A bioinformatics approach for precision medicine off-label drug selection among triple negative breast cancer patients

RECEIVED 5 September 2015
REVISED 10 December 2015
ACCEPTED 10 January 2016
PUBLISHED ONLINE FIRST 23 April 2016

Lijun Cheng,[1,2] Bryan P Schneider,[3,4] and Lang Li[1,2,4]

ABSTRACT

Background Cancer has been extensively characterized on the basis of genomics. The integration of genetic information about cancers with data on how the cancers respond to target based therapy to help to optimum cancer treatment.

Objective The increasing usage of sequencing technology in cancer research and clinical practice has enormously advanced our understanding of cancer mechanisms. The cancer precision medicine is becoming a reality. Although off-label drug usage is a common practice in treating cancer, it suffers from the lack of knowledge base for proper cancer drug selections. This eminent need has become even more apparent considering the upcoming genomics data.

Methods In this paper, a personalized medicine knowledge base is constructed by integrating various cancer drugs, drug-target database, and knowledge sources for the proper cancer drugs and their target selections. Based on the knowledge base, a bioinformatics approach for cancer drugs selection in precision medicine is developed. It integrates personal molecular profile data, including copy number variation, mutation, and gene expression.

Results By analyzing the 85 triple negative breast cancer (TNBC) patient data in the Cancer Genome Altar, we have shown that 71.7% of the TNBC patients have FDA approved drug targets, and 51.7% of the patients have more than one drug target. Sixty-five drug targets are identified as TNBC treatment targets and 85 candidate drugs are recommended. Many existing TNBC candidate targets, such as Poly (ADP-Ribose) Polymerase 1 (PARP1), Cell division protein kinase 6 (CDK6), epidermal growth factor receptor, etc., were identified. On the other hand, we found some additional targets that are not yet fully investigated in the TNBC, such as Gamma-Glutamyl Hydrolase (GGH), Thymidylate Synthetase (TYMS), Protein Tyrosine Kinase 6 (PTK6), Topoisomerase (DNA) I, Mitochondrial (TOP1MT), Smoothened, Frizzled Class Receptor (SMO), etc. Our additional analysis of target and drug selection strategy is also fully supported by the drug screening data on TNBC cell lines in the Cancer Cell Line Encyclopedia.

Conclusions The proposed bioinformatics approach lays a foundation for cancer precision medicine. It supplies much needed knowledge base for the off-label cancer drug usage in clinics.

Keywords: precision medicine, drug selection, bioinformatics

RESEARCH AND APPLICATIONS

BACKGROUND

The landscape of cancer genomics reveals that various cancer types, although having different organ originalities, share many driving mutagenesis mechanisms and their corresponding molecular signaling pathways in several core cellular processes, such as cell fate, cell survival, and genome maintenance.[1] This powerful pan-cancer discovery guides cancer biology research on these essential pathways, while more challenging questions are waiting for answers, such as unexplained drug resistance and tumor recurrence, genomic tumor heterogeneity, and limited knowledge of the drugs and their combination usage. These cancer precision medicine research topics are highly advocated in the recent vision statement and National Institutes of Health of United State Initiative by Drs Collins and Varmus[2] in 2015. In their report, novel clinical trial designs and new bioinformatics tools have also been identified as top prioritized methodology research fields in cancer precision medicine.

Matching the right drugs to the right patients is the primary goal of precision medicine. In the early precision medicine cancer trials, such as Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis[3] and Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination,[4] either multi-gene transcription signatures or candidate genetic mutation or protein expression biomarkers were used to adaptively select a proper cancer patient sub-population for pre-specified drug therapies. Recently, novel precision medicine clinical trials, such as Molecular Analysis for Therapy Choice Trial[5] and Targeted Agent and Profiling Utilization Registry,[6] conducted sequencing for both drug target identification and drug selections. One important rational reason behind the recent trial is to allow patients from various cancer types that share the same targets to receive the same drug therapy. This pan-cancer target identification is well supported by the strong pan-cancer mutations[7] and copy number (CN) alteration[8] patterns observed from the data in the triple Cancer Genome Altar (TCGA) project. Another critical rational reason that drives the precision medicine cancer therapy, either in the clinical practice or trials, is the popular off-label drug use.[9] Although off-label drug use has its potential regulatory[10] challenge and a risk of reimbursement,[11] it has been a common medical practice for both patients and physicians. When some aggressive cancer types have only limited therapeutic options, the off-label drug use will provide much needed options, and become particularly appealing. However, the primary challenge of the off-label drug use is the lack of preclinical and clinical evidence, and guidance for drug selection.[12] With the available TCGA data in the clinical tumor samples and other large-scale cancer cell line drug screening data, such as the Cancer Cell Line Encyclopedia (CCLE),[13] we anticipate that the bioinformatics data integration and analysis will make up this knowledge gap.

Recent bioinformatics research has attempted to develop computational models, which translate the genomic signatures derived from cell line drug screening data to guide drug selections for cancer patients.[14] However, these models through have their own theoretical

Correspondence to Lang Li, PhD, 410 West 10th Street, HITS 5001, Indianapolis, Indiana 46202, USA; lali@iu.edu. For numbered affiliations see end of article.

value, have not yet been designed to address practical needs in target and drug database integration, target selection, and drug-target matching. In this paper, we will use triple negative breast cancer (TNBC) as an example, to build up a roadmap for its precision medicine. This patient population, by definition, is resistant to traditional nonspecific cytotoxic chemotherapy. In addition, by nature of the high risk of relapse, they are clearly in need of additional (and hopefully superior) targeted therapy. Using bioinformatics approaches, we will address several important questions: what are the FDA approved targeted drug therapies for cancer? What are those targeted drugs that can be in off-label use for TNBC patients? Are there any TNBC cell studies and evidence that support the off-label use targeted drug therapies?

