Identification of the key target profiles underlying the drugs of narrow therapeutic index for treating cancer and cardiovascular disease

Jiayi Yin\textsuperscript{a,1}, Xiaoxu Li\textsuperscript{a,1}, Fengcheng Li\textsuperscript{a}, Yinjing Lu\textsuperscript{a}, Su Zeng\textsuperscript{a,b,*}, Feng Zhu\textsuperscript{a,b,c,*}

\textsuperscript{a}College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
\textsuperscript{b}Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Hangzhou 310018, China
\textsuperscript{c}Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 16 December 2020
Accepted 15 April 2021
Available online 21 April 2021

\textbf{Keywords:}
Narrow therapeutic index
Cancer
Cardiovascular disease
Artificial intelligence
Drug mechanism

\textbf{A B S T R A C T}

An appropriate therapeutic index is crucial for drug discovery and development since narrow therapeutic index (NTI) drugs with slight dosage variation may induce severe adverse drug reactions or potential treatment failure. To date, the shared characteristics underlying the targets of NTI drugs have been explored by several studies, which have been applied to identify potential drug targets. However, the association between the drug therapeutic index and the related disease has not been dissected, which is important for revealing the NTI drug mechanism and optimizing drug design. Therefore, in this study, two classes of disease (cancers and cardiovascular disorders) with the largest number of NTI drugs were selected, and the target property of the corresponding NTI drugs was analyzed. By calculating the biological system profiles and human protein–protein interaction (PPI) network properties of drug targets and adopting an AI-based algorithm, differentiated features between two diseases were discovered to reveal the distinct underlying mechanisms of NTI drugs in different diseases. Consequently, ten shared features and four unique features were identified for both diseases to distinguish NTI from NNTI drug targets. These computational discoveries, as well as the newly found features, suggest that in the clinical study of avoiding narrow therapeutic index in those diseases, the ability of target to be a hub and the efficiency of target signaling in the human PPI network should be considered, and it could thus provide novel guidance in the drug discovery and clinical research process and help to estimate the drug safety of cancer and cardiovascular disease.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

\section{1. Introduction}

A narrow therapeutic index (NTI) of a drug implies that a tiny variation in the dosage of the drug might lead to treatment failure or severe adverse drug reactions [1–5]. It also hampers drug development since researchers have to conduct additional studies [6] to modify the compound structure, and some failures in drug research [7] are caused by the NTI of drug candidates. In the course of clinical research, some NTI drugs pose great risks in clinical use due to the lack of clear dose adjustment recommendations [4,8]. It is essential to start mitigation methods to avoid unfavorable traits or to potentially alter resources to alternative candidates by gaining an early consideration of the likely TI value of a certain drug [9–10]. Moreover, this is critical for avoiding clinical trials because TI with low indication specificity may be considered morally unacceptable [2]. Therefore, the molecular mechanisms of NTI drugs play a prominent role in pharmaceutical discovery and clinical research and help to estimate drug safety and efficacy [11].

However, it is complicated to determine and interpret the TI of a drug because this depends not only on the stage of development that affects the available data but also on the properties of the indications for which the drug is being developed [2,12]. A widely used concept of TI is the quantitative relationship between pharmacology and safety toxicology, but the definition of a therapeutic or toxic effect is highly dependent on different therapeutic and toxic effect types [3]. For example, imatinib can allow more toxicity with a smaller TI value when used in cancer in pursuit of higher pharmacological exposure, but there must be a larger and more reasonable TI value when used for pulmonary hypertension [2,13]. This adds complexity to the understanding of the molecular mechanisms of NTI drugs. In fact, of the 161 NTI drugs currently FDA approved, almost half of them belong to cancer and cardiovascular...
disease. Cancer is a group of diseases characterized by uncontrolled cell growth. The cardiovascular disease usually involves narrowed or blocked blood vessels, which can contribute to heart attack, angina, or stroke and is characterized by acute onset, critical condition, and rapid progression [14]. These observations suggest that there may be disease-specific pathology, resulting in different types of the disease each with its characteristics [15], and the molecular mechanisms of NTI drugs in different diseases may also exhibit large variations. Therefore, when designing drugs and conducting clinical research on these two types of diseases, it is necessary to consider the different molecular mechanisms of NTI drugs between them.

