



Therapeutic Response-Based Reclassification of Multiple Tumor Subtypes Reveals Intrinsic Molecular Concordance of Therapy Across Histologically Disparate Cancers

Yue Xu^{1†}, Jie Zheng^{1†}, Zhaoqing Cai¹, Wang Li², Jens Köhler³, Yao Dai¹, Xiaojie Cheng¹, Tao Wu¹, Fan Zhang^{1,2*} and Haiyun Wang^{1*}

¹School of Life Sciences and Technology, Tongji University, Shanghai, China, ²Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China, ³Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, United States

OPEN ACCESS

Edited by:

Chuan-Xing Li,
Karolinska Institutet (KI), Sweden

Reviewed by:

Yanhong Zhou,
Central South University, China
Bairong Shen,
Sichuan University, China

*Correspondence:

Fan Zhang
fzhang@tongji.edu.cn
Haiyun Wang
wanghaiyun@tongji.edu.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Molecular and Cellular Pathology,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 09 September 2021

Accepted: 22 October 2021

Published: 12 November 2021

Citation:

Xu Y, Zheng J, Cai Z, Li W, Köhler J,
Dai Y, Cheng X, Wu T, Zhang F and
Wang H (2021) Therapeutic
Response-Based Reclassification of
Multiple Tumor Subtypes Reveals
Intrinsic Molecular Concordance of
Therapy Across Histologically
Disparate Cancers.
Front. Cell Dev. Biol. 9:773101.
doi: 10.3389/fcell.2021.773101

Cancers that are histologically defined as the same type of cancer often need a distinct therapy based on underlying heterogeneity; likewise, histologically disparate cancers can require similar treatment approaches due to intrinsic similarities. A comprehensive analysis integrated with drug response data and molecular alterations, particularly to reveal therapeutic concordance mechanisms across histologically disparate tumor subtypes, has not yet been fully exploited. In this study, we integrated pharmacological, genomic, and transcriptomic profiling data provided from the Cancer Genome Project (CGP) in a systematic *in silico* investigation of the pharmacological subtypes of cancers and the intrinsic concordance of molecular mechanisms leading to similar therapeutic responses across histologically disparate tumor subtypes. We further developed a novel approach to redefine cell-to-cell similarity and drug-to-drug similarity from the therapeutic concordance, providing a new point of view to study cancer heterogeneity. This study demonstrates how pharmacological and omics data can be used to systematically classify cancers in terms of response to various compounds and provides us with a purely therapy-oriented perspective to view tumor classifications independent of histology subtypes. The knowledge of pharmacological subtypes of 367 drugs are available via our website (<http://www.hywanglab.cn/dtdb/>), providing the resources for precision medicine in the perspective of therapeutic response-based re-classification of tumor.

Keywords: tumor classification, OMICS data, pharmacological subtypes, precision medicine, CGP dataset

INTRODUCTION

Traditional tumor classification based on histopathologic diagnosis and the TNM staging system offers highly practical guidance for surgical resection, regional radiotherapy or chemotherapies. However, due to complex intertumor and intratumor heterogeneity, cancer patients with such histological diagnoses often suffer from receiving effective drug treatment and therapeutic resistance (Bedard et al., 2013; Arozarena and Wellbrock, 2019; Ding et al., 2019). One of the key mechanisms behind this significant challenge is the fact that cancers from the same tissue of origin often present

quite different mechanisms of oncogenesis at the molecular level (Bedard et al., 2013; Arozarena and Wellbrock, 2019). Decades of studies have focused on finding molecular subtypes within histopathologically defined tumor types by analyzing large-scale genomic, transcriptomic, proteomic, and epigenomic alterations (Ko et al., 2020). As a result, a variety of molecular signatures have been identified to distinguish intrinsic molecular subtypes associated with patient survival, prognosis and response to different therapeutic modalities (Gay et al., 2021). For example, BRAF mutation in melanoma (Menzer et al., 2019); EGFR-mutant lung adenocarcinomas (Marcoux et al., 2019); luminal A, luminal B, HER2-enriched, basal-like and normal-like subtypes in breast cancer (Prat et al., 2012; Rueda et al., 2019; Waks and Winer, 2019); and four prominent genetic subtypes in Diffuse large B-cell lymphoma (DLBCL), termed MCD based on the co-occurrence of MYD88L265P and CD79B mutations, BN2 based on BCL6 fusions and NOTCH2 mutations, N1 based on NOTCH1 mutations, and EZB based on EZH2 mutations and BCL2 translocations are several well-established molecular subtypes with explicit clinical significance (Schmitz et al., 2018).

With these advances in challenging cancer heterogeneity by identifying subtypes within the same tissues, emerging studies are uncovering facts in the other direction that cancers across disparate tissues of origin can explicitly share common molecular mechanism of oncogenesis (Hoadley et al., 2014; Lu et al., 2019; Petralia et al., 2020). For example, one study based on a large-scale genomic analysis revealed that lung squamous, head and neck, and a subset of bladder cancers shared highly concordant signatures typified by TP53 alterations, TP63 amplifications, and high expression of immune and proliferation pathway genes, implicating that those different cancers perhaps require similar treatment approaches (Hoadley et al., 2014). Another study based on large-scale genomic data found that TP53 and KRAS were mutually exclusive in COAD, READ, and LUAD, but significantly coexisted in PAAD. These observations reveal the feasibility of considering the same treatment strategy in different tumor types (Drilon et al., 2018). Similarly, HER2-targeted therapy may be applied to other cancer types analogous to breast cancer because ERBB2/HER2, which can be amplified in breast cancer, is also mutated and/or amplified in subsets of glioblastoma and gastric, serous endometrial, bladder and lung cancers (Cancer Genome Atlas Research et al., 2013). In recent years, the FDA granted approval to larotrectinib, which had marked and durable antitumor activity in a variety of patients with RTK fusion-positive cancer, regardless of age or tumor tissue type (Drilon et al., 2018). Such examples explicitly illuminate a distinct avenue for reclassifying multiple tumor types independent of histopathologic diagnosis, highlighting that treatment approaches specifically discovered in one disease can be applied to another due to their intrinsic concordance of molecular patterns.