METHODS

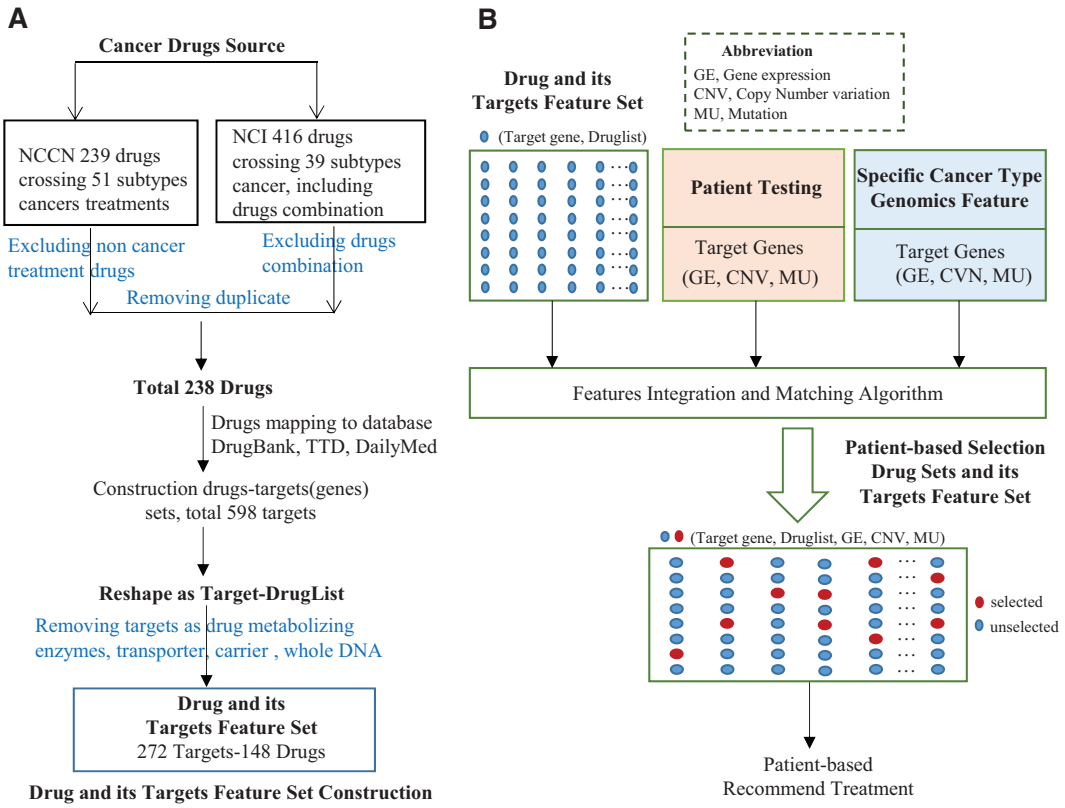
Overview: in our proposed drug-selection computational model for precision medicine, the cancer treatment decision is tailored toward an individual patient. A targeted cancer treatment is recommended based on a patient's genomic and transcriptomic profile and computational searching strategy. Our proposed model is composed of four parts: the information input of an individual cancer patient, the drug-target relationship knowledge dataset, the cancer type specific background knowledge, and the feature integration and matching algorithm for drug-targeted selection. Figure 1B shows how these four parts work together. Using the TNBC as an example, we show how this model works in detail. FDA approved cancer drugs and their targets dataset construction, TNBC individual patients, and patient-specific target and drug selection implementation are shown in the following.

Cancer Drugs Collection. Cancer therapy has many forms: surgery, radiation therapy, chemotherapy, target-based therapy, and, most recently, immunotherapy. The following are two primary data sources - the National Cancer Institute (NCI) and the National Comprehensive Cancer Network (NCCN) - are used for selecting FDA approved cancer drugs.

NCI (http://www.cancer.gov/) is the United States federal government's principal agency for cancer research. NCI played a prominent role in the discovery of cancer mechanisms and anti-cancer drugs. According to the NCI Dictionary (http://www.cancer.gov/publications/dictionaries/cancer-drug), there are 416 drugs related to 39 cancer subtypes approved by the FDA. Of which, there are 45 drugs combination for different types of cancer treatment.

NCCN is an alliance of the world's leading cancer centers (http://www.nccn.org/). It is recognized by NCI in United States. NCCN creates important clinical guidelines for patients, clinicians, and other health care decision-makers. The NCCN Drugs and Biologics Compendium is a mandated reference for medicare coverage decisions about the appropriate use of drugs and biologics in cancer care.[15] In the NCCN 2015 version, there are 239 unique drugs (normalized by the brand names and DrugBank ID in DrugBank, (http://www.drugbank.ca/) approved by FDA for cancer treatment and complementary therapies, it covers 51 subtypes cancers' treatments. In NCCN drugs, many drugs are non-cancer-related drugs as a cancer adjuvant therapy, such as anti-inflammatory action and side effects alleviation of chemotherapy. A pharmacologic adrenocortical steroid is a typical class for anti-inflammatory action as a cancer adjuvant therapy. Mitotane and prednisone are often used in clinic. Pharmacologic

Figure 1: The flowchart for patient-based drug selection modeling construction. (A) The process of drug and its targets collection; (B) the patient-specific target gene and drug selection process.



substances such as p/neurokinin-1 receptor antagonist are for prevention of chemotherapy-induced nausea and vomiting in the clinic. These drugs will be removed from cancer drug datasets here. A total of 85 drugs are, thus, removed.

The drugs from NCI and from NCCN are merged, and duplicates are removed. In total, 238 drugs make up the cancer drug dataset. The detailed process is shown in Figure 1A.