To enhance the understanding of TI, a variety of studies and some approaches have been developed to enhance the ability to reveal the mechanism underlying NTI drugs, such as the exposure-centric TI approach [2], preclinical pharmacology model [16–18], assessment of off-target safety margins [19]. Recently, an article was published in Frontiers in Pharmacology [20] using a target-based approach, combining the profiles of human protein–protein interaction (PPI) network, and biological systems to find features or feature groups that can be used to indicate the drug's narrow TI. It identified 8 features that could collectively indicate that NTI drug targets are tremendously connected and centralized and are related to target drugability in all diseases. Agnieszka Pote˛ga, et al. [21] have shown that this target-based approach to study the mechanisms underlying NTI drugs is important to indicate a well-balanced profile between efficacy and safety. However, no studies have revealed the underlying mechanism behind the complex definition and interpretation of TI in different diseases, and what significantly limits NTI drug design and clinical studies for both specific diseases, and this needs to be solved urgently.

Therefore, in this study, the underlying mechanisms of NTI drugs aimed at cancer and cardiovascular disease were analyzed based on not only the human PPI network features but also the biological system profiles. To discover this underlying mechanism, the NTI and NNTI drug targets were divided into three groups: (i) NTI drug targets of cancer, (ii) NTI drug targets of cardiovascular disease, and (iii) NNTI drug targets for all indications. Next, through the comparative analysis of the target groups (i) and (iii) and the target groups (ii) and (iii), several essential features that could distinguish the two groups were identified, and further studies revealed similarities and differences in the characteristics of cancer and cardiovascular disease. Overall, these findings combined with the newly recognized features can indicate the underlying mechanisms of NTI drugs targeting cancer and cardiovascular disease, respectively, which offer certain guidance in assessing the risks and benefits of drug candidates, as well as drug discovery and clinical research in cancer and cardiovascular disease.

2. Materials and methods

2.1. NTI drugs collection and associated targets and indications identification

The NTI approved drugs and their related drug targets and indications were obtained through the following steps. First, 1,921 FDA approved drugs with their related indications were systematically collected and identified from the orange book of the US FDA [72]. Then, all the corresponding diseases were standardized by the ICD-11 codes (the latest version of the International Classification of Diseases) [73]. Next, the corresponding targets of the approved drugs were authorized by the therapeutic target database (TTD) [74], and 506 corresponding targets of the approved drugs were confirmed. Third, a systematic literature review of all these drugs was performed to confirm their TI value by searching the PubMed database using such keyword combinations as “Drug Name/Synonym” + “Therapeutic ranges” / “Therapeutic index” / “Therapeutic ration” / “Therapeutic window”. Consequently, 36 NTI drugs targeting cancer and 18 NTI drugs targeting cardiovascular disease are discovered, which account for approximately half of all NTI drugs. Moreover, 29 NNTI drugs targeting all indications are also distinguished. The FDA-approved NTI drugs for cancer and cardiovascular disease together with their standardized indication, ICD-11 codes, and targets are provided in Table 1, and the NNTI drugs for all indication together with their standardized indication, ICD-11 codes, and targets are provided in Table 2.

2.2. Assessing the profile of human PPI network properties and biological systems for corresponding therapeutic targets

The human PPI network properties studied in this research consisted of 15,554 proteins and 642,304 interactions between these proteins, and these were created via the information furnished by the STRING database [75]. Only those protein interactions with confidence above 0.95 were selected for further analysis to guarantee the dependability of the analytical data [76–77]. Thus, in this study, a subnetwork consisting of 8,509 proteins, and 40,468 interactions between these proteins was developed for subsequent study. Additionally, the PPI network characteristics of corresponding therapeutic targets were obtained by the PROFEAT [78] and the tool Network Analyzer of Cytoscape [79–80]. In summary, 32 PPI network properties were calculated for further analysis, as shown in Table 3 (six features that are all zeroes were deleted, which are: ‘closeness centrality sum’, ‘bridging centrality’, ‘eigenvector centrality’, ‘page rank centrality’, ‘number of selfloops’, and ‘current flow closeness’). Then, the additional four features of each corresponding target in the biological system profile were estimated.

The first feature is the number of target-affiliated pathways that were collected from the KEGG database [81]. This feature was confirmed by two aspects. On the one hand, the pathway of the corresponding drug targets should be necessary for life not only for patients but also in healthy individuals. On the other hand, the therapeutic target should be upstream and have the ability to regulate the biological function of the pathway. The second feature is the number of each therapeutic drug target distributed in human tissues, which was offered in the TissueDistributionDBs [82] and UniProt [83] databases. The determination of this feature depends on a higher level of total protein (>5%) distributed in a particular tissue or a higher target concentration in that tissue than the average protein concentration. To explore the off-target collateral effect, the third feature was adopted, which is the number of human similarity proteins. This was determined by counting the number of similar proteins that are outside the target protein family for the studied drug target [84–85]. This was calculated using BLAST similarity screening with the cutoff value of e-value < 0.005 [86–87] for the human proteome method furnished in the UniProt database [83]. The differential expression of the target is the fourth feature, which is capable of reflecting the expression differences of the corresponding target between diseased and healthy populations for specific diseases [74,88–89]. The expression data were gathered from TTD [90] and calculated by using the HG-U133 Plus 2.0 platform which was determined by the Gene Expression Omnibus database [91].

