Developing software to build networks and perform data integration in precision oncology

Simina M. Boca

Assistant Professor

Georgetown University Medical Center

Twitter: @siminaboca

BioC Meeting

June 25, 2019

Acknowledgments

Georgetown University:

- Shruti Rao
- Krithika Bhuvaneshwar
- Robert Beckman
- Rebecca Riggins
- Subha Madhavan

University of Maryland College Park:

- Jayaram Kancherla
- Héctor Corrada Bravo (site PI)

Grant Support:

R21CA220398 and R21 CA220398-02S1 (PI: Boca), U41HG009649 (ClinGen), Lombardi Cancer Center Support Grant pilot award

Thank you to the BioC 2019 Conference Committee for inviting me and to the Bioconductor Project for providing awesome open-source software!

Project goal

- Build a tool – **CDGnet** (Cancer Drug Gene networks) – to help researchers and eventually clinicians expand the number of targeted therapies available for individuals with cancer who have specific mutations (DNA changes) or genes/proteins that are overexpressed.

Precision oncology framework

Precision oncology refers to tailoring interventions to patients in ways that go beyond traditional characteristics of age, sex, disease, symptoms etc by considering biomarkers.

Biomarkers may be:

- genetic characteristics: can be either *germline* (inherited, in normal tissue) or *somatic* (in cancer cells but not normal tissue)
- mRNA or protein expression values: refer to expression in tumors, either in comparison to other tumors or to adjacent normal tissues

Tumor molecular profiling and targeted therapies

It is now routine to perform molecular profiling in certain tumor types to check for specific molecular features at diagnosis to decide on a targeted treatment plan eg:

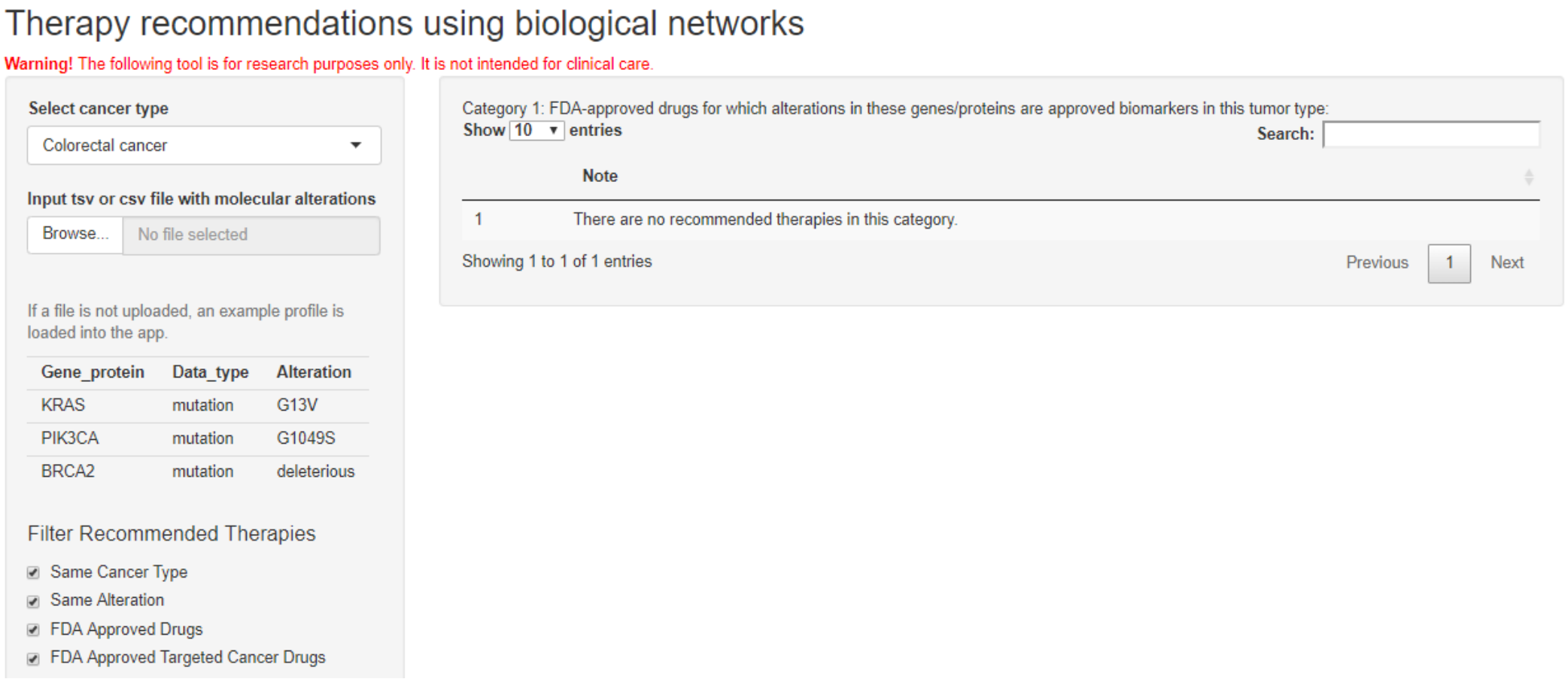
- KRAS-wild type (non-mutated) colorectal cancer is treated with EGFR inhibitors (DNA alteration)
- ER+ breast cancer is treated with tamoxifen or fulvestrant, HER2+ breast cancer is treated with trastuzumab (mRNA/protein expression)

In many cases tumor molecular profiling is used after a patient has progressed on multiple lines of therapy and/or has few/no therapy options left.

- Patient may then receive an off-label therapy that is prescribed for their alteration in another tumor type
- Our goal with CDGnet is to expand the number of possible targeted therapies and prioritize them using drug-gene network approaches

