource that is actively maintained by the NCI (2). Names and symbols for human genes follow the nomenclature standards approved by the HUGO Gene Nomenclature Committee (3). Host mouse strain nomenclature follows the official guidelines established by the International Committee on Standardized Genetic Nomenclature for Mice (4). Drug/compound names use standards from CHEBI (5), CHEMBL (6) and PubChem (7). In addition, we provide feedback to ontology developers to improve these resources for representing the complexities of PDX models.

DATABASE AND SOFTWARE INFRASTRUCTURE

The PDX Finder web site was implemented using a combination of the WordPress content management system and a Java web application. The database for PDX Finder is an instance of a Neo4J graph database. To populate this database, bespoke Extraction, Transformation, Loading (ETL) software pipelines were written in Java to extract relevant attributes corresponding to the PDX minimal information standard from the data provided by the different PDX repositories. Software developed by the PDX Finder team is freely available under an Apache 2 license (https://www.apache.org/licenses/LICENSE-2.0). The PDX Finder source code is available at GitHub (https://github.com/pdxfinder).

A number of key attributes specified in the minimal information standard rely on different vocabularies and ontologies. To address this challenging harmonization issue, PDX Finder employs a semi-automated integration approach using resources at EMBL- EBI including the ZOOMA annotation tool (8) that maps free text annotations to ontology terms based on a curated repositories of annotation knowledge. Transformed metadata is reviewed by local experts and reported to the submitters for validation and approval. Once approved, the transformed metadata describing a model is then loaded into the PDX Finder database and exposed to the users through the web interface.

An example of metadata harmonization includes mapping biologically identical histological concepts harboring different names provided by several sources using their own classification. To harmonize those concepts and support consistent searching across the originating resources, we use different attributes like the original histological term and the primary tissue provided by the source to refine the subtype. Examples of harmonization include mapping source specific histological concepts ("Adenosquamous", "adenosquamous carcinoma" , "Ad and SC carcinoma") that share the same primary tissue ("lung") to the NCIT ontological label "Adenosquamous Lung Carcinoma". Furthermore, ontological association allows aggregation of concepts based on meaningful groupings like cancer by anatomical system or cell morphology. This allows users to search for 'Lung Cancer' models across all subclasses of Lung Cancer models in a single query, without having to look for each subtype individually.

FUTURE DIRECTIONS

PDX Finder is a freely available global catalog of PDX models and associated data available from independent, distributed repositories. Future development of PDX Finder will focus on three areas: addition of new PDX repositories, coordination with other informatics groups, and implementation of user-requested functionality.

Adding new PDX repositories: We will increase the number of PDX models represented in the resource by contacting PDX collection owners. To support scalability of data submission and processing to PDX Finder, we will develop new simplified data submission tools to encourage data submission via spreadsheets or via APIs. The existing ETL modules will be improved by addition of curated knowledge bases that improve the automated data mappings as well as interfaces that facilitate review of the resulting data integration. We will also leverage EBI's Ontology Xref service (OxO; https://www.ebi.ac.uk/spot/oxo/) that captures community defined cross references between semantic standards in our ETL modules such as mappings between MeSH AND NCI Thesaurus (9).

Coordination with other informatics resources: To facilitate data sustainability, maximize use of existing data and avoid redundancy, the PDX Finder team has initiated coordinating activities with existing molecular archives to deposit data generated from PDX Models. The data loading process will have oversight by a PDX Data Coordinator who will broker submission of data to established archives in collaboration with the data providers. Data that have the potential to

allow identification of individuals will be submitted to relevant secure molecular archives such as the European Genome-Phenome Archive (10) or dbGAP (11). These resources provide secure environments where the approval of a submitter defined Data Access Committee is required for data access. Depending on the data type, non-patient identifiable data will be submitted to other archive resources including the Sequence Read Archive (12), the European Nucleotide Archive (13) for nucleotide sequencing information; Gene Expression Omnibus (14) or ArrayExpress (15) for gene expression data or European Variation Archive (https://www.ebi.ac.uk/eva/) for genetic variation information. PDX Finder will also coordinate the submission of sample metadata to BioSamples database (16), or BioSample archive (17), which provides unique identifiers that are used to link varying types of data derived from the same sample. Integration of the PDX Finder datasets linked to their harmonized data in molecular archives will ensure that meta-analysis type of studies combine biologically comparable datasets from multiple sources.