Collectively, these 36 features are valuable and meaningful in revealing human protein–protein interaction data for a given target, including their connectivity, organization, robustness, and stability in the human PPI network [92–94] and the on-target and off-target pharmacology of the studied targets [85,95]. These two aspects are key to enhancing potency for characterizing the underlying mechanisms of NTI drugs [2,96]. In previous publications, including our previous analysis [20],
The Table 1

| FDA Approved Drug (Reference for NTI) | Time of Approval | FDA Approved Indication | KDC-11 Code | Disease Class | Target Name |
|--------------------------------------|------------------|-------------------------|-------------|---------------|-------------|
| Argatroban [22]                      | 2000             | Intracardiac thrombosis | BC46        | Cardiovascular | F2          |
| Axitinib [23]                        | 2012             | Rectum cancer           | 2B92        | Cancer        | KDR         |
| Busulfan [24]                        | 1954             | Chronic myeloid leukemia | 2B33       | Cancer        | hDNA        |
| Capcetabine [25]                     | 1998             | Breast cancer           | 2C50        | Cancer        | TMPI        |
| Carboplatin [26]                     | 1989             | Ovary cancer            | 2C73        | Cancer        | hDNA        |
| Cisplatin [27]                       | 1978             | Ovary cancer            | 2C73        | Cancer        | hDNA        |
| Clonidine [28]                       | 1974             | Hypertension            | BA00        | Cardiovascular | ADR2       |
| Cyclophosphamide [29]                | 1959             | Acute myeloid leukemia  | 2A60        | Cancer        | hDNA        |
| Dalbavancin Sodium [22]              | 2004             | Dendritic thrombosis    | BD71        | Cancer        | ADR2ADMA    |
| Dextran [30]                         | 1980             | Heart failure           | BD10        | Cardiovascular | SPT ATPase |
| Digoxin [31]                         | 1954             | Heart failure           | BD10        | Cardiovascular | SPT ATPase |
| Disopyramide Hydrochloride [32]      | 1977             | Ventricular tachyarrhythmia | BC71       | Cardiovascular | SCNSA       |
| Doceadera [33]                       | 1996             | Breast cancer           | 2C50        | Cancer        | TOP2        |
| Dorzolamide HCl [34]                 | 1974             | Breast cancer           | 2C50        | Cancer        | TOP2        |
| Epinephrine [35]                     | 1951             | Coronary artery disease | BA00        | Cardiovascular | ADRB1      |
| Epirubicin HCl [36]                  | 1999             | Axillary node cancer    | 2D60        | Cancer        | TOP2        |
| Etoposide Phosphate [26]             | 1983             | Testis cancer           | 2C80        | Cancer        | TOP2        |
| Everolimus [38]                      | 2009             | Ovary cancer            | 2C73        | Cancer        | TOP2        |
| Flecainide Acetate [39]              | 1985             | Arrhythmic              | BC64        | Cardiovascular | SCNSA       |
| Fluorouracil [40]                    | 1962             | Colonecral cancer       | 2B91        | Cancer        | TMI         |
| Fondaparinux Sodium [41]             | 2001             | Deep vein thrombosis    | BD71        | Cardiovascular | F10         |
| Gefitinib [42]                       | 2003             | Lung cancer             | 2C25        | Cancer        | EGFR        |
| Gemcitabine HCl [43]                 | 1996             | Pancreatic cancer       | 2C10        | Cancer        | BMRI2       |
| Guanethidine Monosulfate [22]        | 1960             | Hypertensive crisis     | BA03        | Cardiovascular | NET         |