CDGnet landing page and example user inputs

http://epiviz.cbcb.umd.edu/shiny/CDGnet/

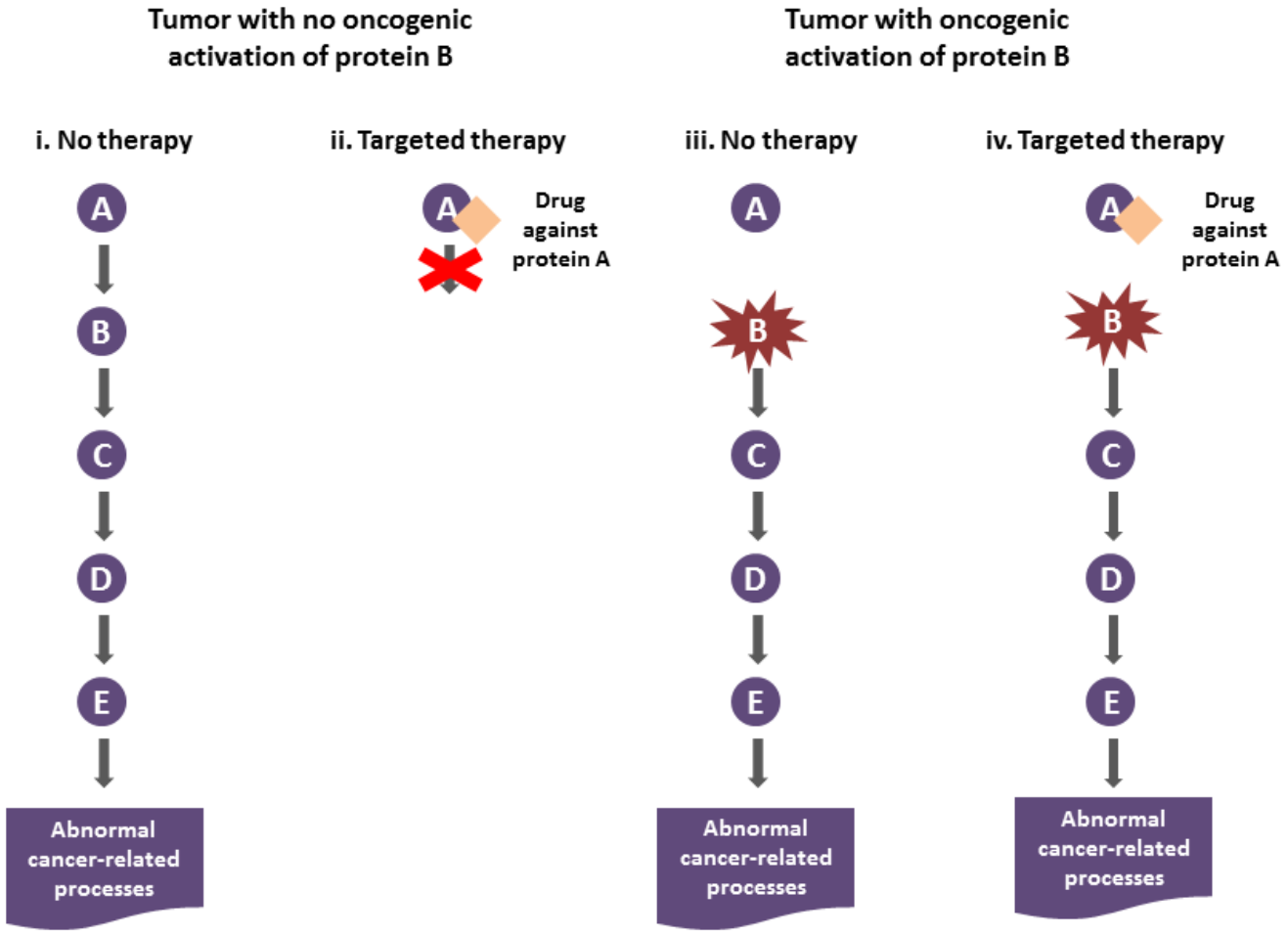


Want to enrich this by including biological pathway information

How do we do this?

How do we include pathway information?

Look at downstream targets of oncogenes



Therapy prioritization: 4 categories

1. FDA-approved drugs for which the patient's alterations/genes/proteins are biomarkers in their tumor type
2. FDA-approved drugs for which the patient's alterations/genes/proteins are biomarkers in other tumor types
3. Drugs which have as targets these alterations/genes/proteins or as biomarkers/targets others that are downstream of input oncogenes when considering the pathway corresponding to this tumor type.*
4. Drugs which have as targets/biomarkers either these alterations/genes/proteins or as biomarkers/targets others that are downstream of input oncogenes when considering the pathways corresponding to other tumor types.*

* Could be targeted drugs prescribed for their tumor type or other tumor types OR any FDA-approved drug OR any drug in DrugBank.

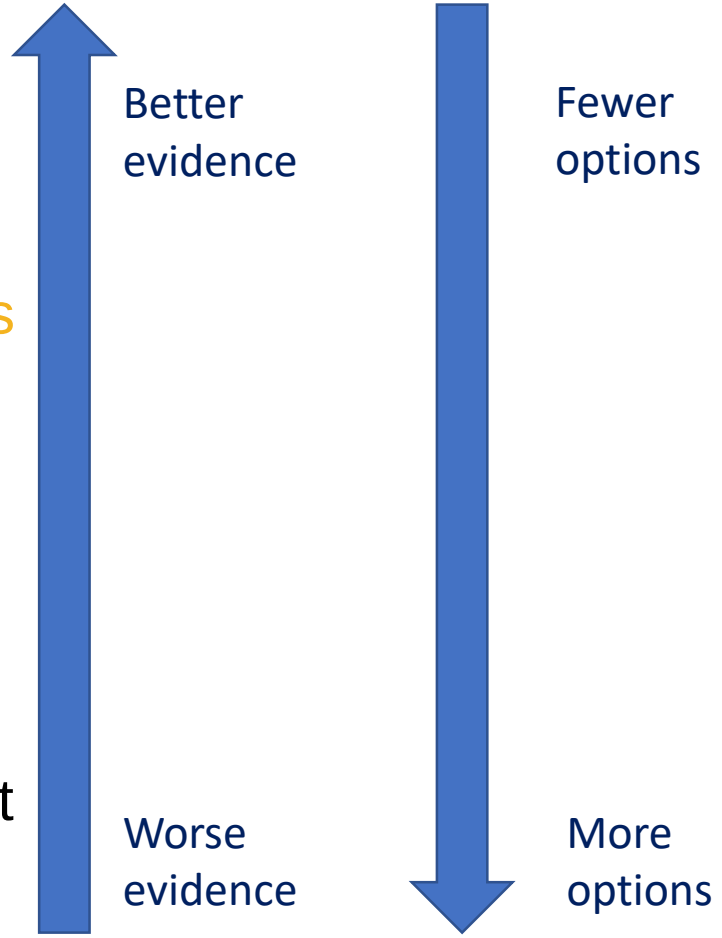
Our tool development goals

Have an approach that is:

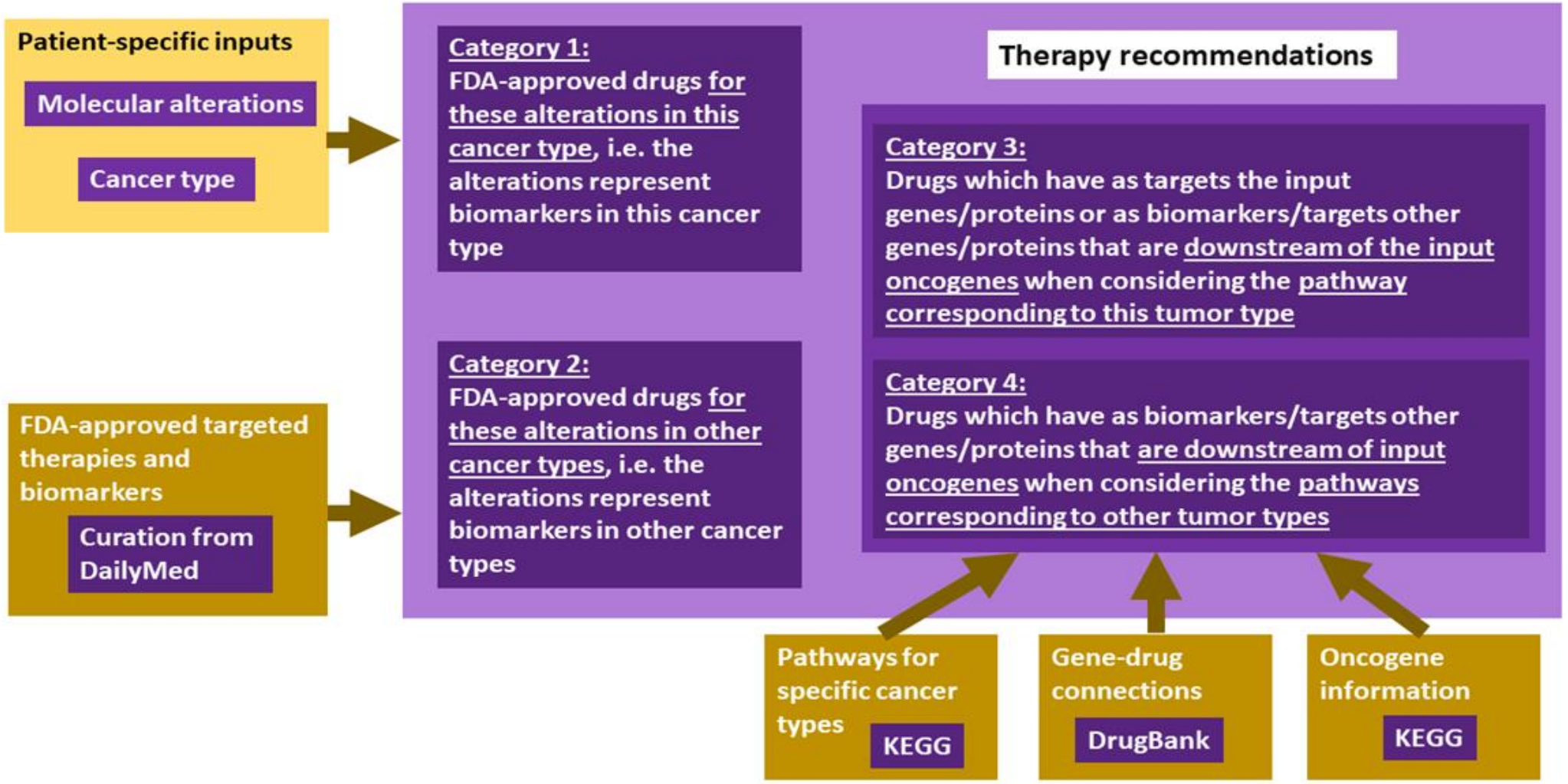
1. Automated
2. Transparent
3. Personalized to individual patients
4. Evidence-based