Currently, emerging efforts, such as pan-cancer analysis projects, are being conducted to comprehensively define commonalities and differences across cancer types and tissues of origin (Cancer Genome Atlas Research et al., 2013). Nonetheless, integrative analysis, particularly towards revealing

intrinsic therapeutic concordance across histologically disparate tumor subtypes, has not yet been fully exploited. Currently, a large-scale pharmacogenomics study, the Cancer Genome Project (CGP), provides high-throughput genomic, expression information and pharmacological profiling of anticancer drugs across hundreds of cell lines that represent explicit molecular subtypes of histologically defined tumors (Garnett et al., 2012). Hence, we integrated drug response information to pharmacological reclassify tumor subtypes regardless of their tissues of origin. Tumor cells in the same class present similar drug responses, and those in different classes show varied drug responses. Furthermore, by integrating genomic alteration and expression information, we unravelled the intrinsic concordant molecular mechanism associated with the common drug response across histologically disparate cancers. Importantly, this research provides us with a purely therapy-oriented perspective to re-examine tumor classifications independent of histology subtypes.

MATERIALS AND METHODS

Data from a large-scale pharmacogenomics study, the Cancer Genome Project (CGP), was accessible from its website: <http://www.cancerrxgene.org>. Gene expression, mutation and drug sensitivity data were downloaded. The CGP dataset includes 987 cell lines, genome-wide analysis of mutations, copy number variations and expression profiling, as well as the presence of commonly rearranged cancer genes, and 367 pharmacological profiles (dataset version 2020) (Garnett et al., 2012). For pharmacological assays, the dose range of drugs varied greatly for the different drugs. Generally, about 96% drugs used the minimal tested concentration below 0.07820 μM and the maximal tested concentrated below 20 μM . In these data, the natural logarithm of the Half maximal inhibitory concentration (IC50) value represents the drug sensitivity value. IC50 is the half maximal inhibitory concentration of an anticancer drug, and a lower value means higher sensitivity. Cell lines cannot be classified according to histological subtypes provided by TCGA, labelled as "UNCLASSIFIED", were excluded from our study.

Generating the Pharmacological Subtypes Tree

The cell lines, IC50 of which are smaller than maximal tested concentrations, are defined as sensitive, otherwise resistant. Our method iteratively splits the cancer cells into two groups in a way that gains the best separation of drug sensitivity between two groups until reaching two terminal conditions: The p value of the difference of two groups' drug sensitivity values is smaller than 0.05 or all cancer cells in a node are sensitive or resistant.

Suppose the set of all cells is S_{total} , and each cell has its drug sensitivity value. The procedure to grow a pharmacological subtype tree is as follows:

- 1) Define the root node. Set S as the set of cancer cells in this node, and assign S_{total} to S .

- 2) Define the root node as the current node.
- 3) N denotes the number of cells in the current node. If N is smaller than 6 or all cells in such node are sensitive or resistant, finish the procedure of growing the tree. Otherwise, a heuristic algorithm splits S into two child nodes, the left node and the right node, which achieves the greatest drug sensitivity difference.
- 4) If no way can be found to split the current node ($p < 0.05$), finish the procedure of growing the tree. Otherwise, the current node is split into two child nodes.
- 5) Define the left node and right node. Suppose the set of cells in the left node is S_l and the set in the right node is S_r . S_l and S_r are the subsets of S .
- 6) Define the left node as the current node. Set S as the set of cells in this node, and assign S_l to S . Repeat steps 3–5.
- 7) Define the right node as the current node. Set S as the set of cells in this node, and assign S_r to S . Repeat steps 3–5.

Algorithm for Splitting the Node

We used a heuristic algorithm to search for a reasonable number of divisions to split S into S_l and S_r . Suppose the number of the cells set is ns . We first sorted the cells by IC50 ascending and divided the cells with higher drug sensitivity into the S_l group and the cells with lower drug sensitivity into the S_r group. Since each node at least 6 cells, a total of $ns - 5$ divisions were considered.

For each division, the Mann-Whitney U test was applied to calculate the p value of drug sensitivity between S_l and S_r . The division with the smallest p value was selected as the optimized one to split the cancer cells in the current node.