New functionality: Other areas of development will be determined by the needs of the PDX research community. We will continue to assess user needs by surveys and testing of the portal. As the number of informatics resources using PDX models grows, the community is best served by sharing solutions to common challenges. The developers of PDX Finder are providing software components to the PDXNet and the EurOPDX data platforms, which is facilitating knowledge exchange and possible reuse of informatics tools. PDX Finder is also funded in part by NCI's ITCR program, a trans-NCI initiative to promote integration of informatics technology development and reuse of software for oncology researchers. We will be working with developers of other ITCR tools to reuse software developed by others to address common needs.

COMMUNITY OUTREACH AND USER SUPPORT

The PDX Finder user help desk is available via email and user outreach is supported through Twitter and the PDX Finder listserv. Researchers interested in listing their PDX repository in PDX Finder should contact the team using the helpdesk email provided below. PDX Finder is available to any group, organization from academia or industry that seeks to openly distribute PDX models and associated resources and information.

- Email access: helpdesk@pdxfinder.org
- Twitter: https://twitter.com/PDXFinder

- PDX Finder Listserve: https://listserver.ebi.ac.uk/mailman/listinfo/pdxfinder-announce

CITING PDX FINDER

For a general citation of PDX Finder, researchers should cite this article. In addition, the following citation format is suggested when referring to specific data about PDX models obtained from the PDX Finder web site: PDX Finder (http://www.pdxfinder.org); data retrieved in March 2018.

ACKNOWLEDGEMENTS

The authors thank the PDX resources who have contributed PDX metadata and information to PDX Finder, including members of PDXNet (https://www.pdxnetwork.org/), members of EurOPDX (http://europdx.eu/), The Jackson Laboratory PDX Resource (http://tumor.informatics.jax.org/mtbwi/pdxSearch.do), and NCI's PDMR (https://pdmr.cancer.gov/). We are particularly grateful to our colleagues, Drs Andrea Bertotti, Enzo Medico (University of Torino, Italy), Aikaterina Chatzipli, Ultan McDermott (Wellcome Sanger Institute, UK) and Yvonne Evrard (NCI Patient-Derived Models Repository, USA) who have provided their expertise and guidance during all stages of the work.

FUNDING

PDX Finder is funded by NIH/NCI R01 CA089713 (CJB) and NIH/NCI U24 CA204781 (HP and TM). Funding for open access charge: NIH/NCI R01 CA089713. Conflict of interest statement: None declared.

LITERATURE CITED

1. 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.
https://doi.org/10.1158/0008-5472.CAN-17-0582
http://www.ncbi.nlm.nih.gov/pubmed/29092942

2. de Coronado,S., Wright,L.W., Fragoso,G., Haber,M.W., Hahn-Dantona,E.A., Hartel,F.W., Quan,S.L., Safran,T., Thomas,N. and Whiteman,L. (2009) The NCI Thesaurus quality assurance life cycle. *J. Biomed. Inform.*, **42**, 530–539.
https://doi.org/10.1016/j.jbi.2009.01.003
http://www.ncbi.nlm.nih.gov/pubmed/19475726

3. Gray,K.A., Yates,B., Seal,R.L., Wright,M.W. and Bruford,E.A. (2015) Genenames.org: The HGNC resources in 2015. *Nucleic Acids Res.*, **43**, D1079–D1085.
https://doi.org/10.1093/nar/gku1071
http://www.ncbi.nlm.nih.gov/pubmed/23161694

4. Davisson,M.T. (1997) Rules and guidelines for genetic nomenclature in mice: excerpted version. Committee on Standardized Genetic Nomenclature for Mice. *Transgenic Res.*, **6**, 309–19.
http://www.ncbi.nlm.nih.gov/pubmed/9322368

5. Hastings,J., De Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N., Muthukrishnan,V., Owen,G., Turner,S., Williams,M., *et al.* (2013) The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. *Nucleic Acids Res.*, **41**.
https://doi.org/10.1093/nar/gks1146
http://www.ncbi.nlm.nih.gov/pubmed/23180789

6. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Krüger,F.A., Light,Y., Mak,L., McGlinchey,S., *et al.* (2014) The ChEMBL bioactivity database: An update. *Nucleic Acids Res.*, **42**.
https://doi.org/10.1093/nar/gkt1031
http://www.ncbi.nlm.nih.gov/pubmed/24214965

7. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A., *et al.* (2016) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
https://doi.org/10.1093/nar/gkv951
http://www.ncbi.nlm.nih.gov/pubmed/26400175

8. Cook,C.E., Bergman,M.T., Cochrane,G., Apweiler,R. and Birney,E. (2018) The European Bioinformatics Institute in 2017: data coordination and integration. *Nucleic Acids Res.*, **46**, D21–D29.
https://doi.org/10.1093/nar/gkx1154
http://www.ncbi.nlm.nih.gov/pubmed/29186510

9. Ong,E., Sarntivijai,S., Jupp,S., Parkinson,H. and He,Y. (2017) Comparison, alignment, and synchronization of cell line information between CLO and EFO. *BMC Bioinformatics*, **18**.
https://doi.org/10.1186/s12859-017-1979-z
http://www.ncbi.nlm.nih.gov/pubmed/29322915

10. Lappalainen,I., Almeida-King,J., Kumanduri,V., Senf,A., Spalding,J.D., Ur-Rehman,S., Saunders,G., Kandasamy,J., Caccamo,M., Leinonen,R., *et al.* (2015) The European Genome-phenome Archive of human data consented for biomedical research. *Nat. Genet.*, **47**, 692–695.
https://doi.org/10.1038/ng.3312
http://www.ncbi.nlm.nih.gov/pubmed/26111507

11. Wong,K.M., Langlais,K., Tobias,G.S., Fletcher-Hoppe,C., Krasnewich,D., Leeds,H.S., Rodriguez,L.L., Godynskiy,G., Schneider,V.A., Ramos,E.M., *et al.* (2017) The dbGaP data browser: a new tool for browsing dbGaP controlled-access genomic data. *Nucleic Acids Res.*, **45**, D819–D826.

https://doi.org/10.1093/nar/gkw1139

12. Leinonen,R., Sugawara,H., Shumway,M. and International Nucleotide Sequence Database Collaboration (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19-21.
https://doi.org/10.1093/nar/gkq1019
http://www.ncbi.nlm.nih.gov/pubmed/21062823

13. Silvester,N., Alako,B., Amid,C., Cerdeño-Tarrága,A., Clarke,L., Cleland,I., Harrison,P.W., Jayathilaka,S., Kay,S., Keane,T., *et al.* (2018) The European Nucleotide Archive in 2017. *Nucleic Acids Res.*, **46**, D36–D40.
https://doi.org/10.1093/nar/gkx1125

14. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M., *et al.* (2012) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
https://doi.org/10.1093/nar/gks1193

15. Kolesnikov,N., Hastings,E., Keays,M., Melnichuk,O., Tang,Y.A., Williams,E., Dylag,M., Kurbatova,N., Brandizi,M., Burdett,T., *et al.* (2015) ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
https://doi.org/10.1093/nar/gku1057

16. Faulconbridge,A., Burdett,T., Brandizi,M., Gostev,M., Pereira,R., Vasant,D., Sarkans,U., Brazma,A. and Parkinson,H. (2014) Updates to BioSamples database at European Bioinformatics Institute. *Nucleic Acids Res.*, **42**, D50–D52.
https://doi.org/10.1093/nar/gkt1081

17. Barrett,T., Clark,K., Gevorgyan,R., Gorelenkov,V., Gribov,E., Karsch-Mizrachi,I., Kimelman,M., Pruitt,K.D., Resenchuk,S., Tatusova,T., *et al.* (2012) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.*, **40**, D57.
https://doi.org/10.1093/NAR/GKR1163
http://www.ncbi.nlm.nih.gov/pubmed/22139929

FIGURE LEGENDS

Figure 1. PDX Finder Homepage

Figure 2. PDX Model Search page

Figure 3. PDX Model Detail page

PDX FINDER

HOME ABOUT SEARCH CONTACT

PDX FINDER IS YOUR OPEN GLOBAL CANCER RESEARCH PORTAL TO PATIENT DERIVED XENOGRAFT MODELS

FIND

By cancer diagnosis

Search (e.g., 'Melanoma')

PDX Finder contains 1910 PDX models

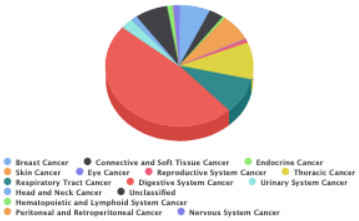
EXPLORE

BY EXACT DIAGNOSIS

639 | Colorectal Carcinoma
207 | Colorectal Adenocarcinoma
141 | Melanoma
114 | Cutaneous Melanoma

