Databases Of The EGFR Variants And Target Drugs, A Mini Review

Sifen Lu¹, Yongzhao Zhou² and Maoshan Chen³*

¹Precision Medicine Center, Precision Medicine Key Laboratory of Sichuan Province, West China Hospital of Sichuan University, Wuhou District National School Lane, Chengdu, Sichuan, China 610041.
²Department of Respiratory and Critical Care Medicine, West China Hospital of Sichuan University, No. 37, Wuhou District National School Lane, Chengdu, Sichuan, China 610041.
³Australian Centre for Blood Diseases, Department of Clinical Haematology, Central Clinical School, Australia

*Corresponding author: Maoshan Chen, Australian Centre for Blood Diseases, Department of Clinical Haematology, Central Clinical School, Melbourne, Victoria, Australia, 3004

ARTICLE INFO

Received: June 11, 2019
Published: June 18, 2019
Keywords: Annotation; EGFR TKI; NSCLC

ABSTRACT

Epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors (TKIs) are highly promising drugs that are well tolerated and have good antitumor activity in non-small cell lung cancer (NSCLC) patients with sensitive mutations in the EGFR gene. Here we reviewed 9 databases, including CKB, CIVIC, PMKB, COSMIC, My Cancer Genome, OncoKB, FDA, ClinicalTrials.gov and NCCN, in which we can obtain the EGFR mutants and related target drug annotations. Some of them have launched the information of EGFR TKIs and related activating EGFR mutations for advanced NSCLC, such as My Cancer Genome, CKB and CIVIC. Although these databases have overlaps in the variants and annotations, their major contents vary. This leads the increase of complexity of annotating target drug information for the EGFR mutants.

Abbreviations: ORR: Objective Response Rate; PFS: Progression-Free Survival; OS: Overall Survival

Introduction

Lung cancer has the highest incidence among all cancers and is the leading cause of cancer-related death worldwide. Non-small-cell lung cancer (NSCLC) accounts for approximately 80% of lung cancer cases [1]. In more than 40% NSCLC patients, the gene encoding epidermal growth factor receptor (EGFR), a transmembrane glycoprotein, is overexpressed and further mutant EGFR is present in more than 50% Asia NSCLC patients [2]. EGFR is a member closely related to growth factor receptor tyrosine kinases (RTKs) that consist of four members: EGFR (HER1/ErbB1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4), which regulate many developmental, metabolic and physiological processes [3]. Activating mutations within the tyrosine kinase domain of the EGFR are found in more than 40% of the Asia lung adenocarcinoma as [4]. The EGFR-tyrosine kinase inhibitors (TKIs) are highly promising drugs that are well tolerated and have good antitumor activity in NSCLC patients with sensitive mutations in the EGFR gene [5]. In particular, the use of EGFR-TKIs in the treatment of advanced NSCLC patients harboring activating EGFR mutations has markedly improved survival outcomes, particularly in Asian descent patients [6]. While, the drugs resistance is the most challenging problem of the EGFR-TKI treatment in NSCLC patients. The major mechanisms giving rise to EGFR-TKI resistance in NSCLC have been demonstrated, including secondary mutation of EGFR T790M and other related positions [7].

Due to the development of sequencing technology, it is getting easier and cheaper to identify mutations in a particular gene or in the genome. Many bioinformatics tools have been developed to characterize gene mutations, such as BWA [8,9], GATK [10] and VarScan [11], and to annotate the mutations, such as ANNOVAR [12] and SnpEff [13]. However, neither provides the target drug information related to gene variants, which is still not working sufficient in clinical sequencing workflow [14]. To obtain such information, users have to search in some other databases, such as MyCancerGenome [15], OncoKB [16], ClinicalTrials.gov [17], CIVIC...
Although all these databases provide clinical information of EGFR mutations and target drug information, they focus on different areas and a combination of these databases can help to understand the EGFR-TKI mutations and target drug information. In this review, we summarized and compared 9 databases (Table 1) that can help to annotate the EGFR-TKI mutations and related target drug information.

**Table 1:** The major contents vary among the 9 databases.

<table>
<thead>
<tr>
<th>ID</th>
<th>Database Name</th>
<th>Database link</th>
<th>Main content</th>
<th>Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CKB</td>
<td><a href="https://ckb.jax.org/">https://ckb.jax.org/</a></td>
<td>complex cancer genomic profiles in the context of protein impact, therapies, and clinical trials</td>
<td>Jackson Laboratory</td>
</tr>
<tr>
<td>2</td>
<td>CIVIC</td>
<td><a href="https://civicdb.org/">https://civicdb.org/</a></td>
<td>the clinical interpretation of variants</td>
<td>Washington University School of Medicine</td>
</tr>
<tr>
<td>3</td>
<td>PMKB</td>
<td><a href="https://pmkb.weill.cornell.edu">https://pmkb.weill.cornell.edu</a></td>
<td>clinical cancer variants and interpretations</td>
<td>Weill Cornell Medicine</td>
</tr>
<tr>
<td>4</td>
<td>COSMIC</td>
<td><a href="https://cancer.sanger.ac.uk">https://cancer.sanger.ac.uk</a></td>
<td>the impact of somatic mutations</td>
<td>Sanger</td>
</tr>
<tr>
<td>5</td>
<td>My Cancer Genome</td>
<td><a href="https://www.mycancergenome.org">https://www.mycancergenome.org</a></td>
<td>related therapeutic implications</td>
<td>Vanderbilt-Ingram Cancer Center</td>
</tr>
<tr>
<td>6</td>
<td>OncoKB</td>
<td><a href="http://oncokb.org/">http://oncokb.org/</a></td>
<td>effects and treatment implications of specific cancer gene alterations</td>
<td>Memorial Sloan Kettering Cancer Center</td>
</tr>
<tr>
<td>7</td>
<td>FDA</td>
<td><a href="https://www.fda.gov">https://www.fda.gov</a></td>
<td>drugs</td>
<td>U.S. Department of Health and Human Services</td>
</tr>
<tr>
<td>8</td>
<td>ClinicalTrials.gov</td>
<td><a href="https://clinicaltrials.gov">https://clinicaltrials.gov</a></td>
<td>publicly funded clinical studies conducted around the world</td>
<td>U.S. National Institutes of Health</td>
</tr>
<tr>
<td>9</td>
<td>NCCN</td>
<td><a href="https://www.nccn.org/">https://www.nccn.org/</a></td>
<td>clinical practice guidelines</td>
<td>National Comprehensive Cancer Network</td>
</tr>
</tbody>
</table>