Cancer Drug Target Collection. Three drug databases are integrated to identify the above-mentioned drugs targets. They are DrugBank (http://www.drugbank.ca/), Therapeutic Target Database (http://bidd.nus.edu.sg/group/cjttd/), and DailyMed (http://dailymed.nlm.nih.gov/dailymed/). These data constitute the critical knowledge base that connects drugs and their targets. Figure 1A shows the detailed process of drug and its target collection. Two hundred and thirty-eight drugs are mapped to dataset DrugBank, Therapeutic Target Database, and DailyMed for their targets searching. A total 598 targets are found. Drugs and their targets (genes) sets are constructed by 238 drugs and 598 targets (genes). However, some genes involved in the drug metabolism and transportation, such as CYP2D6 and OATP1B1, are excluded. These genes are related to pharmacokinetics, but not drug targets. In addition, if the drug target is to a whole DNA, and not a gene, its record is also removed from the drugs-targets dataset. After selection for drugs-targets, the total 148 drugs-272 targets are used in data analysis. Supplementary 1 Table 1 presents all these drugs and their targets.

Triple Negative Breast Cancer Samples. Three types of molecular profiles: mRNA gene expression (GE), DNA CN variation (CNV), and DNA exome sequencing (mutation) were retrieved from CCLE[16] and TCGA[17] databases. This breast cancer study cohort consisted of 85 TNBC tumors and 18 TNBC cell lines. TCGA data and clinic patient annotation were downloaded from website https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm?mode=ApplyFilter with tumor matched selection and level 3 data. In TCGA, DNA exome sequencing data is conducted by the second-generation sequencing Illumina GAIIx platform. Mutation types for missense, nonsense, silent, frame-shift and splice site for each gene are denoted in level 3 data. Here, the gene mutation is signed as 0 or 1 to show if the gene has function loss of mutation or not. The data will be used in drug selection analysis. For mRNA expression, the TCGA for TNBC test uses the AgilentG4502A_07_3 platform. The data is based on its level 2 signals data, where per probe or probe set for each participant's tumor sample has already been normalized according to the description of TCGA data types at the data levels web site (https://tcga-data.nci.nih.gov/tcga/tcgaDataType.jsp). TCGA CN alteration is detected using the Affymetrix 6.0 single nucleotide polymorphism array (SNP-array). CN is measured by a probe corresponding to a segment based upon specific linear calibration curves. The circular binary segmentation algorithm is used to normalize the segmentations (generally, $\log_2^{(CN/2)}$) in the range of [–7,5]. The segmentation mean values of each gene will be used as a background reference to identify the gene focal amplification/deletions and arm-level gains in individuals. From the CCLE database, all data and its annotations can be downloaded from a data portal (http://www.broadinstitute.org/ccle/data/browseData?conversationPropagation=begin). In CCLE, only the GE platform is different from TCGA,[16] which uses Affymatrix HU133 Plus 2.0. For cell line sensitivities with small molecules, it refers to Cancer Therapeutics Response Portal (https://www.broadinstitute.org/ctrp/).[18]

TNBC is definite as the marker deficiency in the estrogen receptor, the progesterone receptor (PR) and the human epidermal growth factor receptor 2 (HER2) (estrogen receptor-/PR-/HER2-). All of these tumors were identified according to the TCGA annotation file. On the other hand, cell line subtypes were defined according to the literature [16]. The summarized 85 patients' characteristics and 18 cell line of TNBC are listed in Supplementary 2 Tables 1 and 2.

In order to detect the differential GE, 9 adjacent non-tumor breast tissue samples in TCGA are defined as normal and are regarded as the control group. These GE profiles from adjacent normal tissue will be used as the mRNA background comparison with new patients' GE. All of the samples are shown in Supplementary 2 Table 3.

Patient-Specific Target and Drug Selection. Personalized medicine can use patients' unique genetic profiles to guide the treatment of disease. Our approach is closely based on this idea. The patient molecular feature includes a GE profile, CNV, and gene mutation that are integrated to identify optimal drug targets from the FDA approved cancer drugs. Figure 1B shows our scheme of patient-specific target and drug selection. Firstly, a drug and its targets are constructed following the flowchart in Figure 1A. It provides important connections between FDA-approved drugs and their targets. Secondly, using the GE profiles of the adjacent normal tissue as the control, a new patient's differential GE will be calculated, including the gene variation folder change and two groups' variation of significant *P*-value. Together with CNV and gene mutation data, they are the input data of our proposed algorithm for target identification, and finally the target-drug matching output. Table 1 shows the algorithmic process. In this algorithm, the chosen target genes shall have a high GE, CN amplification, and no mutation compared to their controls. A stringent selection threshold can warrant that the targeted gene has a significant meaning for the individual gene-based therapy. Here, the threshold of differential expression gene (DEG) selects the baseline corresponding to the top 5% regulation DEG both in the folder-change and the *P*-value; the threshold of CN amplification is 2.62. There is no mutation threshold. Once GE and CNV of the targeted gene are higher than their threshold simultaneously, while the gene is not a mutation, the gene will be chosen as a target for this new patient therapy, at the same time as their matched drugs are searched and output.