The Cell Similarity and the Drug Similarity

The similarity of cells is calculated based on their response to the different drugs. For each drug k , a pharmacological subtypes tree S^k is generated, consisting of n subtypes:

$$S^k = (S_1^k, S_2^k, \dots, S_n^k) = (S_t^k)_{t=1 \rightarrow n}$$

Based on this tree, a matrix that defines a similarity of cancer cells, regarding whether they are in the same subtype and their drug response to drug k , is calculated. We assume there are m cell lines tested for drug k . The cell lines are labelled as sensitive, resistant, or NA (not available). A sensitive cell line has its IC50 value smaller than maximum tested concentration; a resistant cell line has its IC50 value greater than maximal tested concentration. For two given cell lines i and j , their similarity matrix (D_{ij}^k) is defined as:

$$D_{ij}^k = \begin{cases} 1, & \text{if } (i, j) \in S_t^k \text{ and both are labelled as sensitive ones;} \\ -1, & \text{if } (i, j) \in S_t^k \text{ and both are labelled as resistant ones;} \\ 0, & \text{otherwise.} \end{cases}$$

If two cell lines i and j are in the same subtype and both are sensitive to drug k , their score is 1, and -1 for both cell lines with a resistant response, while if they are not in the same subtype or show opposite response to the drug, their score is 0. A $m \times m$ similarity matrix of cell lines is then generated.

Then, the similarity of cell lines i and j can be further calculated as:

$$Cell_{sim} = \frac{\sum_{k=1}^q |D_{ij}^k|}{q}, \quad i \neq j$$

Wherein i and j denote the cell lines tested for q drugs.

Furthermore, for two given drugs a and b , their similarity is calculated as the summarized similarity across the same cell lines which are tested for both drugs.

$$Drug_{sim} = \frac{\sum_{i=1}^m \sum_{j=1}^m D_{ij}^a \cdot D_{ij}^b}{\sum_{i=1}^m \sum_{j=1}^m D_{ij}^a \cdot D_{ij}^a + \sum_{i=1}^m \sum_{j=1}^m D_{ij}^b \cdot D_{ij}^b - \sum_{i=1}^m \sum_{j=1}^m D_{ij}^a \cdot D_{ij}^b}, \quad i \neq j$$

Wherein i and j denote the cell lines tested for both drugs, and m is the total number of the same cell lines.

Connecting Pharmacological Subtypes With Genomic Alterations

The maximum concentration of the drug was used to distinguish between the resistant pharmacological subtypes and sensitive pharmacological subtypes. The chi-squared test was used to calculate the connection between genomic alterations, including mutation and translocation, and pharmacological subtypes. Genomic alterations were determined with p values corrected with the Benjamini-Hochberg method for controlling the false discovery rate (Benjamini and Hochberg, 1995). Here we used the corrected p value < 0.1 as the threshold. We defined two types of connections: sensitive and resistant. If pharmacological subtypes with higher drug sensitivity had more frequent alteration occurrences, we defined this connection as sensitive. If the opposite was true, we defined the connection as resistant.

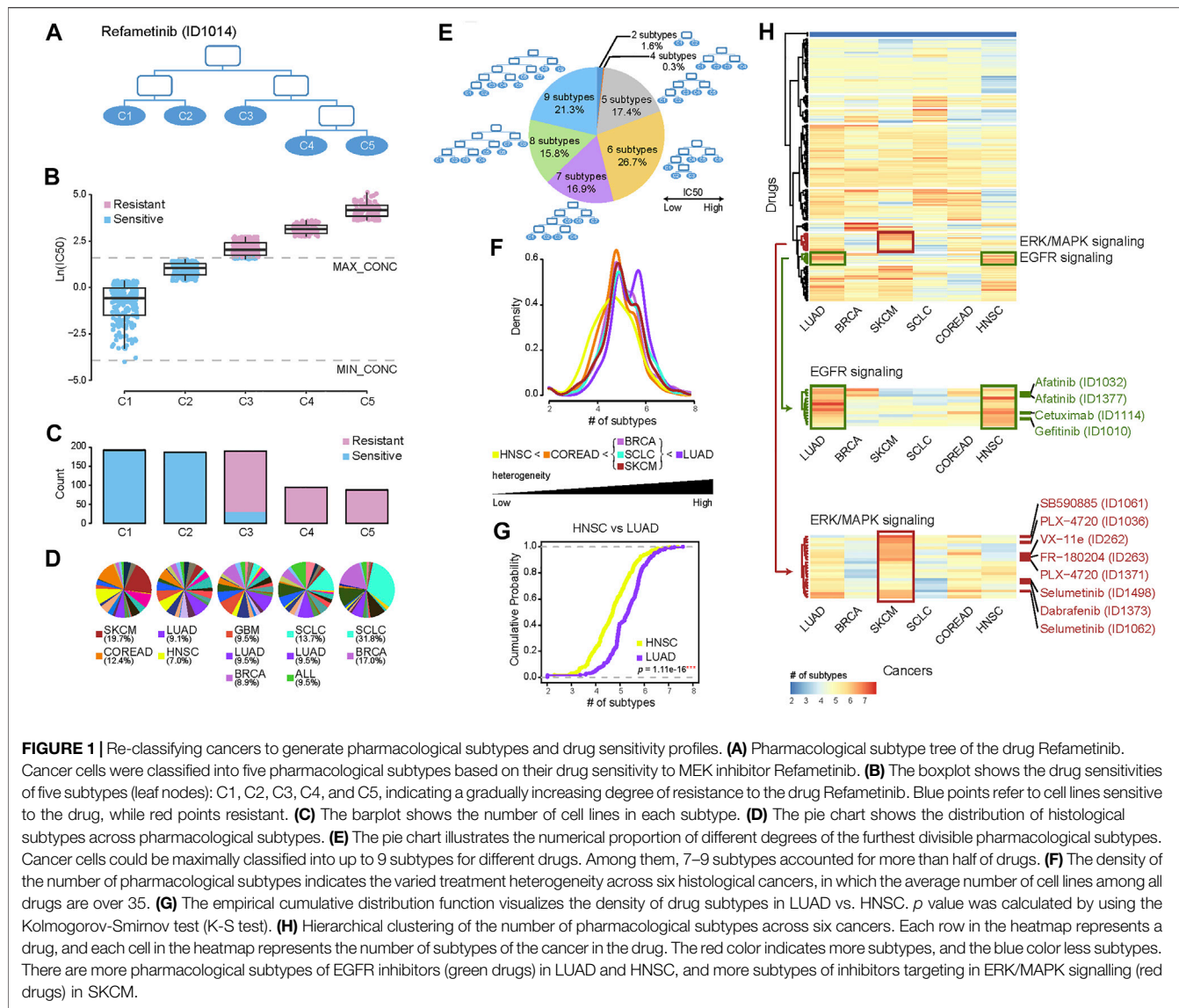
The connection between mRNA expression and pharmacological subtypes was determined with p values from the student t test corrected with the Benjamini-Hochberg method. If pharmacological subtypes with higher drug sensitivity had higher expression, we defined this connection as sensitive. If the opposite was true, we defined the connection as resistant. Here we used the corrected p value < 0.1 as the threshold. To obtain more confident connections, the Pearson correlation coefficient (PCC) was used to measure the correlation between drug sensitivity (IC50) and gene expression in the sensitive subtypes, and only genes with PCC greater than 0.3 were considered. The individual subtypes with sensitive and resistant cells mixed were excluded in the above analyses.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was employed to determine the 333 gene sets from KEGG, enriched by a pre-ranked list of all genes, which were sorted by the statistical significance of differential expression defined by DESeq2 analysis (Love et al., 2014). Gene sets with FDR < 0.05 were statistically significant.