Therapy prioritization: 4 categories

1. FDA-approved drugs for which the patient's alterations/genes/proteins are biomarkers in their tumor type
2. FDA-approved drugs for which the patient's alterations/genes/proteins are biomarkers in other tumor types
3. Drugs which have as targets these alterations/genes/proteins or as biomarkers/targets others that are downstream of input oncogenes when considering the pathway corresponding to this tumor type.
4. Drugs which have as targets/biomarkers either these alterations/genes/proteins or as biomarkers/targets others that are downstream of input oncogenes when considering the pathways corresponding to other tumor types.



General approach and data sources

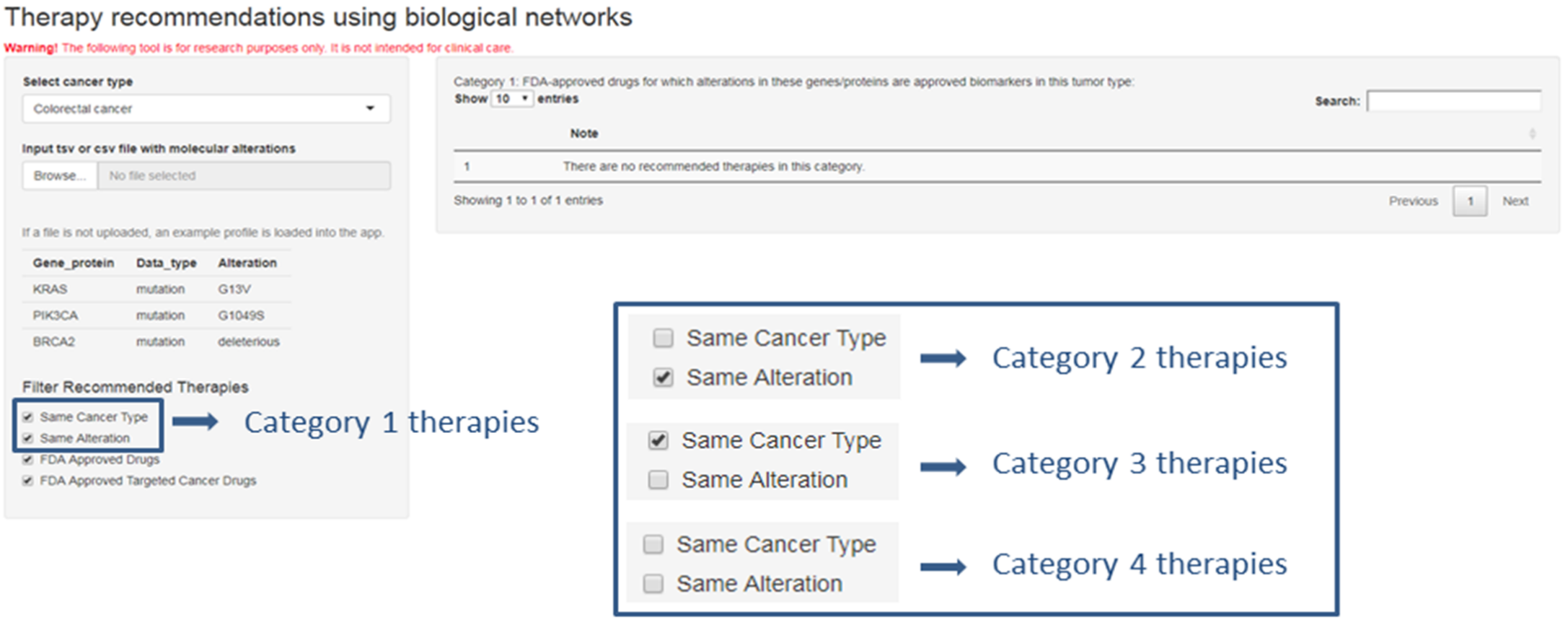


Some important points

- This is currently a data integration, not a statistical project
- Results can be seen as predictions for a single patient or a population of patients with the same molecular profile and cancer type
- We looked for resources that are easy to use and integrate along with being transparent and evidence-based

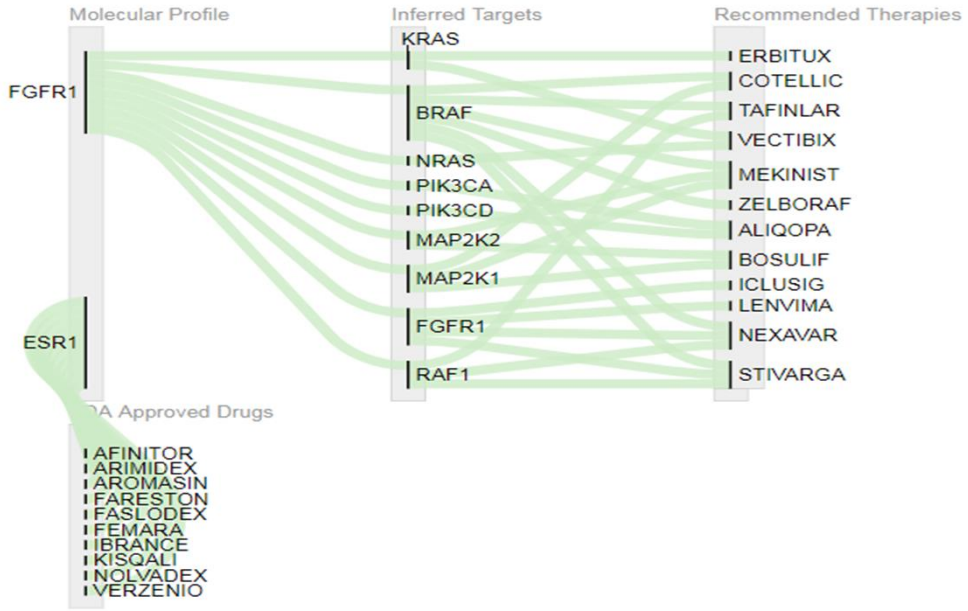
CDGnet landing page and example user inputs