| Interferon Alfa-2B [44]              | 1986             | Melanoma                | 2C30        | Cancer        | IFNA2       |
| Irinotecan HCl [34]                  | 1996             | Colonocral cancer       | 2B91        | Cancer        | TOP1        |
| Lidocaine HCl [45]                   | 1948             | Ventricular tachyarrhythmia | BC71       | Cardiovascular | SCNSA       |
| Mercaptopurine [46]                  | 1953             | Acute lymphocytic leukemia | 2A82       | Cancer        | IMPDH1      |
| Methotrexate Sodium [26]             | 1953             | Breast cancer           | 2C60        | Cancer        | DHFR        |
| Mitomycin [47]                       | 1981             | Stomach cancer          | 2B72        | Cancer        | hDNA        |
| Mitotane [48]                        | 1970             | Adrenal gland cancer    | 2D11        | Cancer        | ESR         |
| Paclitaxel [40]                      | 2002             | Colonocral cancer       | 2B91        | Cancer        | hDNA        |
| Paclitaxel [50]                      | 1992             | Kaposi sarcoma          | 2B57        | Cancer        | TUB; BCL-2  |
| Pazopanib HCl [23]                   | 2009             | Renal cell carcinoma    | 2C90        | Cancer        | c-Kit; KDR; PDGFRB |
| Pemetrexed [51]                      | 2004             | Pleuera cancer          | 2C26        | Cancer        | DHFR; TMPI  |
| Pemetrexed Disodium [51]             | 2004             | Pleuera cancer          | 2C26        | Cancer        | DHFR; TMPI  |
| Phenprocoumon [52]                   | 1957             | Intracardiac thrombosis | BC46        | Cardiovascular | VKORC1      |
| Prasozin HCl [22]                    | 1976             | Hypertension            | BA00        | Cardiovascular | ADR3       |
| Procainamide HCl [53]                | 1950             | Ventricular tachyarrhythmia | BC71       | Cardiovascular | BTK-activated cardiac |
| Propafenone HCl [22]                 | 1989             | Atrial fibrillation     | BC81        | Cardiovascular | ADRB1; ADRB2; ADRB3 |
| Quinidine [49]                       | 1950             | Ventricular tachyarrhythmia | BC71       | Cardiovascular | SCNSA       |
| Regorafenib [23]                     | 2012             | Gastrointestinal stromal cancer | 2B58      | Cancer        | c-Kit; KDR; RET |
| Sorafenib Tozylate [23]              | 2005             | Adrenal gland cancer    | 2D11        | Cancer        | EGFR; c-Kit; KDR; PDGFRB |
| Sotalol HCl [22]                     | 1992             | Ventricular tachyarrhythmia | BC71       | Cardiovascular | ADRB1      |
| Sunitinib Malate [23]                | 2006             | Gastrointestinal stromal cancer | 2B85      | Cancer        | KDR         |
| Teniposide [26]                      | 1992             | Acute lymphopcytic leukemia | 2A82       | Cancer        | TOP2        |
| Thioguanine [46]                     | 1966             | Acute myeloid leukemia  | 2A60        | Cancer        | hDNA        |
| Topotecan HCl [26]                   | 1996             | Ovary cancer            | 2C73        | Cancer        | TOP1        |
| Vandetanib [23]                      | 2011             | Thyroid gland cancer    | 2D10        | Cancer        | EGFR; KDR; RET |
| Vinblastine Sulfate [39]             | 1965             | Hodgkin lymphoma        | 2B30        | Cancer        | TUB         |
| Vincristine Sulfate [54]             | 1963             | Acute lymphocytic leukemia | 2A62       | Cancer        | TUB         |
| Vinorelbine Tartrate [55]            | 1994             | Lung cancer             | 2C25        | Cancer        | TUB         |
| Warfarin Sodium [31]                 | 1954             | Pulmonary thromboembolism | BB00       | Cardiovascular | VKORC1      |