View All Cancer Types

BY ANATOMICAL/BIOLOGICAL SYSTEM



BY SOURCE

639 | IRCC
428 | JAX
316 | PDXNet-MDAnderson
256 | PDXNet-Wistar-MDAnderson-Penn
119 | PDXNet-WUSTL
101 | PDMR
51 | PDXNet-HCI-BCM

NEWS AND EVENTS

LATEST NEWS

5th February 2018
EDIREX KICK OFF MEETING IN TURIN ITALY 5-7TH MARCH 2018

EVENTS

5th February 2018
AACR 2018 MEETING

Hmm, an empty timeline. That's weird.

Check for Tweets

QUICK LINKS

How to cite PDX Finder? (A bioarchive preprint)
PDX Minimum Information standard publication
PDX Finder Terms of Use

CONTACTS

Submit models
Feedback

EMBL-EBI and The Jackson Laboratory are co-developers of PDX Finder. This work is supported by the National Institutes of Health U24 CA204781 01 and the National Cancer Institute MTB R01 R01CA089713.

NIH NATIONAL CANCER INSTITUTE

PDX FINDER

HOME ABOUT SEARCH CONTACT

HOME / SEARCH

FIND

Search (e.g., Melanoma)

PDX Finder contains 1910 models

FILTER BY:

PATIENT / TUMOR

CANCER BY SYSTEM

TUMOR TYPE

SEX

AGE

PDX MODEL

DATASOURCE

MOLECULAR CHARACTERIZATION

SMALL GENETIC VARIATION

MODEL	HISTOLOGY	PRIMARY SITE	COLLECTION SITE	TYPE	AGE	SEX	DATA AVAILABLE
CRC0022LM IRCC	Colorectal Carcinoma	Sigmoid Colon	Liver	Metastatic	70-79	Male	Mutation_TargetedNGS_MUT Dosing Studies
CRC0307LM IRCC	Colorectal Carcinoma	Sigmoid Colon	Liver	Metastatic	60-69	Male	Mutation_TargetedNGS_MUT Dosing Studies
CRC0468PR IRCC	Colorectal Carcinoma	Rectum	Rectum	Primary	50-59	Female	Mutation_TargetedNGS_MUT
WISTAR WM4352 PDXNet-Wistar-MDAnderson-Penn	Cutaneous Melanoma	Not Specified	Not Specified	Metastatic	50-59	Male	
MDA TC-479 PDXNet-MDAnderson	Adenocarcinoma	Lung	Liver	Metastatic	60-69	Male	
TM01212 JAX	Pancreatic Adenocarcinoma	Pancreas	Pancreas	Not Specified	Not Specified	Male	Mutation_CTP
WISTAR WM3991 PDXNet-Wistar-MDAnderson-Penn	Mucosal Melanoma	Not Specified	Not Specified	Primary	50-59	Female	
TM00697 JAX	Synovial Sarcoma	Soft tissue	Lung	Metastatic	20-29	Female	Mutation_Truseq_JAX
TM00887 JAX	Colon Adenocarcinoma	Colon	Liver	Metastatic	Not Specified	Male	Mutation_Truseq_JAX
CRC0207LM IRCC	Colorectal Carcinoma	Sigmoid Colon	Liver	Metastatic	50-59	Female	Mutation_TargetedNGS_MUT

« First 1 2 3 4 5 6 7 8 Last »

Export Results

TM00107: Invasive Ductal Carcinoma Not Otherwise Specified

The Jackson Laboratory

CONTACT PROVIDER

VIEW DATA AT JAX

SUMMARY | GENOMIC DATA (2991) | DOSING STUDIES (1)

PATIENT	
Sex	Female
Age at Time of Collection	48
Race / Ethnicity	White / Not Hispanic or Latino

PATIENT TUMOR	
Histology	Invasive Ductal Carcinoma Not Otherwise Specified
Primary Tissue	Breast
Collection Site	Breast
Tumor Type	Recurrent
Stage / Grade	AJCC IV/Grade:3

PDX MODEL	
Strain	NOD scid gamma
Engraftment Site	Subcutaneous
Implantation Type	Not Specified

MODEL QUALITY CONTROL	
Technology	Fingerprint
Description	Not Specified
Passages	Not Specified