Statistical Analysis

The Fisher's exact test was used to respectively determine whether there is a significant association between pathways and clusters of drugs



determined by the number of drug pharmacological subtypes across 6 cancer types or not, whether any difference of phenotypes (cancers/pathways) in clusters of cell lines/drugs derived from the similarity matrix was significant, and whether cancer censor genes are enriched in a list of genes with the most connections to pharmacological subtypes. A hypothesis test based on hypergeometric distribution was used to determine whether a histological cancer is enriched in the most sensitive pharmacological subtype of a drug. The *p* value <0.05 was regarded as statistically significant. Kolmogorov-Smirnov test (K-S test) was employed to compare cumulative distribution function (CDF) of the number of pharmacological subtypes across the drugs between two histological cancers.

RESULTS

We analysed 367 drugs in the CGP dataset and established their pharmacological subtype trees. The leaf node in each tree

represents a pharmacological subtype of cancers, which is composed of cells from the histological disparate tumor subtypes. We further employed the mRNA expression profile and mutation/fusion profile to connect the molecular alterations with pharmacological subtypes, delineating a novel perspective to re-classify the tumor according to therapeutic response dependent of the intrinsic concordant molecular mechanism, by regardless of tissues of origin. The cell similarity and the drug similarity were also redefined based on the pharmacological subtypes.

Pharmacological Subtypes of Cancers

We built a pharmacological tree for each drug based on the divisibility of the drug sensitivity of cancer cells. Taking the tree of the MEK1/2 inhibitor Refametinib as an example (Figure 1A), all cells in the root node were first divided into left and right child nodes with relatively high and low sensitivity, respectively, and further cells in these two nodes were capable of being divided into

five final subgroups, C1, C2, C3, C4, and C5. Each subgroup had varied drug sensitivity and reached maximum divisibility, thus representing distinct pharmacological subtypes. C1 denotes the most sensitive and C5 the most resistant subtype to the drug Refametinib. Except for C3 where both sensitive and resistant cells were mixed, the other subgroups contained the homologous sensitive or resistant cells (**Figure 1B**). There were much more sensitive cells in C1 and C2 than resistant cells in C4 and C5 (**Figure 1C**). We then investigated how histological subtypes were distributed in the pharmacological subtypes (**Figure 1D**). C1 was composed of 25 histological subtypes, with top 2 cancers, SKCM (19.7%) and COREAD (12.4%). C5 was composed of 22 histological subtypes, with leading cancers including SCLC (31.8%) and BRCA (17.0%).

Our analysis established 367 pharmacological trees. The number of pharmacological subtypes with distinct drug sensitivities varied from 2 to 9 across 367 trees (**Figure 1E**). This suggested that heterogeneity of treatment effects across histological subtypes existed widely, and one pharmacological subtype presenting concordant drug response indeed comprised histological disparate cells to some extent. A high degree of treatment heterogeneity, meaning more than seven pharmacological subtypes, was observed for 54.0% (16.9% of the drugs had 7; 15.8% of the drugs had 8; and 21.3% of the drugs had 9 pharmacological subtypes) of the drugs. These drugs included Dabrafenib (BRAF inhibitor), Selumetinib (MEK1/2 inhibitor), Erlotinib (EGFR inhibitor), and Alectinib (ALK inhibitor). 44.1% of the drugs had 5 or 6 subtypes, showing a moderate degree of treatment heterogeneity, while 1.9% of the drugs had subtypes lower than 4, showing a low degree of heterogeneity related to drug therapy (**Figure 1E**).

