



An Open Access Database of Licensed Cancer Drugs

Pan Pantziarka^{1,2*}, Rica Capistrano I¹, Arno De Potter³, Liese Vandeborne¹ and Gauthier Bouche¹

¹The Anticancer Fund, Brussels, Belgium, ²The George Pantziarka TP53 Trust, London, United Kingdom, ³Faculty of Medicine, University of Leuven, Leuven, Belgium

A global, comprehensive and open access listing of approved anticancer drugs does not currently exist. Partial information is available from multiple sources, including regulatory authorities, national formularies and scientific agencies. Many such data sources include drugs used in oncology for supportive care, diagnostic or other non-antineoplastic uses. We describe a methodology to combine and cleanse relevant data from multiple sources to produce an open access database of drugs licensed specifically for therapeutic antineoplastic purposes. The resulting list is provided as an open access database, (<http://www.redo-project.org/cancer-drugs-db/>), so that it may be used by researchers as input for further research projects, for example literature-based text mining for drug repurposing.

OPEN ACCESS

Keywords: drug repurposing, drug licensing, antineoplastic drugs, licensed drugs, database, list of cancer drugs

Edited by:

Claudia Cerella,
Fondation De Recherche Cancer Et
Sang, Luxembourg

Reviewed by:

Xiangqian Guo,
Henan University, China
Jing Tang,
University of Helsinki, Finland

*Correspondence:

Pan Pantziarka
anticancer.org.uk@gmail.com

Specialty section:

This article was submitted to
Pharmacology of Anti-Cancer Drugs,
a section of the journal
Frontiers in Pharmacology

Received: 09 November 2020

Accepted: 26 January 2021

Published: 11 March 2021

Citation:

Pantziarka P, Capistrano I R,
De Potter A, Vandeborne L and
Bouche G (2021) An Open Access
Database of Licensed Cancer Drugs.
Front. Pharmacol. 12:627574.
doi: 10.3389/fphar.2021.627574

INTRODUCTION

The availability of curated datasets across diverse areas of medical research provides input to machine learning algorithms, natural language processing and the data mining of published scientific literature. In our own work on drug repurposing in oncology, our literature-based methodology relies heavily on database tables of drug repurposing candidates, molecular targets and pathways, the World Health Organization Essential Medicines List and so on (Pantziarka et al., 2018). A key input to this process is the list of drug candidates. For ‘hard repurposing’ projects, in which we are interested only in the potential of non-cancer drugs to be repurposed as cancer therapeutics, the ReDO_DB list provides a starting point for building a dataset for literature mining. But repurposing exists on a spectrum, with ‘soft’ repurposing representing the strategy of testing existing, licensed anticancer medications in cancers for which they are not currently approved (Pantziarka et al., 2020a). For a number of cancer-specific projects all forms of repurposing candidates, including cancer drugs that could be used outside of their licensed indications, needed to be included. In these cases we have not been able to identify a curated list of cancer drugs analogous to the ReDO list of repurposing candidates.

A similar issue arose in connection with automated searches of clinical trial registries to build datasets of clinical trials for manual assessment. In some cases these assessments are part of the process of identifying repurposing candidates for specific cancers, in other cases these searches are undertaken to support cancer patients looking for new therapeutic options. In both types of cases the unavailability of a curated list of licensed cancer drugs is an impediment.

Graph based methods are increasingly being used in the drug discovery process, including for repurposing (Gramatica et al., 2014). One recent example used such methods to search for new therapeutic avenues to explore in diffuse intrinsic pontine glioma (MacLean et al., 2020). Here too the availability of curated data enables rich networks to be developed in which multiple data sources

can be used to construct models which embed implicit knowledge and new connections. Knowledge of which drugs are licensed as cancer medications can of course be discovered using such models. However, having a curated list reduces the need to do this and can lead to model simplification.

Given the need outlined above, a number of criteria for inclusion of a drug in the list emerge:

- The drug is approved in the treatment of one or more malignancies by one or more regulatory agency (e.g. EMA or FDA)
- The drug is used for anticancer effects rather than for supportive care, diagnostic use or for cancer-related comorbidities

These simple criteria therefore exclude medicines commonly used in supportive care (analgesics, anti-emetics etc), and, more importantly, drugs which are in clinical development and which are therefore not approved. This includes drugs in late-stage development, including those in phase III randomized controlled trials.