**Comparison of 9 Public Databases**

The major contents vary among the 9 databases (Table 1). Basic information of related drug (e.g., name) can be searched using the name of EGFR mutant in all databases. While, detailed information about the variants or the drugs are selected to present in the search results with different database (Table 2). Among them, CKB, CIVIC, PMKB, COSMIC and MyCancerGenome can provide some information of the EGFR mutant, such as the transcript, effect of the mutation, protein domain, pathways and ethnic frequency in population. While OncoKB, FDA, ClinicalTrials.gov and NCCN are more focusing the drugs. In the FDA website users can obtain the most detailed information about target drugs using the name of EGFR mutant.
Table 2: Detailed information about the variants or the drugs are selected to present in the search results with different database.

<table>
<thead>
<tr>
<th>Database/Tool</th>
<th>Link</th>
<th>Name</th>
<th>Transcript</th>
<th>Position</th>
<th>Effect</th>
<th>Protein Domains</th>
<th>Pathways</th>
<th>Somatic Status</th>
<th>Variant description</th>
<th>Ethnic frequency</th>
<th>Drug name</th>
<th>Response Type</th>
<th>Evidence level</th>
<th>Efficacy Evidence</th>
<th>ORR</th>
<th>PFS</th>
<th>OS</th>
<th>Adverse Reaction</th>
<th>Ethnic information</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKB</td>
<td><a href="https://ckb.jax.org/">https://ckb.jax.org/</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CIVIC</td>
<td><a href="https://civicdb.org/">https://civicdb.org/</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PMKB</td>
<td><a href="https://pmkb.weillcornell.edu">https://pmkb.weillcornell.edu</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>COSMIC</td>
<td><a href="https://cancer.sanger.ac.uk">https://cancer.sanger.ac.uk</a></td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>My Cancer Genome</td>
<td><a href="https://www.mycancergenome.org">https://www.mycancergenome.org</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OncoKB</td>
<td><a href="http://oncokb.org/">http://oncokb.org/</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FDA</td>
<td><a href="https://www.fda.gov/">https://www.fda.gov/</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ClinicalTrials.gov</td>
<td><a href="https://clinicaltrials.gov">https://clinicaltrials.gov</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NCCN</td>
<td><a href="https://www.nccn.org/">https://www.nccn.org/</a></td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: Position refers to genomic coordinates, such as chr7:55249071.55249071.

Note: Effect refers to gain-of-function or loss-of-function.

Note: Variant description refers to the details, such as protein domain, biochemically, protein function, report information of the variant.

Note: Response Type refers to reaction to the drug, such as sensitive or resistant.

Note: Evidence level refers to reliability, such as FDA approved or clinical study and so on.
The characteristics of each database are as follows:

a. The main display information of CKB database is classified by three parts – gene variants, molecular profiles and gene level evidence. It provides information as comprehensive as possible and uses a clear information presentation way for users to understand. But complete information of EGFR is not directly available for free.

b. The CIVIC database provides information as comprehensive as possible, which includes much data from timely public reference papers. While the ethnic information is missing, and it takes time to filter the results from a lot of hits with low-level evidence.

c. The PMKB database is mainly classified by three parts – gene, variant and interpretation. However, no detailed drug related information can be found in this database, such as ORR, median PFS, median OS and adverse reaction.

d. The COSMIC database is a relative complete database for gene variants. It can be better if some more information of the drugs is provided, rather than the names only.

e. The MyCancerGenome database provides relative full information. The annotated information can be chosen by disease, gene and variant. But the information of drugs is not as detailed as the FDA.

f. The OncoKB database provides information classified by evidence level. It links the gene variants with related drugs. However, detailed information is missing for the variants and the drugs, such as effect for variants, ORR, median PFS, median OS and adverse reaction for drugs.

g. The FDA database releases comprehensive and authoritative for approved drugs, but no detail information for gene variants can be found, such as effect, protein domains and pathways.

h. The Clinical Trials database provides clinical trial information from most countries in the world.

i. The NCCN database releases guidelines for clinical practice of malignant tumors, which includes therapy choice, such as how to choose drugs according to situations. But detail information for gene variants is relatively lack, such as effect and pathways.

Conclusion

The current public target drug annotation databases all have advantages and disadvantages. When doing annotations for specific genetic variants, users need to integrate the information of them. And we can see that there is no public tool to provide EGFR TKI target drug annotation of gene variants automatically for non-small cell lung cancer samples for the moment. This is a place worth exploring in the future.

References