We also particularly examined how pharmacological subtypes were constituted within the same histological tumor subtypes, including LUAD, BRCA, SKCM, SCLC, COREAD, and HNSC, in which the average number of cell lines per drug was more than 35 (**Figures 1F–H**). We built approximately 310 pharmacological trees for each type of the above cancers. The density distribution of the number of pharmacological subtypes indicated the varied treatment heterogeneity across 6 histological subtypes (**Figure 1F**). The shape of the distribution curve in LUAD was characterized by double kurtosis and significantly positively biased, and oppositely the curve in HNSC was negatively biased, suggesting the highest treatment heterogeneity in LUAD and lowest treatment heterogeneity in HNSC. Furthermore, K-S test was applied to evaluate if two different histological subtypes had the same level of treatment heterogeneity by comparing their cumulative distributions. The results showed the cumulative distribution curve of LUAD is significantly different from that of HNSC ($p = 1.11e-16$, K-S test) (**Figure 1G**). The pairwise comparisons between the curves of any two histological subtypes showed BRCA, SCLC, and SKCM had the statistically similar distributions ($p > 0.05$, K-S test), and other cancers had not ($p < 0.05$, K-S test) (**Supplementary Figure S1**).

To observe the varied treatment heterogeneity of the different histological subtypes in detail, the number of pharmacological subtypes of drugs across 6 histological cancers were shown in the

heatmap (**Figure 1H**). Six histological tumors were arranged in columns according to their treatment heterogeneity, from high to low. Interestingly we observed that EGFR inhibitors (Afatinib, Cetuximab, and Gefitinib) targeting in EGFR signalling pathway (green drugs in **Figure 1H**) were clustered together ($p < 0.001$, Fisher's exact test). And in LUAD and HNSC there were more pharmacological subtypes of EGFR inhibitors than other drugs, suggesting high treatment heterogeneity for EGFR inhibitors in LUAD and HNSC. In addition, drugs targeting ERK/MAPK signalling (red drugs in **Figure 1H**) were clustered together ($p < 0.001$, Fisher's exact test). There was high treatment heterogeneity for these drugs in SKCM.

Genomic Alterations Associated With Pharmacological Subtypes

To identify the molecular alterations associated with varied drug sensitivity, we examined whether the genomic alterations, including mutations and translocations, significantly changed across the pharmacological subtypes. The statistical significance was determined by p values calculated with the Chi-squared test and corrected with the Benjamini-Hochberg method for controlling the false discovery rate. If the pharmacological subtypes with higher drug sensitivity had more frequent mutation or translocation events, we defined their association as a sensitive connection. If the opposite was true, we defined their association as resistant. Consequently, genes with specific genomic alterations were associated with the respective drugs. Since one gene could have multiple connected drugs and one drug could have multiple connected genes, these gene-drug connections ultimately composed a network that allowed us to investigate the contribution of either a drug or gene to the holistic understanding of how altered genes connect to drug responses (**Figure 2A**). The hub genes with the most connections to the drugs, including known cancer genes KRAS, RB1, BRAF, TP53, NOTCH2, and other genes LTB, TNFRSF9, DERA et al. were highlighted in the network.

We first investigated the distribution of genes on their connected drugs (**Figure 2B**). Our analysis identified 624 genes whose genomic alterations were connected with at least two drugs' pharmacological subtypes. Most of the genes (70.67%) had fewer than 3 connections with drugs, but 40 (6.41%) genes whose genomic alterations were associated with more than 5 drugs. By ranking genes by their number of connections to drugs, we found that the top ranked genes were enriched in cancer genes (**Figure 2C**; **Supplementary Figure S2**), a catalogue of genes with mutations or fusions that are causally implicated in cancer provided by the COSMIC database (Tate et al., 2019). The top 10 genes connected to more than 12 drugs, 30% of which were remarkably annotated as cancer genes (**Figure 2C**). These cancerous percentages for the top 10 ($p = 3.62e-03$, Fisher's exact test), top 20 ($p = 1.27e-07$, Fisher's exact test), and top 30 genes ($p = 1.62e-09$, Fisher's exact test) were significantly higher than the background when using genome-wide 21,972 genes as a reference (**Figure 2C**). Among the top ranked oncogenes, ABL1_BCR ranked first with connections to 28