Table 2

| Drug Target | Biological Action |
|-------------|-------------------|
| SCN5A       | Voltage-gated sodium channel alpha Nav1.5 |
| SCN11A      | Voltage-gated sodium channel alpha Nav1.9 |
| EGFR        | Epidermal growth factor receptor |
| IFNA2       | Interferon alpha-2 |
| KDR         | Tyrosine-protein kinase Kit |
| mTOR        | Serine/threonine-protein kinase |
| NET         | Norepinephrine transporter |
| PDGFRB      | Platelet-derived growth factor receptor |
| RRM2        | Ribonucleoside-diphosphate reductase |
| mTOR        | Serine/threonine-protein kinase |

a series of analyses have been performed by these 36 features. And these 36 features (30 features are described in Table 3, excluding the 6 features that the calculated values are zero) are still adopted in this study to further explore the different features of NTI drug targets between two representative disease classes (cancers and cardiovascular diseases). Their calculation formulas and biological descriptions are separately reflected in Supplementary Table S1.

2.3. NTI drug characteristic identification in two diseases by an artificial intelligence-based algorithm

Artificial intelligence (AI) has seen significant advancement in recent decades for aiding drug treatment [97–101], predicting drug-target or drug-drug interactions [5,102–103], and optimizing treatment protocols [104–106], including machine learning algorithms [107–109], deep learning methods [110–112], and
cognitive-computing [113]. In this study, to better understand the underlying mechanisms of NTI drugs, one of the most widely used artificial intelligence algorithms, Boruta, which was based on a random forest classifier [18,114], was adopted. This method compares artificial intelligence algorithms, Boruta, which was based on a random forest classifier, with the performance of other algorithms. However, none of them can be used separately as an important index to identify target drug features that can be combined to further explore the different features of NTI drug targets between two representative diseases (cancer and cardiovascular disease). Their calculation formulas and biological descriptions are separately reflected in Supplementary Table S1. The average and median values of 30 features for cancer NTI drug targets, cardiovascular disease NTI drug targets, and NNTI drug targets were also calculated (removing six characteristics equal to 0), as shown in Table 3. These 30 features were classified into three categories according to the attributes inherent in each feature, that is, the connectivity/adjacency-based properties, the shortest path length-based properties, and the human biological system properties, as also shown in Table 3.

The mean and median values between the two groups of targets (NTI and NNTI drug targets) for each disease in Table 3 show a significant difference between the two groups of targets in many features. However, none of them can be used separately as an indicator to distinguish between NTI drug targets and NNTI drug targets. Only through collective combination can NTI drug targets be more effectively distinguished from NNTI drug targets [20]. Therefore, in the next part of the study, we integrated the feature selection method based on artificial intelligence to select some important indexes from these features that can be combined to determine the drug targets of NTI and the drug targets of NNTI. However, this approach seems to introduce a very strong bias when 36 features are directly used for feature selection because of the significant dependence between 19 of these features [20].

3. Results and discussion

3.1. Merging the human PPI network and biological system properties for artificial intelligence-based algorithm

The drug risk-to-benefit ratio (RBR) is mainly determined by the drug target profile of the network properties and biological system [84,119–121]. Network characteristics are inherent to drug targets in human PPI networks, and biological system properties can mirror the pharmacology of on-target and off-target. In this paper, the most comprehensive sets of characteristics belong to the human PPI network properties and biological system profiles were chosen to further explore the different features of NTI drug targets between two representative diseases (cancer and cardiovascular disease). Their calculation formulas and biological descriptions are separately reflected in Supplementary Table S1. The average and median values of 30 features for cancer NTI drug targets, cardiovascular disease NTI drug targets, and NNTI drug targets were also calculated (removing six characteristics equal to 0), as shown in Table 3. These 30 features were classified into three categories according to the attributes inherent in each feature, that is, the connectivity/adjacency-based properties, the shortest path length-based properties, and the human biological system properties, as also shown in Table 3.

The mean and median values between the two groups of targets (NTI and NNTI drug targets) for each disease in Table 3 show a significant difference between the two groups of targets in many features. However, none of them can be used separately as an indicator to distinguish between NTI drug targets and NNTI drug targets. Only through collective combination can NTI drug targets be more effectively distinguished from NNTI drug targets [20]. Therefore, in the next part of the study, we integrated the feature selection method based on artificial intelligence to select some important indexes from these features that can be combined to determine the drug targets of NTI and the drug targets of NNTI. However, this approach seems to introduce a very strong bias when 36 features are directly used for feature selection because of the significant dependence between 19 of these features [20].