There remain questions as to what constitutes a drug—is it the active pharmaceutical ingredient (API) or a specific drug product? How should different formulations of an API or drug be handled? Do products that include a number of anticancer APIs count as one or many drugs? Our starting point is that regulators grant approval to drug products rather than APIs. Secondly, we are interested in the generic names of the drugs rather than the commercial trade names which are used to differentiate competing products or which may be market-specific. In most, but not all, cases the generic drug name is the same as the API name. Therefore, we have adopted a pragmatic approach:

- We are interested in licensed drug products rather than in commercial products—in other words cyclophosphamide or nivolumab rather than Cytoxan or Opdivo
- Different pharmaceutical forms and strengths of the same drug (i.e., oral tablet or solution for intravenous injection) are not included as separate entries
- Different commercial versions of the same drug (i.e., Doxil or Lipidox) are not included as separate entries
- Different formulations of the same drug, for example paclitaxel and nab-paclitaxel, are counted as separate medicines as the licensing and approval for specific medical indications are different

For each drug entry we also list the anticancer APIs. For example paclitaxel and nab-paclitaxel are separate entries but they feature the same API. Other APIs included in a drug to alter bioavailability or enhance response, for example hyaluronidase in trastuzumab/hyaluronidase, are not listed with the anticancer APIs although they may be included in the drug product.

Here, we describe the methodology we have adopted to generate a new database of licensed anticancer drugs that we have published in an open access format online. We also highlight a number of existing data sources which are relevant to the problem we are solving, but which do not provide a complete solution to it.

METHODOLOGY

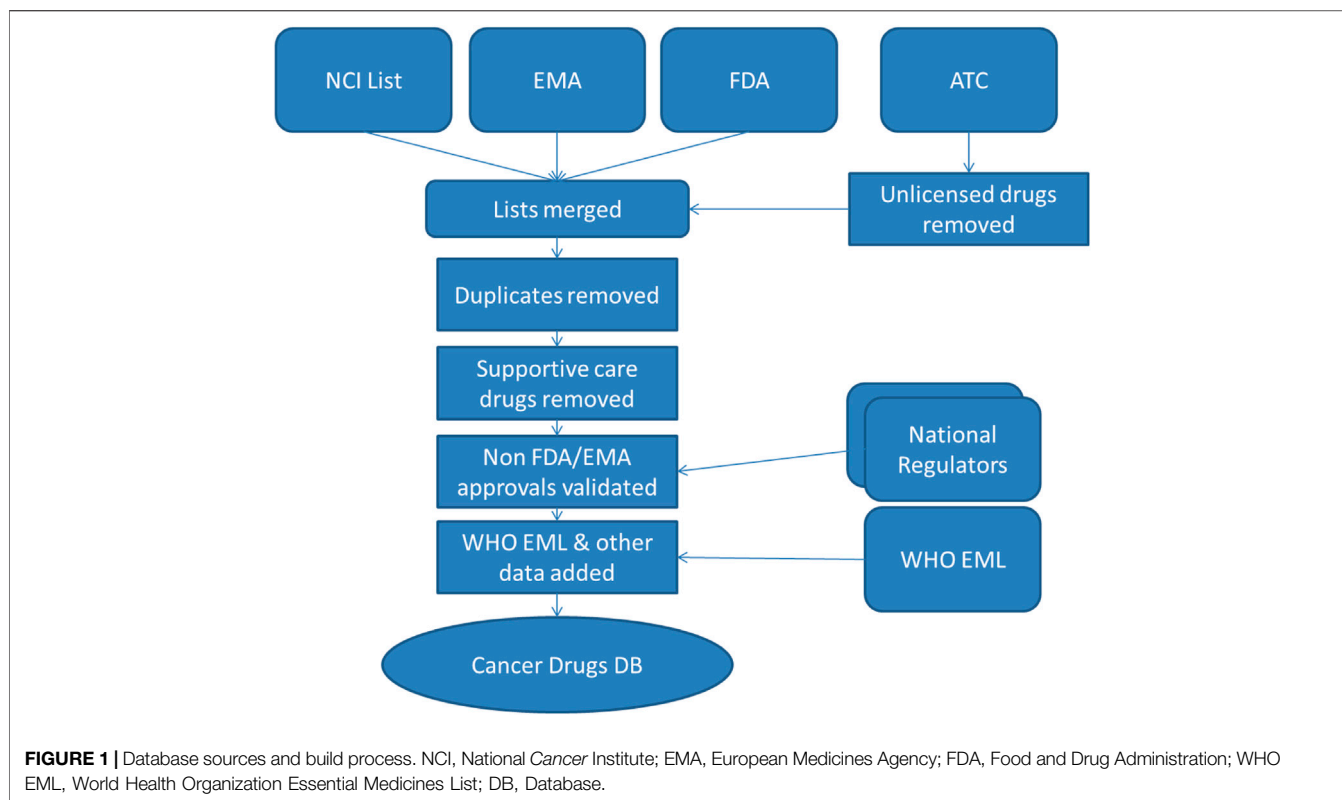
An initial list of drugs used in oncology was created by extracting the list of cancer drugs from the NCI website (<https://www.cancer.gov/about-cancer/treatment/drugs>) on February 10, 2020, and combined with the informal lists we have previously developed for use in our repurposing projects. The NCI list contains drugs routinely used in cancer care for non-anti-neoplastic activity, for example anti-emetics for chemotherapy-induced nausea and vomiting. These drugs were filtered out to leave only the anticancer agents. A list of EMA-approved drugs was derived from the European Public Assessment Reports (EPARs) published on the EMA website (<https://www.ema.europa.eu/en/medicines/download-medicine-data>) on February 19, 2020. The list was filtered to remove non-oncology drugs and those not approved for human use. FDA approvals were cross-referenced with the Drugs@FDA website (<https://www.accessdata.fda.gov/scripts/cder/daf/>).