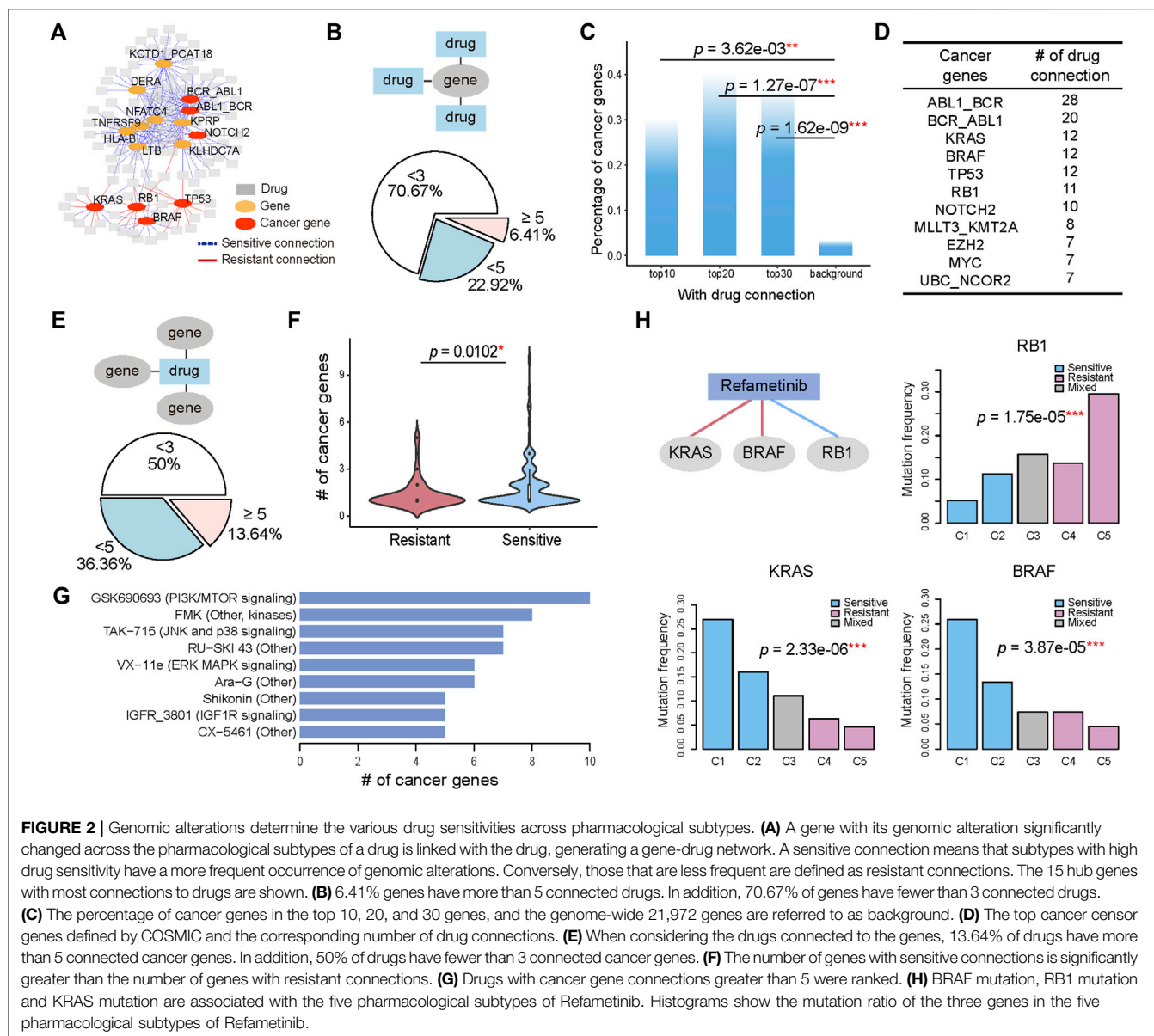


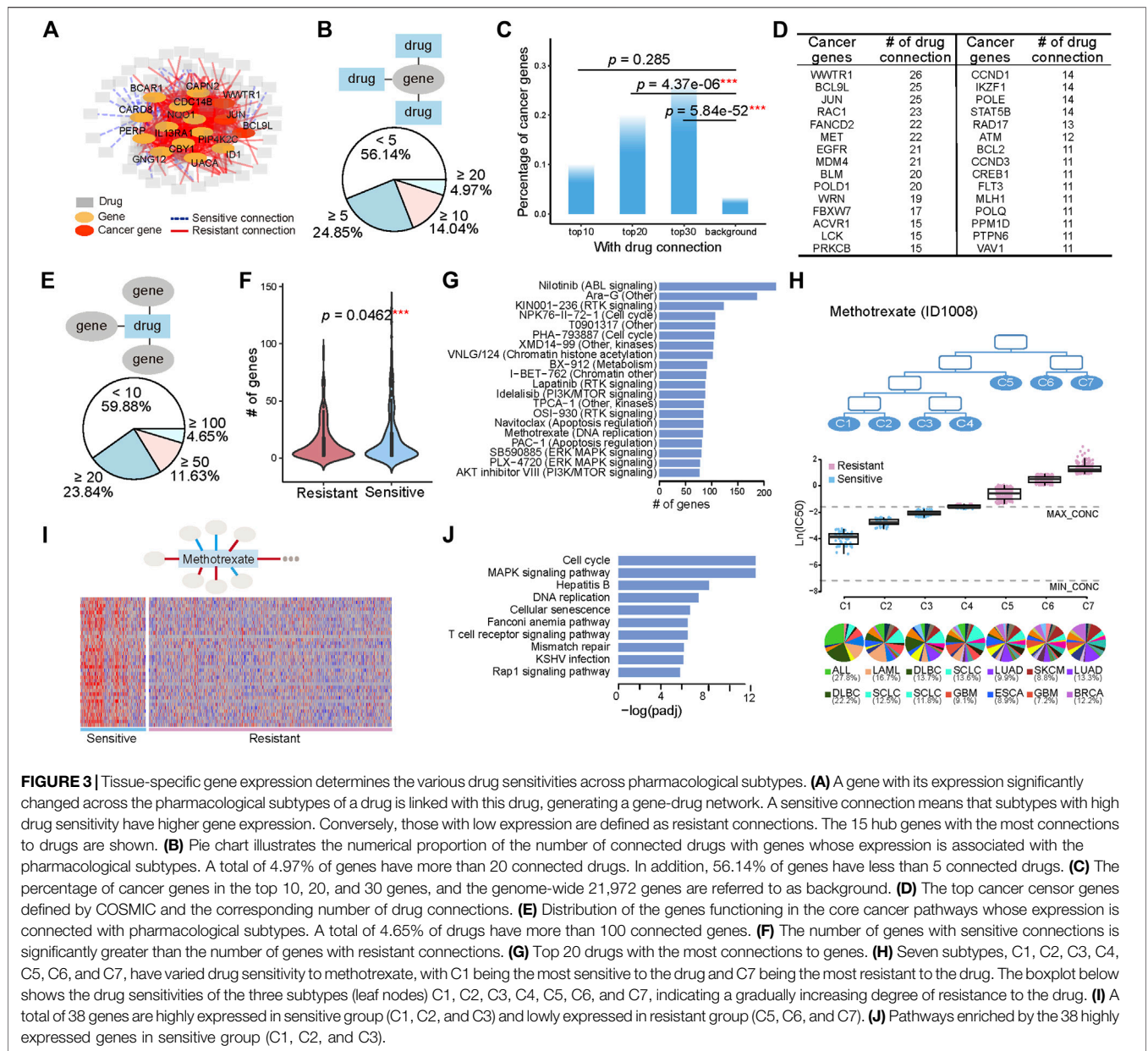
FIGURE 2 | Genomic alterations determine the various drug sensitivities across pharmacological subtypes. **(A)** A gene with its genomic alteration significantly changed across the pharmacological subtypes of a drug is linked with the drug, generating a gene-drug network. A sensitive connection means that subtypes with high drug sensitivity have a more frequent occurrence of genomic alterations. Conversely, those that are less frequent are defined as resistant connections. The 15 hub genes with most connections to drugs are shown. **(B)** 6.41% genes have more than 5 connected drugs. In addition, 70.67% of genes have fewer than 3 connected drugs. **(C)** The percentage of cancer genes in the top 10, 20, and 30 genes, and the genome-wide 21,972 genes are referred to as background. **(D)** The top cancer genes defined by COSMIC and the corresponding number of drug connections. **(E)** When considering the drugs connected to the genes, 13.64% of drugs have more than 5 connected cancer genes. In addition, 50% of drugs have fewer than 3 connected cancer genes. **(F)** The number of genes with sensitive connections is significantly greater than the number of genes with resistant connections. **(G)** Drugs with cancer gene connections greater than 5 were ranked. **(H)** BRAF mutation, RB1 mutation and KRAS mutation are associated with the five pharmacological subtypes of Refametinib. Histograms show the mutation ratio of the three genes in the five pharmacological subtypes of Refametinib.

drugs, followed by BCR_ABL1 and KRAS with 20 and 12 drugs (Figure 2D), implying that the highly drug-connected genes that are not currently identified as cancer genes could actually be causally implicated in cancer therapy, such as KCTD1_PCAT18 fusion, TNFRSF9, LTB, and NFATC4, which showed connections to over 18 drugs and ranked as the top five genes (Supplementary Figure S2).