Table 2

| FDA Approved Drug (Reference for NTI) | Time of Approval | FDA Approved Indication | ICD-11 Code | Disease Class | Target Name |
|-------------------------------------|------------------|-------------------------|-------------|--------------|-------------|
| Apixaban [56]                       | 2012             | Deep vein thrombosis    | BD71        | Cardiovascular | F10         |
| Aripiprazole [57]                   | 2002             | Schizophrenia           | 6A20        | Mental disorder | D2R         |
| Atorvastatin HCl [58]               | 2002             | Hyperlipidemia          | 6A05        | Metabolic disorder | NET         |
| Clopidogrel [59]                    | 2011             | Epilepsy or seizures    | 8A60        | Nervous system | GABRA1; GABRG3 |
| Clonazepam [22]                     | 1975             | Epilepsy or seizures    | 8A60        | Nervous system | GABRA1      |
| Enalapril Maleate [60]              | 1985             | Hypertension            | 8A00        | Cardiovascular | ACE         |
| Ethosuximide [22]                   | 1960             | Epilepsy or seizures    | 8A60        | Nervous system | CACNA1G     |
| Ezogabine [59]                      | 2011             | Epilepsy or seizures    | 8A60        | Nervous system | KCNQ2; KCNQ3 |
| Felbamate [22]                      | 1993             | Epilepsy or seizures    | 8A60        | Nervous system | NMDAR       |
| Gabapentin [61]                     | 1993             | Epilepsy or seizures    | 8A60        | Nervous system | CACNA2D2; CACNA2D3 |
| Gabapentin Enacarbil [61]           | 1993             | Epilepsy or seizures    | 8A60        | Nervous system | CACNA2D2; CACNA2D3 |
| Lacosamide [62]                     | 1990             | Epilepsy or seizures    | 8A60        | Nervous system | DPPSL2      |
| Lamivudine [63]                     | 1995             | HIV infection           | 1G62        | Infection     | HIV RT      |
| Lamotrigine [22]                    | 1994             | Bipolar disorders       | 6A60        | Mental disorder | SCN11A     |
| Leviteracetam [22]                  | 1999             | Epilepsy or seizures    | 8A60        | Nervous system | SV2A        |
| Linagliptin [64]                    | 2011             | Type 2 diabetes mellitus | 5A11    | Metabolic disease | DPP4      |
| Mechlorethamine HCl [65]            | 1949             | Multiple sclerosis      | 8A40        | Nervous system | TOF2        |
| Mitoxantrone HCl [66]               | 1987             | Multiple sclerosis      | 8A40        | Nervous system | TOF2        |
| Montelukast Sodium [67]             | 1998             | Asthma                  | CA23        | Respiratory system | CYSLTR1 |
| Oxcarbazepine [22]                  | 2000             | Epilepsy or seizures    | 8A60        | Nervous system | SCN11A      |
| Perampanel [56]                     | 2012             | Epilepsy or seizures    | 8A60        | Nervous system | GRIA        |
| Pimecrolimus [68]                   | 2001             | Atopic eczema           | EA80        | Skin disease  | PPP3CA      |
| Pregabalin [69]                     | 2004             | Epilepsy or seizures    | 8A60        | Nervous system | CACNA2D1    |
| Rivaroxaban [59]                    | 2011             | Deep vein thrombosis    | BD71        | Cardiovascular | F10         |
| Rolapitant HCl [70]                 | 2015             | Nausea or vomiting      | DD90        | Digestive system | TACR1     |
| Rufinamide [62]                     | 2008             | Epilepsy or seizures    | 8A60        | Nervous system | N.A.        |
| Topiramate [22]                     | 1996             | Epilepsy or seizures    | 8A60        | Nervous system | GABRA1      |
| Vigabatrin [71]                     | 2009             | Types of seizures       | 8A68        | Nervous system | ABAT        |
| Zonisamide [22]                     | 2000             | Epilepsy or seizures    | 8A60        | Nervous system | SCN1A       |
Therefore, after a thorough investigation of 36 features, the 19 features were eventually merged into five features due to their innate independence. Considering the remaining 17 relatively independent features, a total of 22 features for each target were applied for further feature selection. The method of feature integration referred to previous research by our group [20], as well as the biological description and equation of those 36 properties in human PPI networks and biological system profiles, provided in Supplementary Table S1.

### 3.2. Revealing the essential properties of NTI drug targets in cancer by an artificial intelligence-based algorithm

The Boruta algorithm was built by an artificial intelligence method, which is particularly suitable for low-dimensional data sets compared to other available strategies because of its strong stability in variable selection [20,116]. Specifically, by setting the key parameters (described in detail in the Materials and Methods section), the R package Boruta was used to selecting the key difference features from 22 target profiles. The feature selection result is shown in Fig. 1, which means that 13 features were selected to collectively reflect the underlying mechanism of cancer NTI drugs, including ‘interconnectivity’, ‘bridging coefficient’, ‘average shortest path length’, ‘average closeness centrality’, ‘radiality’, ‘topological coefficient’, ‘number of affiliated pathways’, ‘number of similarity proteins’, ‘stress’, ‘number of tissues’, ‘degree’, ‘neighborhood connectivity’, and ‘number of triangles’. The violin districts colored dark blue and light blue refer to the NTI drug targets in cancer and NTTI drug targets in all indications, respectively. Among these 13 selected features, some important features
displayed an upward trend from NTI to NNTI drug targets (such as ‘interconnectivity’), while others showed a downward trend (such as ‘average closeness centrality’). In particular, the ‘average closeness centrality’ is defined as the reciprocal of the average shortest path length of the studied target. It measures how fast information spreads from a studied drug target to other reachable proteins in the PPI network [122], and the ‘interconnectivity’ is a connection metric that indicates the quality or status of the studied targets connected [123]. It was reported that a higher value of ‘average closeness centrality’ and a higher level (lower value) of ‘interconnectivity’ of the target demonstrated a greater lethality risk [20,124], which meant that a protein with tremendous centrality and connectivity carries a greater lethality risk. The results from our study proved that the capabilities of the applied Boruta algorithms in determining essential features of cancer NTI drug targets were due to the trends of the values of features in NTI and NNTI in Fig. 1, in agreement with these previous studies. Moreover, what we found also suggested that some features could be indirectly relevant to the drug risk-to-benefit ratio [124–125], and NTI drug targets of cancer in the biological network were not only inclined to be hub proteins [126] but also to have high centrality and connectivity.