As for DEG analysis, the following shows the process. For a patient, the GE*p* is the *P*-value of a 2sample-*t*-test. This test compared the patient GE with the normal breast cancer GE as the following:

Table 1: Target Identification and Drug Matching Algorithm

Input	Patient (GE*p*, CNV*p*, MU*p*)
Knowledge Base Comparison	**Visiting all of the targeted genes, for each of genes**
	GE*p* is compared with these normal tissue GE
	by *t*-test, where *P*-value is less than a threshold and
	folder change value is higher than another threshold.
	CNV*p* means that CNV is higher than a CNV threshold
	MU*p* means no mutation
	If all these conditions are met, the gene is chosen as the drug target.
	Search drugs for the target in drug-target knowledge base
Output	Patient-specific recommendation drugs list

Given an adjacent normal sample $y_j \sim N(\beta, \delta^2)$, $j = 1, 2, n$, its mean $\overline{y}$ distribution is $\overline{y} \sim N(\beta, \delta^2/n)$, n is the number of normal tissue samples. For a new sample x_i from patient tumor, its distribution is assumed to be independently distributed with a homogeneity variance, $x_i \sim N(u_i, \delta^2)$. A statistical hypothesis is to test the similarity between x_i and $\overline{y}$ by a *t*-test method. Here the null hypothesis is H0: $\beta = u_i$; the alternative hypothesis is a one-sided H1: $\beta < u_i$;

The difference between $x_i$and $\overline{y}$ can be tested in $T = \boldsymbol{x_{i-}}\overline{y}/\sqrt{\boldsymbol{\delta}^2 + \boldsymbol{\delta}^2/\boldsymbol{n}} \sim N(0, \delta^2(1 + 1/n))$, and its *t*-statistics can be calculated as

$$t = \frac{\boldsymbol{x_{i-}}\overline{y}}{S\sqrt{1 + 1/\boldsymbol{n}}}, \quad \text{where} \quad s = \sum_{j=1}^{n} \frac{(y_j - \overline{y})^2}{n - 1}$$

This is a one-sided *t*-test with a degree of freedom (*df*) of n and its significance level of 1%. In our TNBC data analysis, the *df* is 9.

Cell Line-Based Target Selection Algorithm using Genomics and Transcriptomics Data. Unlike primary tumor samples that have normal controls, cell lines do not have normal controls. Therefore, highly expressed genes are judged if their expressions are higher than the medians. These highly expressed genes are selected as drug targets. Like the primary tumor samples, genes with CN amplifications (amplification threshold larger than 2.62) and without mutations are selected as the drug targets.

Drug-Target and Drug Potency Correlation Analysis in Cancer Cell Lines. An area under the concentration curve (AUC) is used as the drug potency measurement for its effect on a cancer cell line. In a cell line, the presence of a drug targeted gene is defined by a CN amplification >2.62 and no mutations, or the drug targeted GE is higher than its median expression level; otherwise the drug targeted gene is absent. We hypothesized that cell lines with the presence of a drug targeted gene will have higher drug potency than the cell lines that have the absence of the targeted genes. The potency comparisons are compared using a two-sample *t*-test. This cell line based drug-target potency analysis will supply much needed additional evidence for the TNBC off-target drug use.

RESULTS

Drug Target Distribution. Targeted genes usually are the easiest way to understand cancer mechanisms. Our first set of analysis focuses on the molecular functions and signaling pathways of these targeted genes. Using the analysis software, David (https://david.ncifcrf.gov/), we select the KEGG pathway mapping result (see Figure 2). The target gene distribution in the molecular function is displayed in Figure 2A. The most popular gene function is the phosphoprotein. Among our 272 target genes, 148 are phosphoproteins. Other top functional classes include 96 membrane proteins, 84 nucleotide binding proteins, etc. Looking into the enriched signaling pathways (Figure 2B), top pathways include MAPK, cytokine-cytokine receptor interactions, ErbB, B- and T-cell receptor, Jak-STAT, VEGF, GnRH, mTOR, DNA replication, and P53.

Drugs and Their Pharmacologic Class. Chemotherapeutic agents primarily inhibit mitosis (cell reproduction), induce apoptosis (cell death), and sometimes inhibit metabolic functions of the cancer cell. For example, vinca alkaloids block microtubule assembly, taxanes block microtubule disassembly, alkylating with or without platinum prevents DNA crosslinking, DNA replication inhibitor, etc.; these drugs are called "cytotoxic." The main targets of these chemotherapeutic agents are DNA or cell cycle proteins. Figure 3 clearly displays the distributions of these chemo agents based on their pharmacology mechanisms. The other drug subgroup is to target one or several proteins, in order to disrupt the growth of tumor cells. Monoclonal antibodies, cytokines, antisense, peptide molecules, and kinase inhibitors are the major sub-categories (Figure 3). According to NCCN data collection from different cancer tissues, there are 3604 drug treatment records involving 238 drugs across 124 cancer subtypes (51 different tissues). The drug distribution is calculated and analyzed. Figure 3 shows the distribution of these 238 drugs, of which 63% drugs are chemo agents, while the other non-chemotherapy are more target-based.

Drug targets and Drug Selections for TNBC Patients. In selecting drug targets for TNBC patients, the threshold of GE*p* *P*-value is 0.01, folder change is higher than five comparing TNBC sample to normal sample, and CNV*p* threshold is 0.4 in log-scale (2.62 times CN

Figure 2: Two hundred and seventy-two drug targets distribution and their involvement in the genes function pathway in drugs-targets knowledge. (A) Genes function distribution to these drug targets; (B) The targets' genes involved in the function pathway by KEGG analysis.