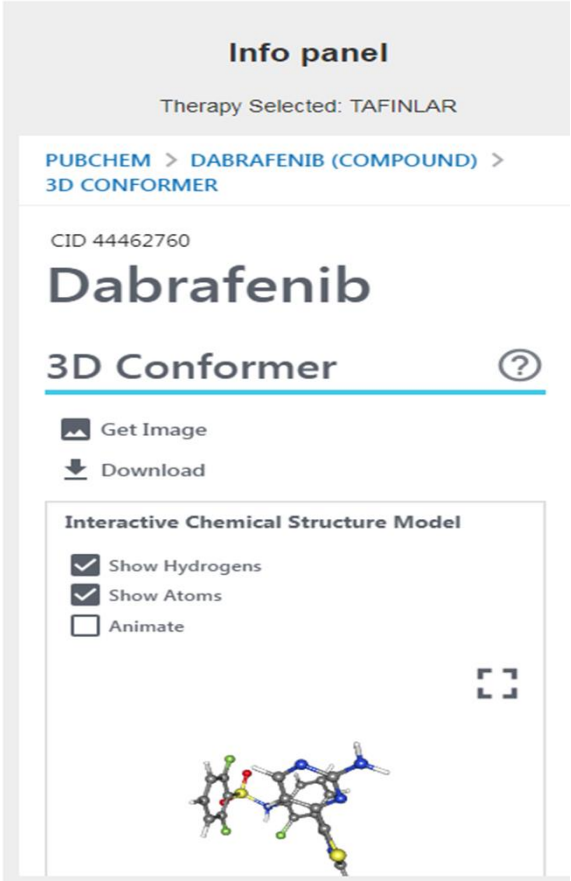
http://epiviz.cbcb.umd.edu/shiny/CDGnet/



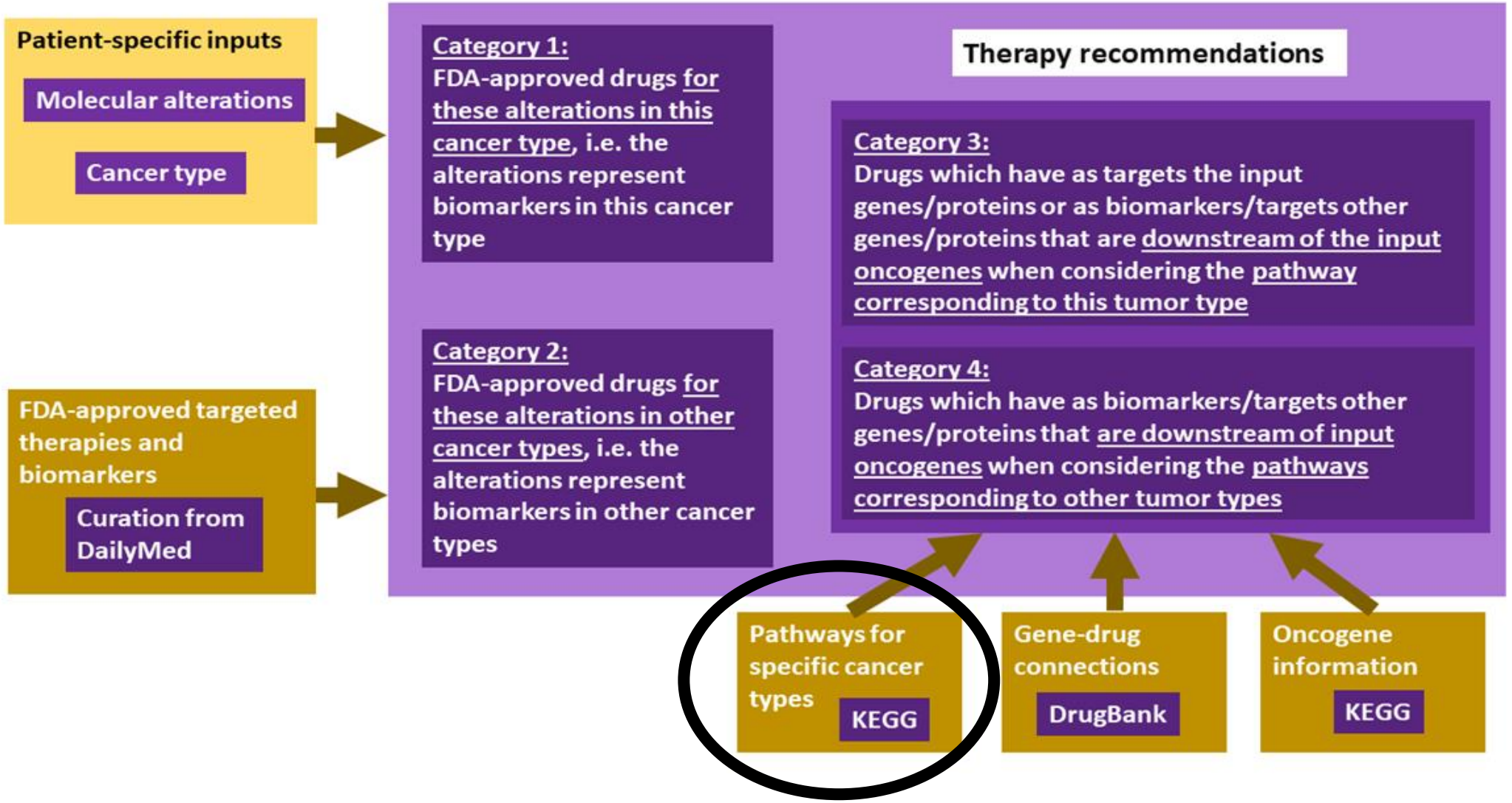
CDGnet flow diagram example for category 3 recommendations

http://epiviz.cbcb.umd.edu/shiny/CDGnet/



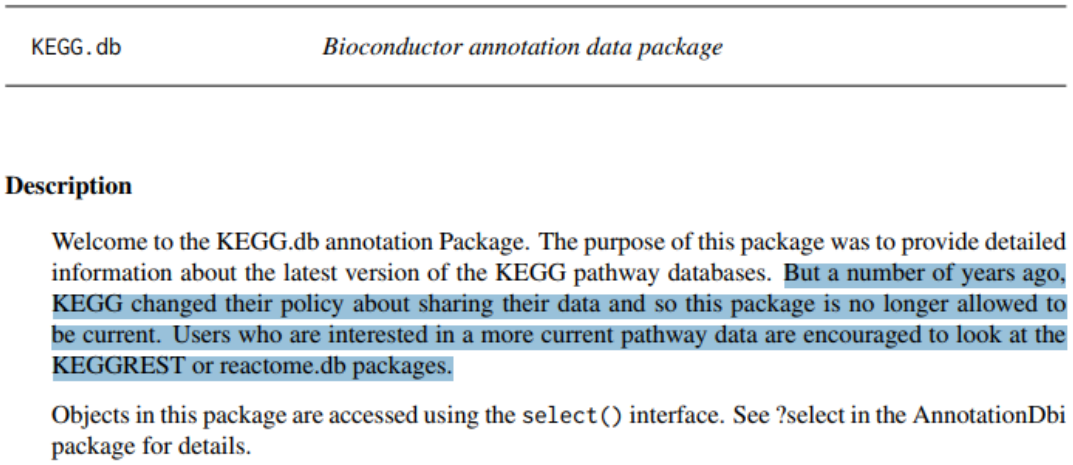


General approach and data sources



How did we choose the pathways?