3.3. Discovering the basic characteristics of NTI drug targets in cardiovascular disease by artificial intelligence-based algorithm

To identify the features of NTI drugs treating cardiovascular disease, the Boruta algorithm was adopted. The results are shown in Fig. 2. Eleven features were selected to collectively reflect the underlying mechanism of cardiovascular disease NTI drugs, including ‘interconnectivity’, ‘bridging coefficient’, ‘average shortest path length’, ‘average closeness centrality’, ‘radiality’, ‘number of affiliated pathways’, ‘number of similarity proteins’, ‘topological coefficient’, ‘stress’, ‘number of tissues’, and ‘distance deviation’. In Fig. 2, the violin districts colored dark orange and light orange refer to the NTI drug targets in cardiovascular disease and NNTI drug targets in all indications, respectively. Similar to drug targets in cancer, some important features displayed an upward trend from NTI to NNTI drug targets (such as ‘average shortest path length’). In contrast, some displayed a downward trend (such as ‘radiality’). The ‘average shortest path length’ describes the average length of shortest paths between the studied drug target and all other proteins in the studied PPI network [127], and the ‘radiality’ is the reachability level of the studied nodes through diverse shortest paths throughout the network [128]. Moreover, the trend of these features in NTI and NNTI drug targets meant that NTI drug targets of cardiovascular disease were likely to have more links with other proteins. [129].

3.4. Exploring the shared/differential characteristics of NTI drug targets between cancers and cardiovascular diseases

The shared/differential features of NTI drug targets between cancers and cardiovascular diseases identified in the study are provided in Fig. 3. The boxes of pink background are the feature class of Connectivity, the boxes of light green background are the feature class of Centrality, and the box of yellow background provides the feature class of Biological System Profile. Besides, the dark blue bars indicated the characteristics of NTI drug targets for cancers, and the orange bars denoted the characteristics of NTI drug targets for cardiovascular diseases. Those 10 features in the first layer are shared by both cancers and cardiovascular diseases. Seven of these 10 features are the same as those identified by the previous report [20], which include ‘average shortest path length’, ‘bridging coefficient’, ‘closeness centrality’, ‘interconnectivity’, ‘number of affiliated pathways’, ‘number of similarity proteins’ and ‘radiality’. These features indicated that the NTI drug targets of cancers and cardiovascular diseases were greatly connected and centralized in human PPI networks, and shared a biological system of the large number of similar proteins and target-affiliated pathways [20]. Moreover, these results validated the capability of Boruta...
in determining the features of NTI drug targets in different disease classes.

The second layer is unique characteristics identified for cancers (dark blue) and cardiovascular diseases (orange). For cancers, the identified features included 'neighborhood connectivity', 'degree', and 'number of triangles'. For cardiovascular diseases, the discovered feature was 'distance deviation'. As reported, drug targets tend to have a lower 'clustering coefficient' in cancer [15]. The 'clustering coefficient' denoted neighborhood connectivity [130], and the 'clustering coefficient' decreased with the increase in the number of affiliated pathways.
number of interacting proteins [131]. Therefore, a lower ‘clustering coefficient’ indicates higher ‘neighborhood connectivity’. The ‘degree’ means the total number of edges connected to the studied node, and the ‘number of triangles’ that referred to the percentage of the triangle between a node and its neighbors. The higher the ‘degree’ and ‘number of triangles’, the higher the centrality of the drug target, and the more likely it is to lead to adverse drug reactions [132]. Thus, the computational discoveries of our result that cancer has a higher value of ‘degree’, ‘number of triangles’, and ‘neighborhood connectivity’ are consistent with literature reports. These phenomena indicate that the molecular mechanisms underlying of NTI drug targets in cancer require greater attention for the higher level of ‘degree’, ‘number of triangles’, and ‘neighborhood connectivity’.

3.5. Clinical implication of the identified features underlying NTI drug targets

Based on the above analyses, a total of ten features from three feature groups were identified as common features for both disease classes (cancers & cardiovascular diseases) in distinguishing NTI from NNTI drug targets, and there were another 4 features from two feature groups that were singled out by one of those 2 disease classes. Those shared feature groups identified in this study were consistent with our previous publication [20], which reaffirmed the importance of these shared features in differentiating NTI drug targets from NNTI ones. Since the vast majority of all NTI drugs were from those two disease classes (cancers & cardiovascular diseases), it was not surprising to have such similarity in the shared feature groups. Such shared features could also provide a new direction for optimizing the drug efficacy-safety balance [20]. Particularly, the importance of these shared features in the prediction of drug-induced hepatotoxicity has already been reported by the previous publication [133].