We then investigated the distribution of drugs on their connected genes. In this analysis, we focused only on cancer genes. For the majority of drugs (86.36%), the number of connected genes ranged from 2 to 5 (Figure 2E). Among the pharmacological subtypes, genes whose alteration was sensitively connected with drug resistance were significantly more abundant than those that were connected in a resistant manner ($p = 0.0102$, student t test). (Figure 2F). We ranked the

drugs with more than 5 genes' connection with their pharmacological subtypes (Figure 2G). Drugs with gene connections greater than 5 included GSK690693, affecting the PI3K/MTOR signaling pathway, TAK-715, affecting the JNK and p38 signaling pathway, VX-11e, affecting the ERK MAPK signaling pathway, and IGFR_3801, affecting the IGF1R signaling pathway.

Next, we observed how the mutation rate of genes associated with the variability of drug sensitivity changed among drug-sensitive and drug-resistant subtypes. Taking the MEK1/2 inhibitor Refametinib as an example, three genomic alterations were found to connect to the drug and be associated with pharmacological subtypes. As shown in Figure 2H, BRAF mutations and KRAS mutations occurred significantly more frequently (over 25%) in the sensitive subtype C1 and then



decreased gradually in the subtypes C2, C3, C4, and C5. Here one subtype C3 with sensitive and resistant cells mixed was excluded in the analysis. Conversely, RB1 mutation, occurred more frequently in the resistant subtypes. Therefore, C1 group, consisting mainly of SKCM (19.7%) and COREAD (12.4%) (Figure 1D), was significantly sensitive to Refametinib due to its high occurrence of BRAF and KRAS mutations and low occurrence of RB1 mutations (Figure 2H). Other types of MEK inhibitors, including PD0325901, Trametinib, Selumetinib, and CI-1040, further confirmed that BRAF, KRAS, and RB1 mutations were robustly connected to pharmacological subtypes, contributing to the drug sensitivity of MEK inhibitors (Supplementary Figure S3). Our analysis revealed that the varied distribution of genomic alterations across

pharmacological subtypes could lead to their differences in response to anticancer therapies.