- We are currently using the KEGG pathways
- **Benefits:**
  - Well-known, include pathways for many individual cancers
  - Expert-curated pathways
- **Issues:**
  - May not always agree with papers chosen by curators, some pathways are somewhat out of date
  - Not all cancer types included



KEGG.db *Bioconductor annotation data package*

Description

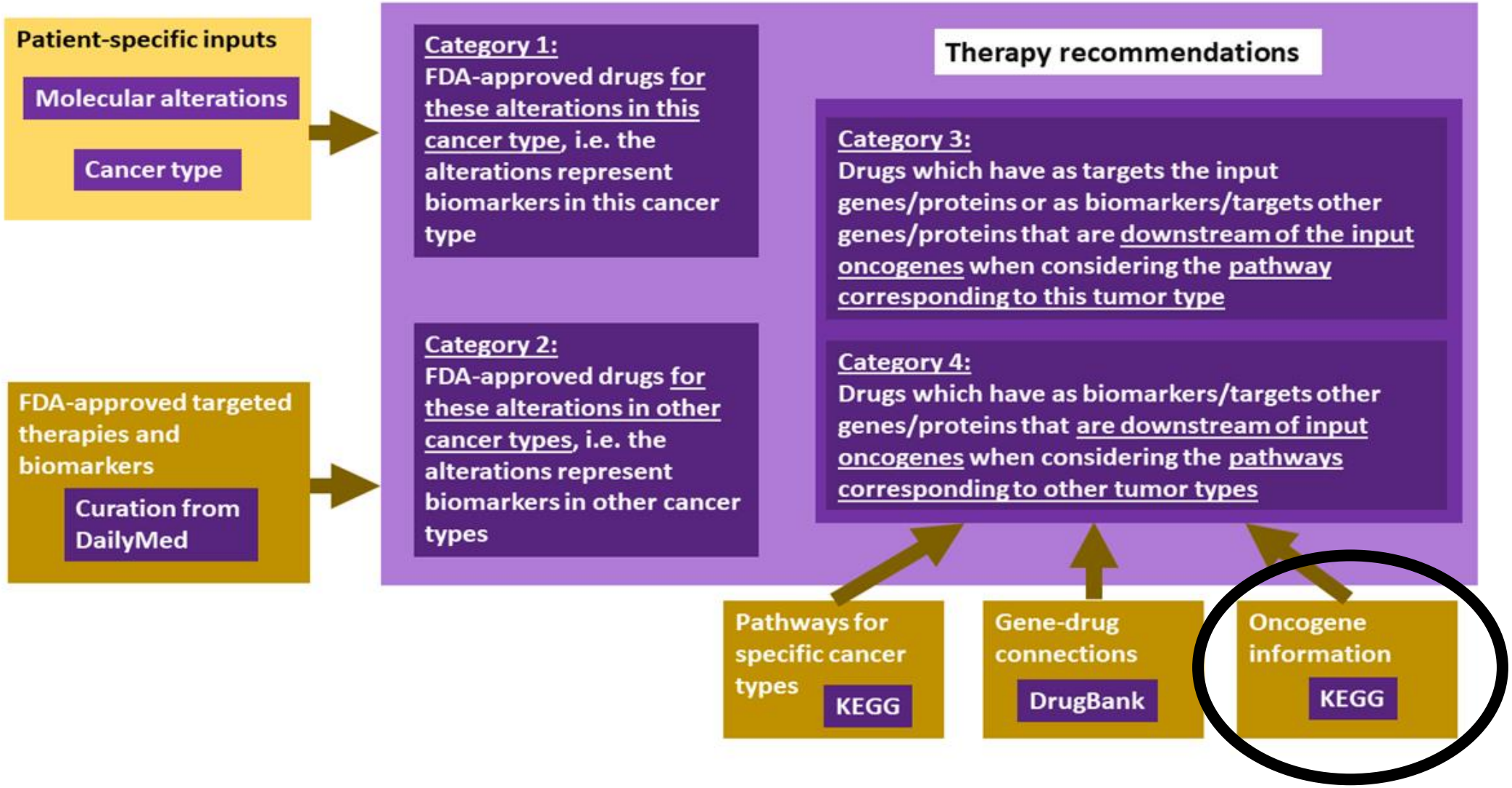
Welcome to the KEGG.db annotation Package. The purpose of this package was to provide detailed information about the latest version of the KEGG pathway databases. But a number of years ago, KEGG changed their policy about sharing their data and so this package is no longer allowed to be current. Users who are interested in a more current pathway data are encouraged to look at the KEGGREST or reactome.db packages.

Objects in this package are accessed using the `select()` interface. See ?select in the AnnotationDbi package for details.

How did we choose the pathways?

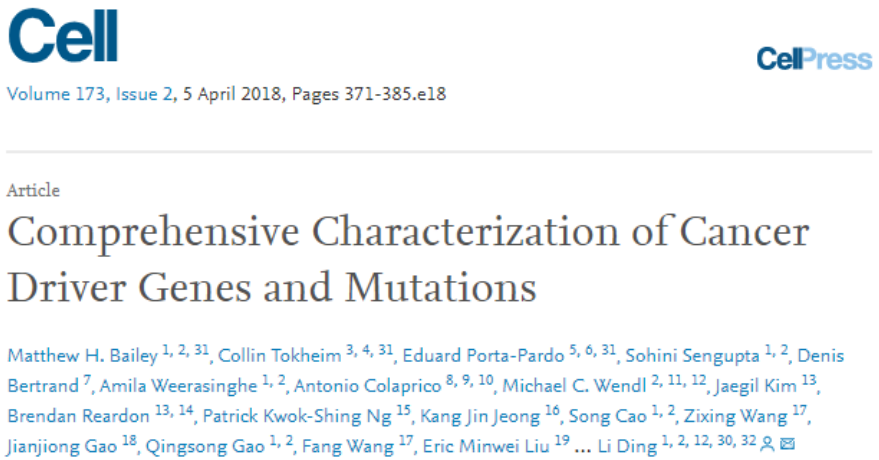
- We are currently using the KEGG pathways
- **Benefits:**
  - Well-known, include pathways for many individual cancers
  - Expert-curated pathways
- **Issues:**
  - May not always agree with papers chosen by curators, some pathways are somewhat out of date
  - Not all cancer types included
  - Due to KEGG's policy, have the following approach to pulling in their data:
    - Use KEGGREST package to get access to KEGGREST API
    - Download all human pathways as KGML files (KEGG-specific XML files)
    - Use KEGGgraph package to convert KGML files into data frames
    - Use org.Hs.eg.db package to convert KEGG gene IDs to gene names
  - This approach makes it difficult to perform updates and harder for others to reproduce our work