An additional source of drugs came from the World Health Organization (WHO) Anatomical Therapeutic Chemical (ATC) Classification System (<https://www.whocc.no/>). A downloadable version of the most recent list (August 3, 2020) is available for download at the Kyoto Encyclopedia of Genes and Genomes (KEGG) website (https://www.genome.jp/kegg-bin/get_htext?br08303.keg). All drugs in the ‘L Antineoplastic and Immunomodulating Agents’ class were extracted and cross-referenced with the EMA and NCI lists in order to identify and remove duplicates. The remaining drugs were manually assessed to identify those which are licensed anticancer medications and all others were excluded.

The drugs from these different sources were merged into a single list and duplicates identified and removed, along with any remaining non-antineoplastic agents. All drugs were identified by their International Non-proprietary Name (INN). For drugs lacking FDA or EMA approval we noted whether they were approved in Europe (from the period preceding the mandatory EMA centralized approval for cancer drugs) or elsewhere. In some cases this included ‘classical’ cytotoxic drugs such as doxorubicin and cyclophosphamide which are widely used in Europe but which were approved before the EMA central licensing process was instituted. These drugs were flagged as having European ‘national approval’. In a few other cases, such as the immunotherapeutic monoclonal antibody racotumomab, the drug is only approved in a few countries, in this case Cuba and Argentina. In such cases these approvals were listed explicitly after independent verification via the relevant regulatory agency website.

Finally, the anticancer API(s) of each drug was also identified, in the majority of cases this being the same as the drug INN. Each entry in the consolidated drug list was then cross-referenced with WHO Essential Medicines List (WHO EML) to identify those which are included in that list.

This process of data merging and cleansing, illustrated in **Figure 1**, has been performed iteratively and the list has been updated on December 21, 2020 to reflect recent drug approvals. A proactive approach has been adopted wherein monthly searches of relevant EMA and FDA publications listing drug approvals are manually scanned to identify new drugs which fit the criteria for inclusion in the database. It is also intended that there will be a



twice-yearly scan to identify all drugs in the database which have been withdrawn.

Data on the medical uses for drugs is derived from the DrugBank database (Wishart et al., 2018). All medical indications are downloaded from DrugBank and filtered off-line to exclude non-cancer indications before being incorporated into the *Cancer Drugs* database. Drug target information is downloaded from the Drug-Target Interaction database (DGIdb), which integrates target information from multiple sources, including DrugBank, the Drug Target Commons (Tang et al., 2018), PharmGKB (Thorn et al., 2013) and other databases (Cotto et al., 2017).