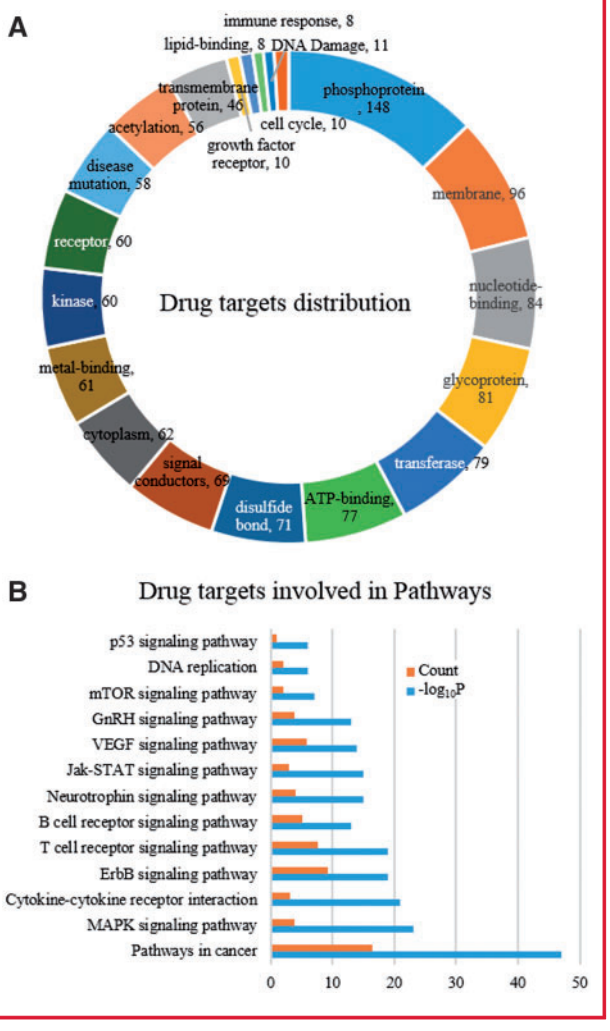


Figure 3: All 238 drugs in distribution in different cancer subtypes by NCCN statistics.

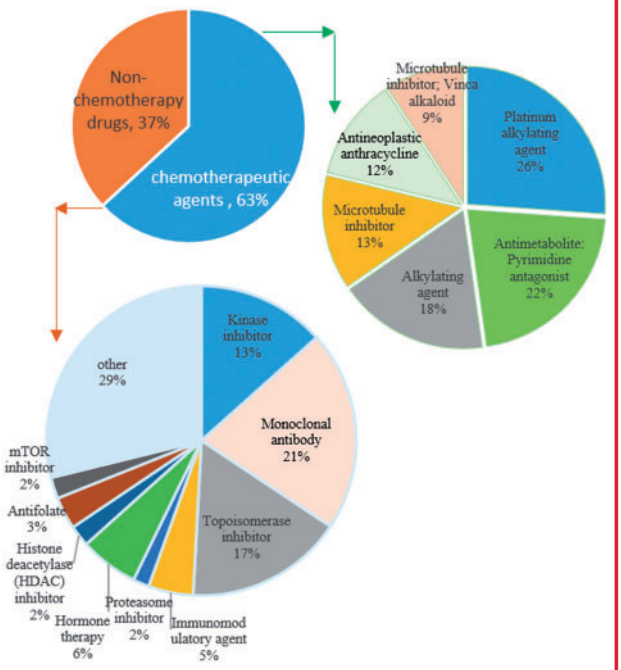


Table 2: The Number of Drug Targets per TNBC Patient

Number of drug targets	TNBC patient frequency
0	24
1	18
2	17
3	11
4	7
5	8
6	4
7	4

amplification in raw scale). Using genomic profiling data in CNV, mutation, and GE, 61 out of 85 TNBC patients (71.8%) have at least one identified drug target, 44 out of 85 patients (51.8%) have at least two drug targets, while 24 patients (28.2%) do not have drug targets (Table 2). The average number of drug targets per patient is 2.05 with standard deviation of 1.94. Some drug-targeted genes, whose molecular genomic data do not show significant up-regulation or CN amplification, were removed as drug-targets.

We then focus on those 61 patients who have identified drug targets. Table 3 presents these selected drug targets and their selection frequency. PARP1 comes out as the highest frequently selected targets, 20 times among 61 TNBC patients who have identified drug targets. The second highest selected is GGH at the frequency of 14, while PTK6 and TOP1MT tied at 7 next. There are totally 62 selected drug targets.

Using our constructed drug-target relationship table (Supplementary 1 Table 1), the targeted drugs were then selected for each patient. Table 4 shows the drug selection frequency for the 61 TNBC patients. Methotrexate has the highest selection frequency, 22; Olaparib ranks second, 20; and the next two are Vandetanib and Regorafenib, which were selected 12 and 11 times, respectively. In total, there are 88 selected drugs based on their targets. The top ranked drugs and the top targets match very well. Comparing Tables 3 and 4, Olaprib is a well-known PARP1 inhibitor; methotrexate is a well-established inhibitor for GGH and TYMS; and Vandetanib inhibits genes PTK6 and epidermal growth factor receptor (EGFR).

There is one patient who has seven identified drug targets, which are (BRAF, CYP51A1, LIMK1, PLA2G4A, POR, PRKCQ, SMO). The potential targeted drugs include: Dabrafenib, Regorafennib, Vemurafenib, Sorafenib, Ketoconazole, Epirubicin, Aldesleukin, Daunorubicin, Doxorubicin, Nilutamide, Mitomycin, tamoxifen, and Vismodegib.

TNBC Cell Drug Screening Data Support Off-Label Precision Medicine Drug Usage. Using the TNBC cell line drug response and genomic profiles in the CCLE database, we investigated drug sensitivity based on our proposed target-gene and drug selection algorithm. Figure 4A illustrates the selection of ABL1-Bosutinib, and its TNBC cell line-based drug sensitivity data. AUC represents the total percent of cell toxicity due to the drug treatment. The larger the AUC is, the stronger the drug sensitivity. Figure 4A suggests that TNBC cells that have high ABL1 GE, more amplification, and less mutation, are more sensitive to Bosutinib than the others ($P = .0013$). Similarly, TNBC cells that have high FLT1 GE, more amplification, and less mutation, are more sensitive to Lenvatinib than the others ($P = .0013$). While neither Bosutinib nor Lenvatinib is indicted for TNBC therapy, these data would support the efficacy of the off-label usage of these two drugs.