General approach and data sources

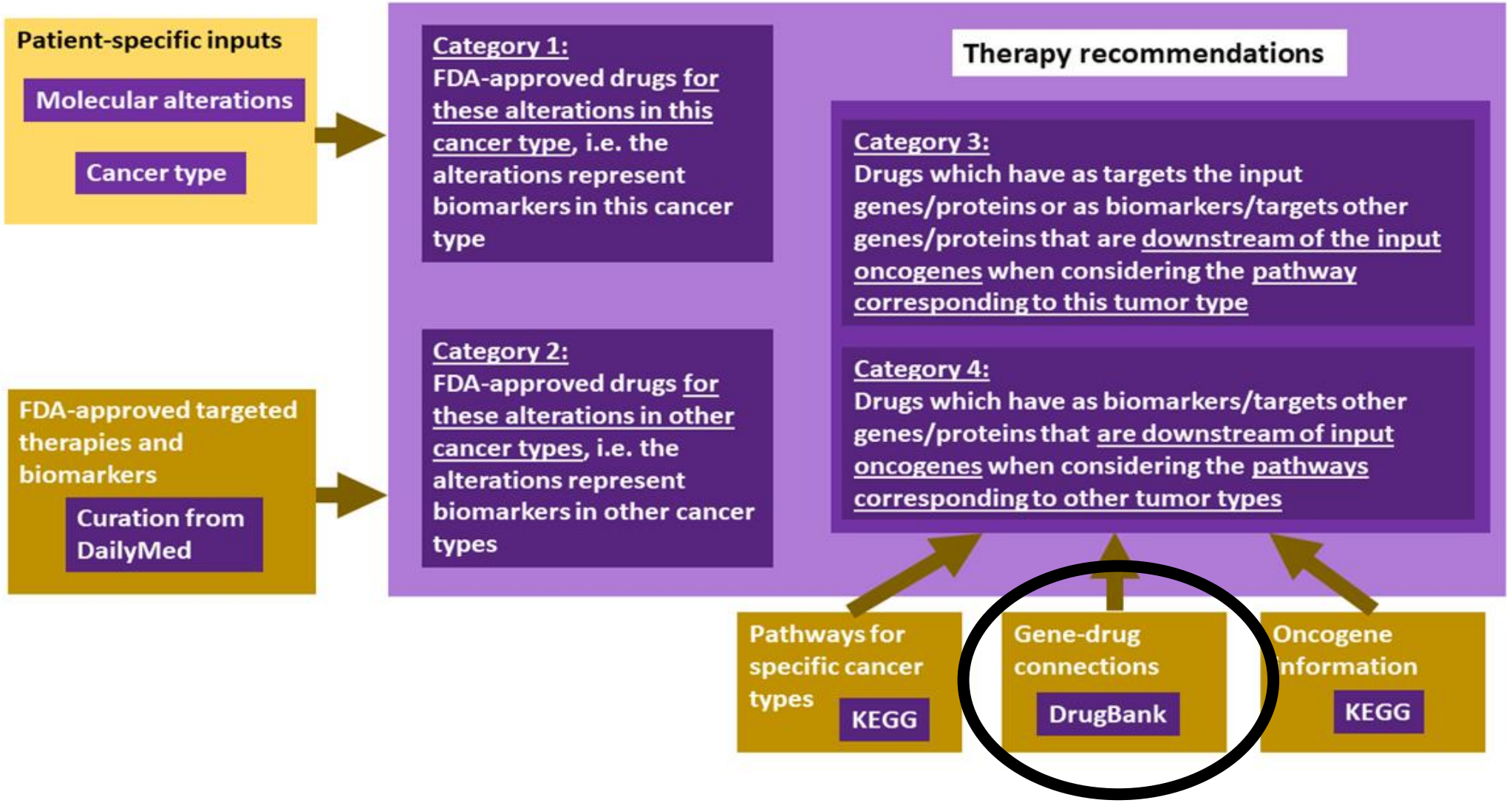


How did we choose the oncogenes?

- We are currently using the oncogenes provided by KEGG for each cancer type
  - This ensures that we are considering the same cancer types for pathways and oncogenes
- Note that the same gene may not always be expressed or have the same function in all tissues so oncogenic activity often differs by tumor time
- Being classified as an “oncogene” is based on applying a threshold to results from a mathematical or statistical model
- For internal use, we are also considering the candidate oncogenes from the pan-TCGA paper:

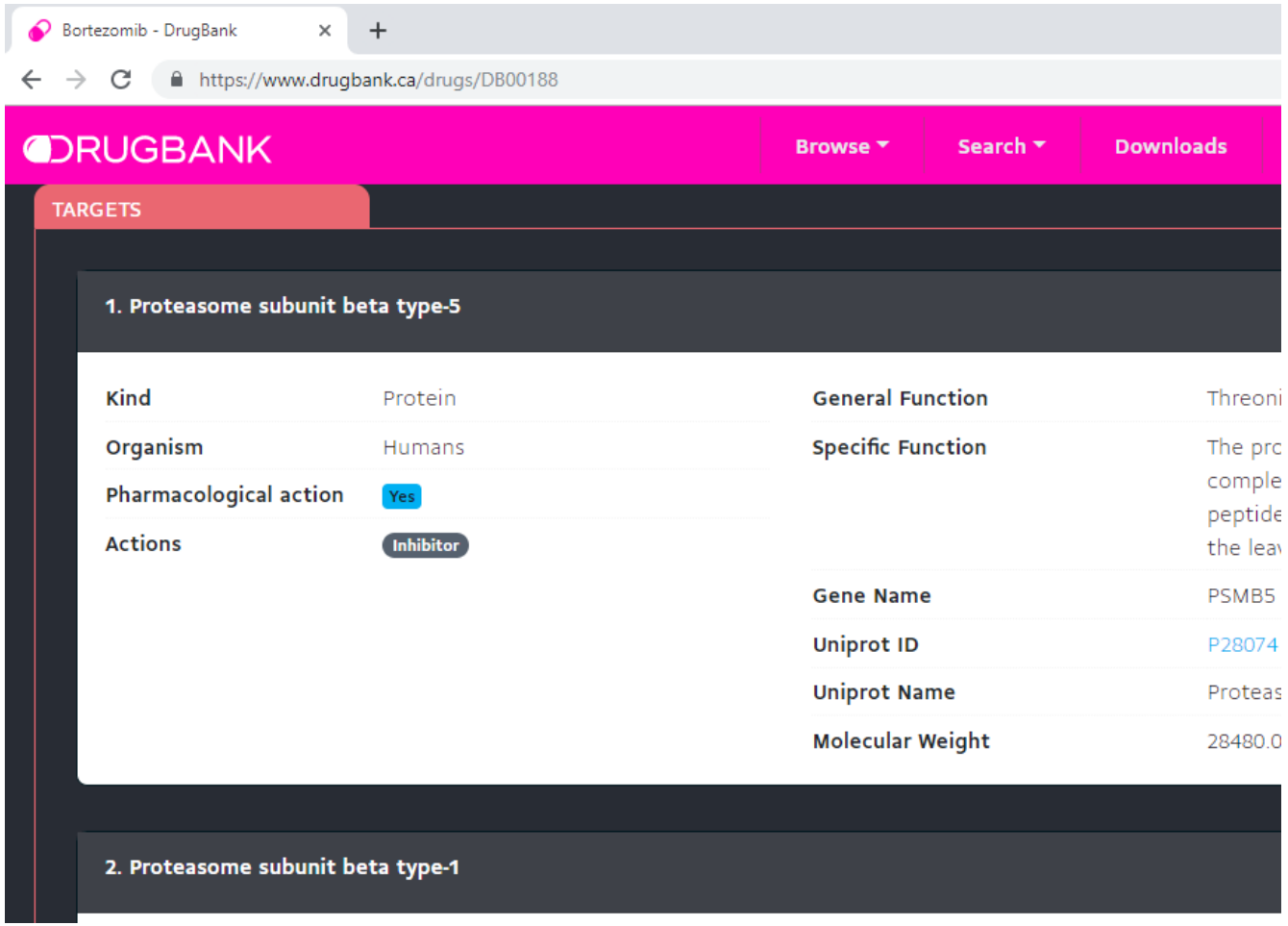


General approach and data sources



How did we choose the drug targets?