The patent status of a drug is a key consideration when selecting prospective repurposing candidates—for example in terms of support for clinical trials from manufacturers (Pantziarka et al., 2020b). A useful proxy for patent status is the year of first approval for a drug—a drug first approved in 1980, for example, is beyond the patent protection and drug exclusivity periods and is therefore unlikely to gain commercial sponsorship. Information on year of first approval is derived from a number of sources: KEGG (Kanehisa et al., 2010), DrugCentral (Avram et al., 2021) and the ChEMBL database (Cook et al., 2018). Where there are differences in first year approval date, for example a drug may be approved by FDA prior to EMA approval, the earliest approval year is used.

RESULTS

As of January 13, 2021 we have identified 313 drugs, of which 41 were approved for supportive care and two for cancer diagnostic

purposes only. The total number of licensed anticancer drugs was 270, of which 243 (90%) were approved by the FDA, 168 (62%) had EMA approval and 50 (19%) were approved for use in different European countries through national approvals only. A further nine drugs (3%) had approvals in other parts of the world but no approval from the FDA, EMA or national agencies within Europe.

The vast majority of drugs, 268 (99%), have a single anticancer API, and only two drugs contain a combination of anticancer APIs.

Finally, 52 (19%) are included in the WHO EML.

The full list of drugs is shown in **Supplementary Appendix**.

First approval year data is available for 257 (95%) of the 270 cancer drugs. The number of drugs approved per decade is shown in **Figure 2**.

There is drug target data for 257 (95%) of cancer drugs, with the total number of unique drug targets being 963.

DISCUSSION

To our knowledge, we have generated the first comprehensive list of approved anticancer drugs. This list is available at <http://www.redo-project.org/cancer-drugs-db/> and will be updated regularly to include new approvals and to exclude drugs that are no longer available. This open access list can be used as a source for multiple purposes, such as cancer drug meta-research or soft drug repurposing efforts.

This manually curated listing of anticancer drugs has a number of strengths and weaknesses. To our knowledge this is

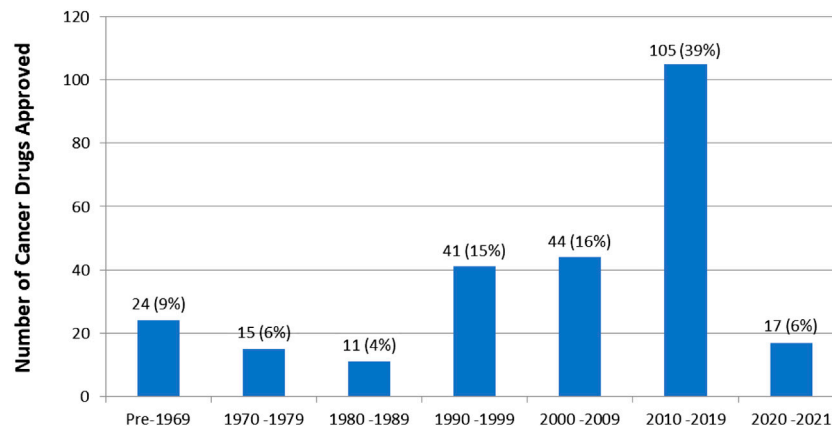


FIGURE 2 | Number of drug approvals per decade. Note that this is the year of first approval of the drug regardless of indication. Some drugs, such as dexamethasone, were first approved for non-cancer indications.

the only open access listing of medicines specifically approved for oncological use. While there are many partial listings of such drugs they are limited in that they have national or regional biases (e.g., NCI or EMA data), they often include drugs for diagnostics or supportive care and many include drugs which have yet to be approved (i.e., drugs still in development). Additionally, our process has removed duplicate entries and identified combination products.

An additional strength is that the listing is being published in an open access format, with a machine-readable download option to facilitate use of the data in further research work. For example, to facilitate the identification of new ‘soft repurposing’ opportunities which may have been previously over-looked.

There are a number of weaknesses also, the most obvious being the lack of information on drugs approved outside of Europe and the United States. While we have captured some of those instances (e.g. racotumomab approved in Cuba and *Argentina* or belotecan in South Korea), our methodology did not allow us to capture such cases systematically. This is particularly problematic for drugs only approved in countries with strong national drug development efforts. Japan has been a point in case for decades (Fujiwara, 1999), and more recently, and importantly for our study, the People’s Republic of China has also become very prominent (Wu et al., 2015). The National Medical Products Administration (NMPA), the Chinese drugs regulator, does not provide an English-language database or downloadable listing of approved medications. A number of drugs, for example the tyrosine kinase inhibitor icotinib, are reported by websites such as Wikipedia to have NMPA approval for cancer treatment, but without official verification this cannot be confirmed. According to the English-language version of the NMPA website a database is currently being implemented—when this is online the relevant data will be incorporated into our database.