More importantly, those features that were unique in different disease classes were concentrated in two feature classes of Connectivity and Centrality. Particularly, the features unique to cancer included ‘degree’, ‘neighborhood connectivity’, and ‘number of triangles’. The ‘degree’ denoted the number of proteins in the human PPI network that interacted with the studied drug target [134]. The ‘neighborhood connectivity’ indicated the average number of interacting proteins of all the studied drug target’s neighbors [135]. The ‘number of triangles’ showed the number of triangles that included the studied target as a vertex [136], and this triangular relationship includes the studied drug target and its interacting proteins, as well as the interactions among the interacting proteins. In fact, these three features could collectively represent whether the studied drug target acted as a hub in the human PPI network, the higher their values the stronger the core position of the studied targets [137]. These findings represent an emphasis by cancers in differentiating NTI drug targets with respect to the target’s ability to be a hub in the PPI network. In other words, the narrow therapeutic index of an anticancer drug may originate from its interaction with hub protein [138–139]. To improve the situation of the narrow therapeutic index of anticancer drugs, it is necessary to impose more requirements on target selection. The hub proteins in the human PPI network should be avoided when designing anticancer drugs.

Different from cancers, the unique feature singled out in cardiovascular diseases is ‘distance deviation’, which belongs to the connectivity feature group. The ‘distance deviation’ indicated the absolute difference between the sum of all shortest paths starting from the studied target to all other proteins and the mean shortest path length of all the proteins in the human PPI network [135]. This implied an emphasis on the efficiency of inter-target signaling in NTI drug targets for cardiovascular diseases [135,140], which may indicate the needs for a more in-depth study of target signaling pathways when designing drugs for cardiovascular diseases. All in all, this study identified the key target features indicating the NTI drugs for cancers and cardiovascular diseases, which has great clinical implications in the drug designs for both disease classes.

4. Conclusion

This work is the first practice to reveal the underlying mechanism behind the complex definition and interpretation of NTI between different disease classes. Ten shared and four unique features were identified for both disease classes (cancers & cardiovascular diseases) to distinguish NTI drug targets from NNTI ones. This work suggested that in the clinical study of avoiding narrow therapeutic index in those diseases, the ability of target to be a hub and the efficiency of target signaling in the human PPI network should be especially considered.

Author contributions

Feng ZHU, Su ZENG conceived the idea and supervised the work. Jiaji YIN, Xiaoxiu LI performed the research. Jiaji YIN, Xiaoxiu LI and Fengcheng LI prepared and analyzed the data. Jiaji YIN, Xiaoxiu LI, Yingjing LU, wrote the manuscript. All authors have read and approved this manuscript.

Funding

Funded by National Natural Science Foundation of China (81872798 & U1909208); Natural Science Foundation of Zhejiang Province (LR21H300001); National Key R&D Program of China (2018YFC0910500); Leading Talent of the ‘Ten Thousand Plan’ – National High-Level Talents Special Support Plan of China; Fundamental Research Fund for Central Universities (2018QNA7023); Key R&D Program of Zhejiang Province (2020C03010). This work was supported by Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; Information Technology Center of Zhejiang University.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.035.

References

[1] Yu LX, Jiang W, Zhang X, Lionberger R, Makhloof F, et al. Novel bioequivalence approach for narrow therapeutic index drugs. Clin Pharmacol Ther 2015;97 (3):286–91.
[2] Muller PY, Milton MN. The determination and interpretation of the therapeutic index in drug development. Nat Rev Drug Discov 2012;11 (10):751–61.
[3] Tamargo J, Le Heuzey JY, Mabo P. Narrow therapeutic index drugs: a clinical pharmacological consideration to flecaainide. Eur J Clin Pharmacol 2015;71 (5):549–67.
[4] Krens SD, Laxxhe G, Janssen FGA, Desar IME, et al. Dose recommendations for anticancer drugs in patients with renal or hepatic impairment. Lancet Oncol 2019;20(4):e200–7.
[5] Chen L, Chu C, Zhang Y-H, Zheng M, Zhu L, et al. Identification of drug-drug interactions using chemical interactions. Curr Bioinform 2017;12(6):526–34.
[6] Han H, Jain AD, Truica MI, Izquierdo-Ferrer J, Anker JF, et al. Small-molecule MYC inhibitors suppress tumor growth and enhance immunotherapy. Cancer Cell 2019;36(5):483–97.