- We considered the DrugBank database, due to its comprehensiveness and expert-curation
- Note that DrugBank is not tissue-specific



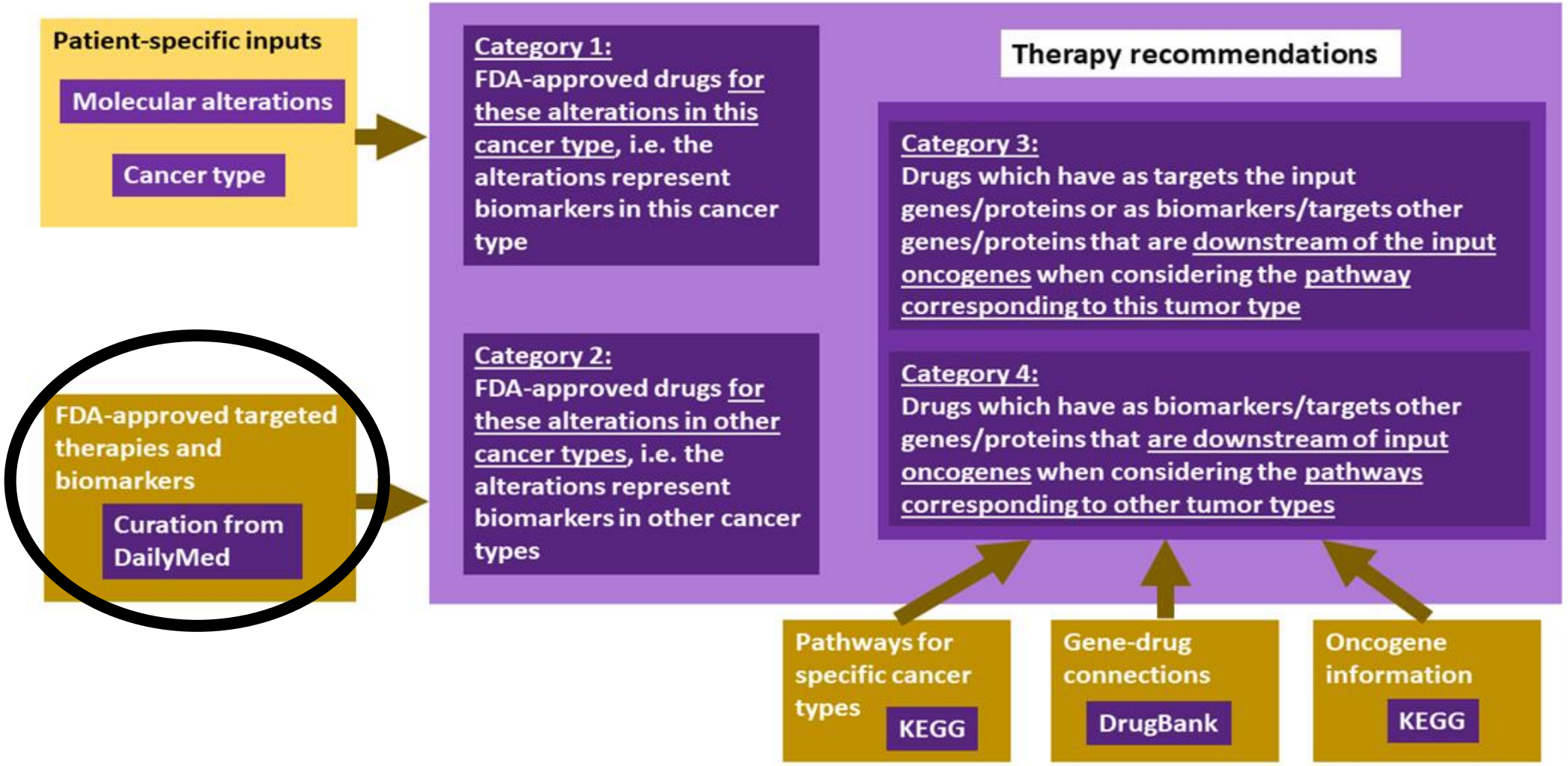
How did we choose the drug targets?

- We considered the DrugBank database, due to its comprehensiveness and expert-curation

- Note that DrugBank is not tissue-specific

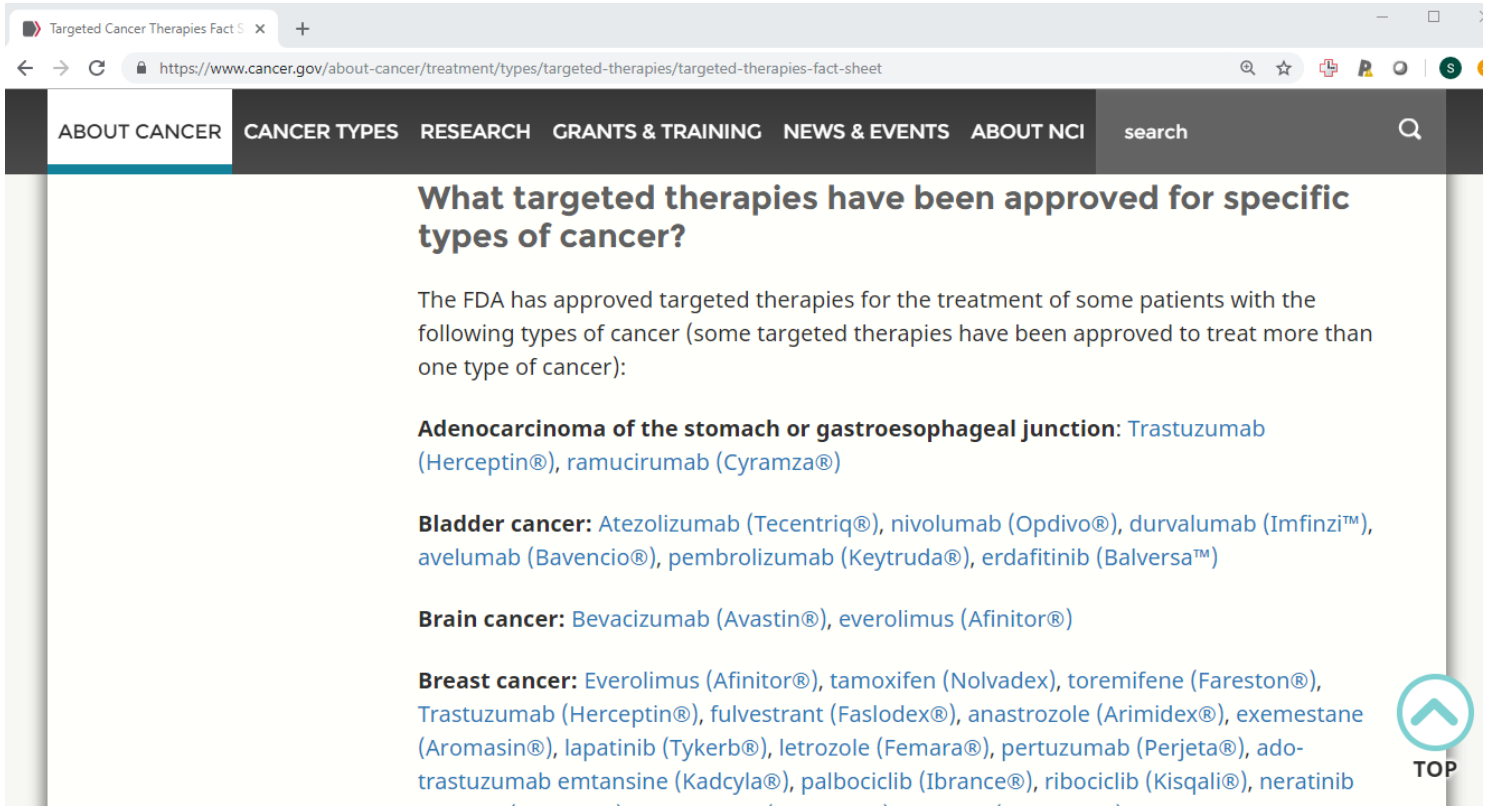
- We downloaded all of DrugBank and read in the data, but will look into the ChemmineDrugs and paxtoolsr packages for the next iteration of our tool