A number of drugs which are known to be used off-label as cancer treatments have not been included. Zoledronic acid, for example, is used on-licence to prevent bone-related events in advanced cancer and is therefore classed as a drug for supportive care in our database. However, it is also used off-label to prevent cancer recurrence in breast cancer (Ottewell and Wilson, 2019). If the license is updated to include this new medical use then zoledronic acid would be included as an anticancer drug.

In addition to maintaining the currency of the data through periodic monitoring of new drug approvals, we intend to enhance and expand the range of data collected for each entry in the database.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://osf.io/nk8pg/> <http://www.redo-project.org/cancer-drugs-db/>.

AUTHOR CONTRIBUTIONS

PP and GB conceived this work. All authors took part in the data analysis and screening of drugs. PP wrote the manuscript and all authors edited and approved the final draft.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2021.627574/full#supplementary-material>.

REFERENCES

- Avram, S., Bologna, C. G., Holmes, J., Bocci, G., Wilson, T. B., Nguyen, D.-T., et al. (2021). DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res.* 49, D1160–D1169. doi:10.1093/nar/gkaa997
- Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., et al. (2017). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 46, D1068. doi:10.1093/nar/gkx1143
- Curpan, 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. doi:10.1093/nar/gkx1154
- Fujiwara, Y. (1999). Current status of oral anticancer drugs in Japan. *J. Clin. Oncol.* 17, 3362–3365. doi:10.1200/jco.1999.17.10.3362
- Gramatica, R., Di Matteo, T., Giorgetti, S., Barbiani, M., Bevec, D., and Aste, T. (2014). Graph theory enables drug repurposing - how a mathematical model can drive the discovery of hidden mechanisms of action. *PLoS One* 9, e84912. doi:10.1371/journal.pone.0084912
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38, D355–D360. doi:10.1093/nar/gkp896
- MacLean, F., Pantziarka, P., Nazarian, J., and Wilson, J. (2020). A network polypharmacology approach to drug repurposing for diffuse intrinsic pontine glioma. bioRxiv. doi:10.06.14.15071410.1101/2020.06.14.150714
- Ottewell, P., and Wilson, C. (2019). Bone-targeted agents in breast cancer: do we now have all the answers?. *Breast Cancer* 13, 117822341984350. doi:10.1177/1178223419843501
- Pantziarka, P., Verbaanderd, C., Sukhatme, V., Capistrano, R., Crispino, S., Gyawali, B., et al. (2018). ReDO_DB: the repurposing drugs in oncology database. *Ecancermedicalscience* 12, 886. doi:10.3332/ecancer.2018.886
- Pantziarka, P., Meheus, L., Rombauts, K., Vandeborne, L., and Bouche, G. (2020a). “Drug repurposing for cancer therapy—an introduction,” in *Drug repurposing in cancer therapy* Editors K. K. W. To and W. C. S. Cho (London: Elsevier), 1–14.
- Pantziarka, P., Verbaanderd, C., Huys, I., Bouche, G., and Meheus, L. (2021b). Repurposing drugs in oncology: from candidate selection to clinical adoption. *Semin. Cancer Biol.* 68, 186. doi:10.1016/j.semcancer.2020.01.008
- Tang, J., Tanoli, Z.-u.-R., Ravikumar, B., Alam, Z., Rebane, A., Vähä-Koskela, M., et al. (2018). Drug target Commons: a community effort to build a consensus knowledge base for drug-target interactions. *Cel Chem. Biol.* 25, 224–229. doi:10.1016/j.chembiol.2017.11.009
- Thorn, C. F., Klein, T. E., and Altman, R. B. (2013). PharmGKB: the pharmacogenomics knowledge base. *Methods Mol. Biol.* 1015, 311–320. doi:10.1007/978-1-62703-435-7_20
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037
- Wu, Y.-L., Zhang, H., and Yang, Y. (2015). Cancer drug development in China: recent advances and future challenges. *Drug Discov. Today* 20, 766–771. doi:10.1016/j.drudis.2014.11.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Pantziarka, Capistrano I, De Potter, Vandeborne and Bouche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.