DISCUSSION

Comparing our reported drugs and targets (Tables 3 and 4) to the existing FDA approved TNBC drugs and other approved drugs, we found many consistent results in the recent research work on the drug selections for the TNBC patients.[19,20] These selected inhibitors for TNBC by our method were: PARP1 inhibitors (eg, Olaparib), CDK6 inhibitors (eg, Palbociclib), EGFR inhibitors (eg, Gefitinib, Erlotinib, Lapatinib), are very well documented. On the other hand, we have identified a number of drug targets that have not yet been studied in the TNBC patient population, such as GGH and TYMS and their inhibitor Methotrexate, PTK6 and its inhibitor Vandetanib, TOP1MT and its inhibitor Irinotecan, SMO and its inhibitor Vismodegib, etc. However, our method also missed a few drug targets, such as for phosphatidylinositide 3-kinases CD (PI-3Ks family), mTOR, and AKT. PIK3CA is the oncogene, and belongs to one of important PI-3Ks member involved in cellular functions such as cell growth, proliferation, and differentiation. PIK3CA shows the highest frequency of gain-of-function mutations in breast cancer, nearly 33% mutation frequency in breast cancer patients and 37% in TNBC.[21,22] These gain function mutations indicate the benefit of PI3K/mTOR pathway inhibitors. The PIK3CD inhibitor idelalisib (marketed as Zydelig, CAL-101) was an unique drug approved by US FDA as a treatment for patients with chronic lymphocytic leukemia in 2014,[23] and it exists in our knowledge base. However, our current algorithm shields all mutations and target selection in the downstream signaling. Hence, PIK3CA is missed in our target selection. This issue shall be addressed in the future. With regard to mTOR, its GE is absent in the GE data in TCGA, although our knowledge base has it and its targets. Hence, our drug selection missed it. AKT, which

Table 3: Selected Drug Targets and Their Frequencies for TNBC Patients

Involved pathway	Target	Frequency	Involved Pathway	Target	Frequency
	PARP1	20		RALBP1	2
	GGH	14	ErbB, VEGF, GnRH signaling pathway	SRC	2
	PTK6	7		TPMT	2
	TOP1MT	7		YES1	2
	SMO	6		ABCA3	1
	TUBB	6		ATIC	1
	TYMS	6		DGUOK	1
Jak-STA, p53 signaling pathway	CCND1	5	ErbB signaling pathway	ERBB2	1
p53 signaling pathway	RRM2	5		FCGR2A	1
MAPK, ErbB, Neurotrophin and mTOR signaling pathway	BRAF	4		FMO3	1
MAPK signaling pathway	FGFR2	4		FOLR1	1
	GPRC5A	4		GART	1
	LIMK1	4		HDAC1	1
	PSMD2	4	Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway	IFNAR2	1
	TUBB3	4		LIG3	1
	DNMT1	3	GnRH, MAPK signaling pathway	MAP3K2	1
MAPK, ErbB signaling pathway	EGFR	3		NDUFS2	1
	IMPDH1	3	VEGF signaling pathway	NOS3	1
	MAP2	3		ORM1	1
DNA replication	POLE2	3		ORM2	1
	PPAT	3	Cytokine-cytokine receptor interaction, MAPK signaling pathway	PDGFRA	1
T cell receptor signaling pathway	PRKCQ	3		POLB	1
p53 signaling pathway	CDK6	2	DNA replication	POLE	1
	CYP51A1	2		POR	1
	EBP	2		PSMB10	1
	FRK	2	p53 signaling pathway	RRM2B	1
	HDAC2	2		SULT1A1	1
Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway	IL2RA	2		TOP2A	1
	LYN	2		TUBG1	1
MAPK, VEGF, GnRH signaling pathway	PLA2G4A	2		TXNRD1	1
VEGF signaling pathway	PTGS2	2		UPP1	1

include AKT1, AKT2, and AKT3, is now known to play a central role in regulating cellular processes. If deregulated, it contributes to the development or progression of cancer.[24] In our differential gene between TNBC and its adjacent normal, AKT1 has only 1.68 fold higher expression in the TNBC than the normal. It has CN deletion. Therefore, these conditions led to the missing of AKT and its drug selection.

Our current methodology also has some limitations. One limitation is that we exclude all the genes with loss function mutations. Although these function loss mutated genes themselves cannot serve as drug targets, full functional genes in their downstream signaling can still be validated as drug targets. For example, PTEN mutations usually indicate its downstream activation of PI3CA and/or mTOR. Mutated PTEN itself is not a valid drug target, but PI3CA and/or mTOR can very well be. Another example is the BRCA1 loss function mutation that often suggests an active PARP1[22] in its downstream. Our current bioinformatics algorithm cannot yet integrate and query the pathways of these mutated genes. Hence, it remains a potential research topic. The second limitation is the drug selection for TNBC patients who have multiple targets and drugs. Our data report that 51.7% of TNBC patients