General approach and data sources



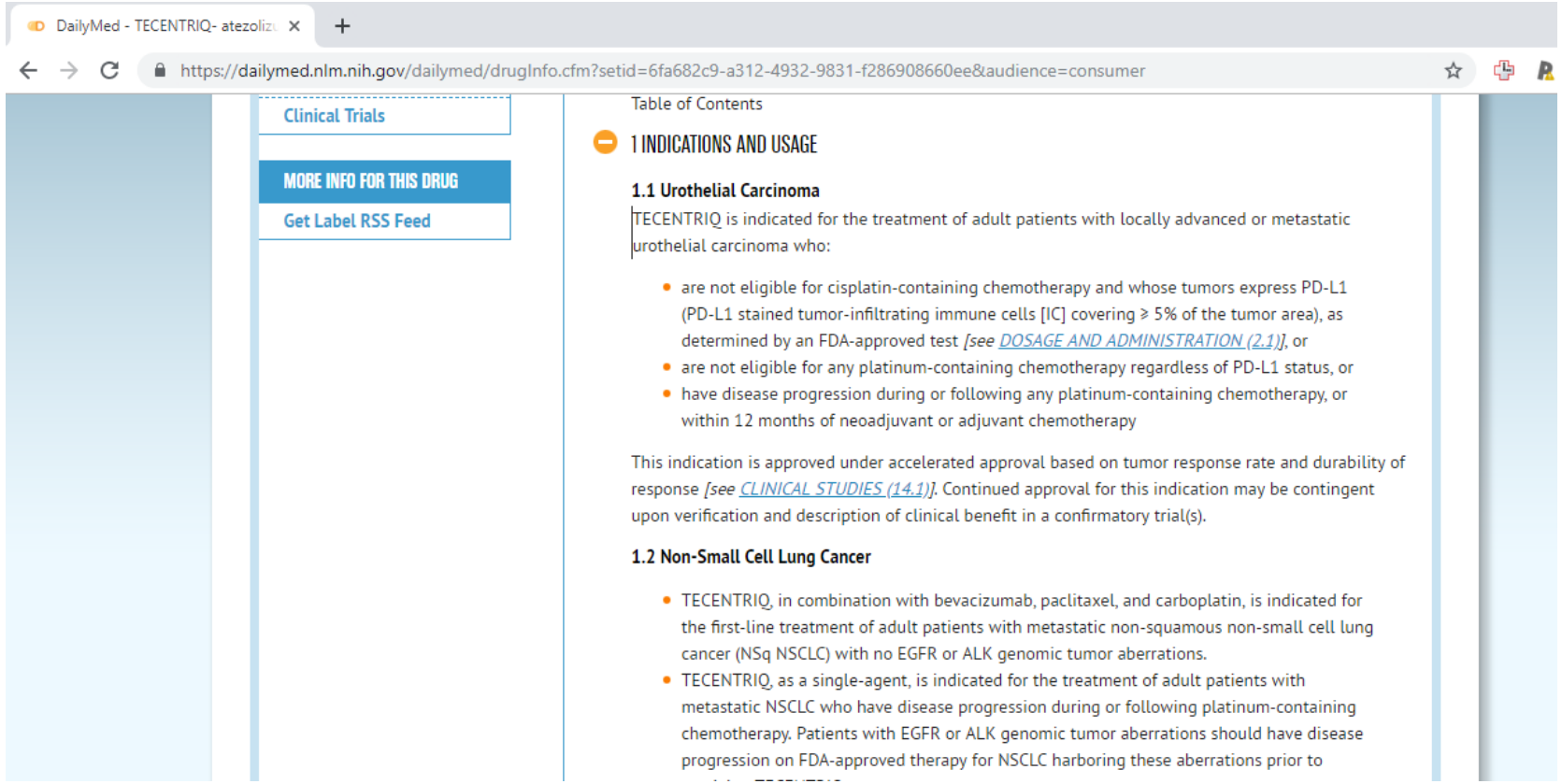
How did we obtain the FDA-approved targeted therapies and biomarkers?

- We performed manual curation of targeted cancer therapies:
  - We first got the list of targeted therapies from the NCI:



How did we obtain the FDA-approved targeted therapies and biomarkers?

- We performed manual curation of targeted cancer therapies:
  - We first got the list of targeted therapies from the NCI
  - We then followed the links to get to the drug labels, which are available through NLM's DailyMed and curated the therapy, biomarker, disease from the section "Indications and Usage":

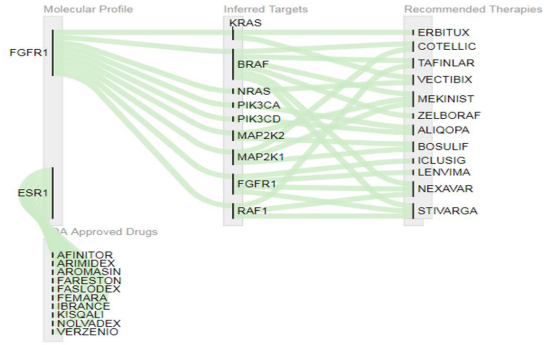


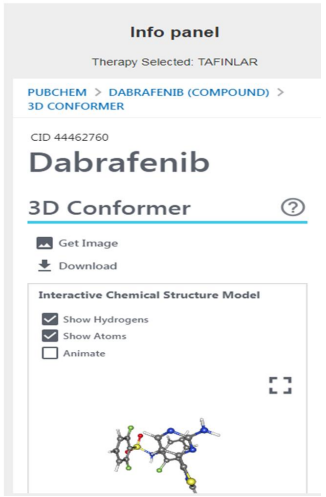
How did we obtain the FDA-approved targeted therapies and biomarkers?

- We performed manual curation of targeted cancer therapies:
  - We first got the of targeted therapies list from the NCI
  - We then followed the links to get to the drug labels, which are available through NLM's DailyMed and curated the therapy, biomarker, disease from the section "Indications and Usage"
- Clearly this does not scale up well for maintaining our tool, especially given the large number of recent approvals
- Databases exist with some of this information, but they were either difficult to use or had restrictive permissions
- Currently developing a package to do at least some preliminary curation via natural language processing on the DailyMed API

Relevant links

- Shiny-based web application available at http://epiviz.cbcb.umd.edu/shiny/CDGnet/





- Code available at https://github.com/SiminaB/CDGnet and https://github.com/jkanche/nfpmShinyComponent (package for interactive visualization)
- Preprint available at https://www.biorxiv.org/content/10.1101/605261v1 (Kancherla, Rao et al)

Future development goals

- Categorize therapies by therapy classes, approval type etc, to rank them within categories and reduce the number of disjoint recommendations
- Consider other pathways besides KEGG, potentially allow users to upload their own set of pathways
- Automate more of the current steps
- Drill deeper into 1-2 different cancer types

Thank you!

Email me at:
smb310@georgetown.edu

Tweet to:

@siminaboca