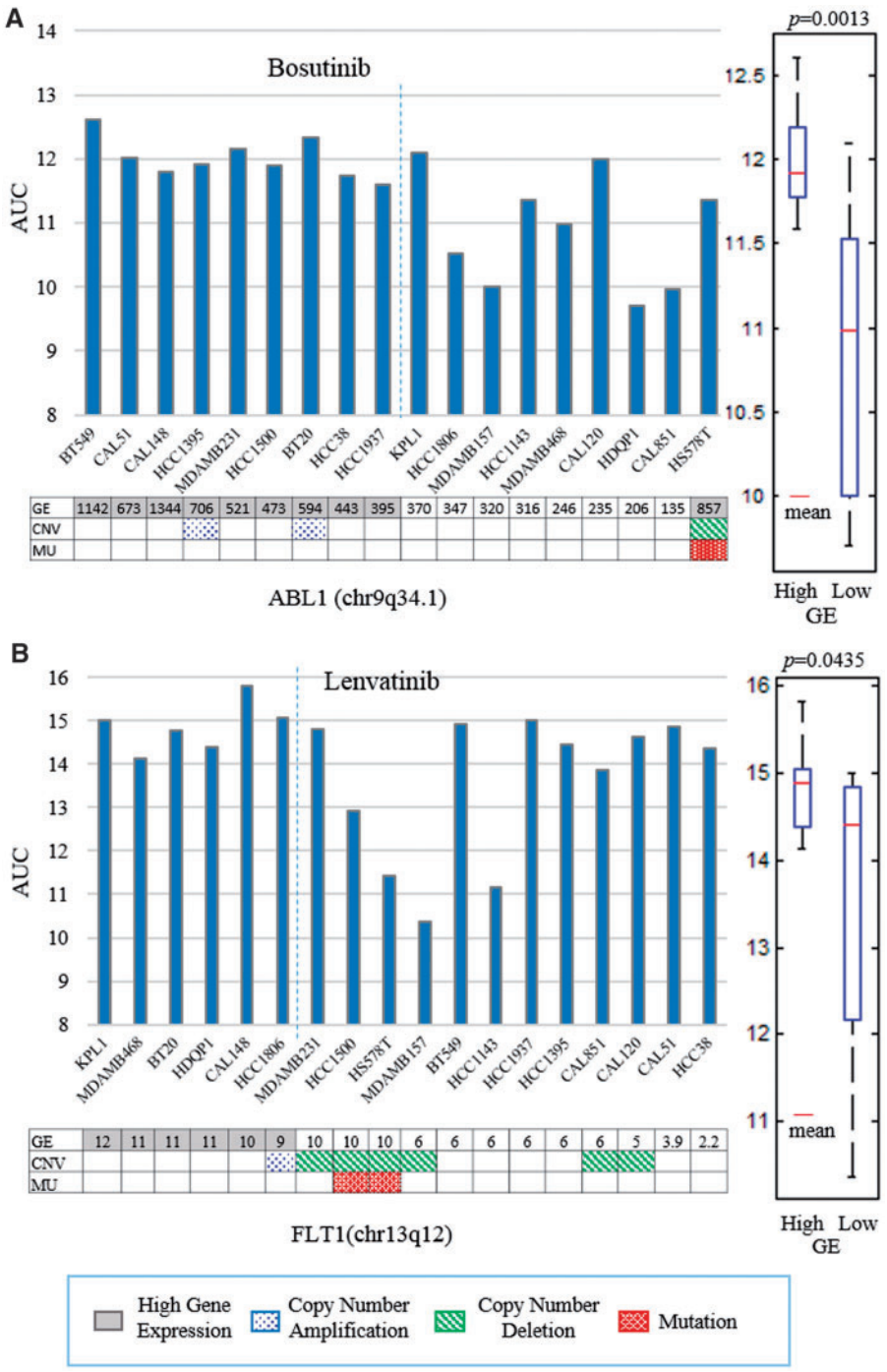
Table 4: Selected Drugs and Their Frequencies for TNBC Patients

Target	Frequency	Target	Frequency	Target	Frequency
Methotrexate	22	Afatinib	4	Alemtuzumab	1
Olaparib	20	Bortezomib	4	Bevacizumab	1
Vandetanib	12	Cetuximab	4	Bleomycin	1
Regorafenib	11	Erlotinib	4	Carfilzomib	1
Cladribine	10	Gefitinib	4	Cytarabine	1
Fluorouracil	10	Ixabepilone	4	Dactinomycin	1
Ponatinib	9	Lapatinib	4	Dexrazoxane	1
Thalidomide	9	Lenvatinib	4	Gemtuzumab	1
Dabrafenib	8	Palifermin	4	Ibritumomab	1
Mercaptopurine	8	Tretinoin	4	Idarubicin	1
Pemetrexed	8	Azacitidine	3	Mitomycin	1
Vincristine	8	Decitabine	3	Mitoxantrone	1
Irinotecan	7	Docetaxel	3	Nelarabine	1
Tamoxifen	7	Epirubicin	3	Nilutamide	1
Topotecan	7	Etoposide	3	Panobinostat	1
Vinblastine	7	Imatinib	3	Pazopanib	1
Vismodegib	7	Paclitaxel	3	Pertuzumab	1
Aldesleukin	6	Panitumumab	3	Rituximab	1
Arsenic	6	Vorinostat	3	Romidepsin	1
Capecitabine	6	Cisplatin	2	Sunitinib	1
Doxorubicin	6	Daunorubicin	2	Tositumomab	1
Floxuridine	6	Denileukin	2	Valrubicin	1
Gemcitabine	6	Ketoconazole	2	ado-trastuzumab	1
Pralatrexate	6	Lenalidomide	2	Alfa-2a	1
Sorafenib	6	Palbociclib	2	Alfa-2b	1
Vinorelbine	6	Peginterferon	2	Emtansine	1
trioxide	6	Pomalidomide	2	Ozogamicin	1
Bosutinib	5	diftitox	2		
Dasatinib	5				
Trastuzumab	5				
Vemurafenib	5				

have at least two drug targets. This will require more advanced ranking schemes for the drug targets and drugs. The third limitation is the drug-targets knowledge of cancer should be filtered. A number of drug off-targets were in our current database. In TNBC patient drug selections, some drugs were not selected based on their main targeted genes. Using trastuzumab for an example, it is a monoclonal antibody directed against the extracellular domain of the tyrosine kinase receptor HER2. Trastuzumab is used for HER2-overexpressing breast cancers in the clinic.[25] However, in our knowledge base, trastuzumab can target 15 genes except EGFR and HER2. Most of these are off-targets of trastuzumab. In our current drug-target selection algorithms for TNBC, trastuzumab is selected due to its off-targets.

Among the current characterization of the molecular mechanisms of the TNBC patients and their target drugs,[26] the primary approach was to use GE signatures to cluster the TNBC patients into subtypes, such as Basal-like 1, Basal-like 2, immunomodulatory, mesenchymal, and mesenchymal stem-like. Then the corresponding drugs were selected based on the pathway enrichment among these selected subtypes, rather than the direct drug target selection. Our proposed method, on the other hand, focuses more on the drug target selections, and then their corresponding drugs. This strategy fits well with the available genomic data on CNV, mutation, and GE, and is much easier for off-label drug usage in clinical practice. However, in order to reduce the practical challenge of our multi-target and multi-drug problem, we shall also integrate the strength of pathway information. On the other hand, we also believe the multi-target and multi-drug problem will not go away even if we have the optimal ranking system. Multiple drug selection, ie, drug combinatory prediction, is

Figure 4: Target-based drugs screening in TNBC cell lines.



an unavoidable reality. It motivates additional cancer biology and computational biology research alone this line.

In this paper, for the first time, we have shown a bioinformatics approach that integrates various drugs, targets, and genomic data sources, and automatically predicts and selects drugs for the identified drug targets among TNBC individuals. We then further demonstrate the value of the cell line drug screening data for the off-label drug use in the TNBC precision medicine scenario by the CCLE and Cancer Therapeutics Response Portal. Using the cell line-based drug screening data, we have found significant evidence that drug targets having high GE, more amplification, and less mutation will be more sensitive to their corresponding drugs.

CONTRIBUTORS

Entire paper design and writing by L.L. and L.J.C.; algorithm design and development for precision medicine tools L.J.C.; and knowledge base data collection by L.J.C. and B.P.S.

FUNDING

This work was funded by the National Institutes of Health Research Foundation grant numbers DK102694, GM10448301, and LM011945.

ACKNOWLEDGEMENTS

We acknowledge Sai Mounika Inavolu, Varshini Vasudevaraja, and Yunlong Liu for their assistance in the research and with our study participants.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Vogelstein B, Papadopoulos N, Velculescu VE, *et al.* Cancer genome landscapes. *Science.* 2013;339(6127):1546–1558.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372(9):793–795.
3. Barker AD, Sigman CC, Kelloff GJ, *et al.* I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther.* 2009;86(1):97–100.
4. Kim ES, Herbst RS, Wistuba II, *et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discovery.* 2011;1(1):44–53.
5. Do K, O'Sullivan Coyne G, Chen AP, An overview of the NCI precision medicine trials-NCI MATCH and MPACT. *Chin Clin Oncol.* 2015; 4(3):31.
6. Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. *J Clin Oncol.* 2015; 33(23):2563–77.
7. Kandoth C, McLellan MD, Vandin F, *et al.* Mutational landscape and significance across 12 major cancer types. *Nature.* 2013;502(7471):333–339.
8. Zack TI, Schumacher SE, Carter SL, *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet.* 2013;45(10):1134–1140.
9. Conti RM, Bernstein AC, Villaflor VM, *et al.* Prevalence of off-label use and spending in 2010 among patent-protected chemotherapies in a population-based cohort of medical oncologists. *J Clin Oncol.* 2013;31: 1134–1139.
10. Fairman KA, Curtiss FR. Regulatory actions on the off-label use of prescription drugs: Ongoing controversy and contradiction in 2009 and 2010. *J Manag Care Pharm.* 2010;16:629–639.
11. American Society of Clinical Oncology. Reimbursement for cancer treatment: coverage of off-label drug indications. *J Clin Oncol.* 2006;24: 3206–3208.
12. Poole SG, Dooley MJ. Off-label prescribing in oncology. *Support Care Cancer.* 2004;12:302–305.
13. http://www.broadinstitute.org/ccle/home
14. Daemen A, Griffith OL, Heiser LM, *et al.* Modeling precision treatment of breast cancer. *Genome Biol.* 2013;14(10):R110.
15. Ettinger DS, Kuettel M, Malin J, *et al.* NCCN roundtable: what are the characteristics of an optimal clinical practice guideline? *J Natl Compr Canc Netw.* 2015;13:640–642.
16. Barretina J, Caponigro G, Stransky N, *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012;483(7391):603–607.
17. Tomczak K, Czerwinska P, Wiznerowicz M, *et al.* The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncol.* 2015;19(1A):A68–A77.
18. Basu A, Bodycombe NE, Cheah JH, *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell.* 2013;154:1151–1161.
19. O'Toole SA, Beith JM, Millar EK, *et al.* Therapeutic targets in triple negative breast cancer. *J Clin Pathol.* 2013;66(6):530–542.
20. Ellis ML, Perou CM. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discovery.* 2013;3(1):27–34.
21. Forbes SA, Tang G, Bindal N, *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* 2010;38(Database issue):D652–D657.
22. Cancer Genome Atlas N: Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70.
23. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm406387.htm.
24. Altomare DA, Testa JR. Perturbations of the AKT signaling pathway in human cancer. *Oncogene.* 2005;24:7455–7464.
25. Valabrega G, Montemurro F, Aglietta M. Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer, *Ann Oncol.* 2007;18(6):977–984.
26. Lehmann BD, Bauer JA, Chen X, *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750–2767.

AUTHOR AFFILIATIONS

[1]Center for Computational Biology and Bioinformatics, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

[2]Department of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

[3]Division of Hematology and Oncology, Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN 46202, USA

[4]Indiana Institute of Personalized Medicine, School of Medicine, Indiana University, Indianapolis, IN 46202, USA