Expression Alterations Associated With Pharmacological Subtypes

We also associated mRNA expression with the pharmacological subtypes. The connection between gene expression and pharmacological subtypes was determined by *p* values calculated with the student *t* test and corrected with the Benjamini-Hochberg method for controlling the false discovery rate. If the subtypes with higher drug sensitivity had higher gene expression, we defined this connection as sensitive. If the opposite was true, we defined the connection as resistant. The

connections between genes and drugs constituted a network (Figure 3A). To narrow down the genes whose expression levels were directly associated with drug sensitivity, we only kept genes which act in core cancer pathways (Supplementary Table S1), and whose Pearson correlation between drug sensitivity and gene expression were greater than 0.3 in the sensitive subtypes. The hub genes with the most connections to the drugs, including known cancer genes WWTR1, JUN, BCL9L and other genes NQO1, CAPN2, and PERP et al., were highlighted in the network.

We first investigated the distribution of genes on their connected drugs (Figure 3B). Our analysis identified 684 genes whose expression was connected to at least two drugs' pharmacological subtype. Among them, 4.97%, 14.04%, and 24.85% of genes respectively had connections with greater than 20, 10, and 5 drugs. The top 30 genes connected to more than 20 drugs, 25% of which were annotated as cancer genes (Figure 3C; Supplementary Figure S4). These cancerous percentages for the top 20 ($p = 4.37e-06$, Fisher's exact test), and top 30 genes ($p = 5.84e-52$, Fisher's exact test) were significantly higher than the background. The top genes with their expression alterations associated with subtypes were different from those with genomic alterations associated with subtypes (Figures 2D, 3D). Among them, WWTR1, BCL9L, and JUN ranked the top three with connections to 26 drugs and 25 drugs (Figure 3D).

We then investigated the distribution of drugs on their connected genes. The number of genes whose expression was associated with the subtypes varied from 0 to 223 across the drugs, with 4.65% drugs had more than 100 gene connections (Figure 3E). Comparing with genomic alterations, there were more genes whose expression was connected to the drugs. Moreover, genes whose expression was sensitively connected with drug resistance across pharmacological subtypes were significantly more abundant than genes that were resistant connected ($p = 0.0462$, student t test) (Figure 3F). The top 20 drugs with the most connections to genes included Nilotinib, affecting the ABL signaling, PHA-793887 and NPK76-II-72-1, affecting cell cycle, Lapatinib and KIN001-236, affecting RTK signaling pathway, Navitoclax, affecting Apoptosis regulation, and some compounds such as Ara-G, XMD14-99 et al., affecting multiple or unknown pathways (Figure 3G).

Methotrexate, as an example, is a chemotherapy that specifically acts during DNA and RNA synthesis, and cancer cells were classified into seven subtypes (Figure 3H) in terms of drug sensitivity to methotrexate. The first subtype, C1, was most sensitive to the drug, and the seventh subtype, C7, was the most resistant to the drug (Figure 3H). C1 included 12 histological subtypes, with the top 2 shown in the bottom of Figure 3H, ALL and DLBC. C2 and C3 were composed of 18 and 19 histological subtypes, respectively, with SCLC both included in the top 2 (Figure 3H). A total of 38 genes with high expression in sensitive group (C1, C2, and C3) and low expression in resistant group (C5, C6, and C7), which had sensitive connections to pharmacological subtypes of methotrexate, were selected (Figure 3I). Pathway enrichment analysis showed that these genes were enriched in pathways including cell cycle and

DNA replication et al. (Figure 3J), suggesting that the ectopic activation of cell cycle led to the sensitive response of ALL, DLBC, LAML, and SCLC to Methotrexate.

Redefining the Similarity of Cells Based on Pharmacological Subtypes

Pharmacological subtypes provide us a new point of view to redefine the similarity of cells purely from therapeutic concordance. Thus, we calculated the similarity of pairwise cells based on pharmacological subtypes of drugs. If there are quite a few cases that two cells are in the same sensitive or resistant pharmacological subtypes of drugs, these cells gain a high similarity. Otherwise, they gain a low similarity. The similarity of cells defined a hierarchical cluster as shown in Figure 4A. We further applied the Fisher's exact test to observe if the cells in the same clusters were from the same histological subtypes (Figures 4B–D). The points with black border in the figure represent statistical significance ($p < 0.05$), with the smaller p values shown in red.

The results showed that the cancer cell lines were respectively divided into 2, 4, and 15 clusters at the different hierarchical cluster levels (Figure 4A). Two clusters at the high level were reflected by two distinct patterns of similarity (Figure 4A). One cluster (Cluster H1) was significantly correlated with solid tumors including BRCA, COREAD, LUAD, MESO, OV, and PAAD, and the other (Cluster H2) hematological tumors including ALL, DLBC, LAML, LCML, and MM (Figure 4B). Here, the point with black border represents statistical significance ($p < 0.05$, Fisher's exact test) of the association between the cluster and cancer type. So in a holistic view, the different types of hematological tumors shared a similar drug sensitivity profile, which was quite different from the profile of solid tumors. However, for a given cancer type (the right panel of Figure 4B), such as ACC, BLCA, CESC, CLL, or ESCA et al., its cells were unbiasedly distributed in two clusters ($p > 0.05$, Fisher's exact test), showing the treatment heterogeneity for the cells from the same tissue origin. At the median hierarchical levels, 4 clusters were identified (Figure 4C). Among them, Cluster M3 was specifically associated with haematological tumors ($p < 0.001$, Fisher's exact test) including ALL, DLBC, LAML, LCML, and MM. The treatment heterogeneity was still observed in cancers.

When cell lines were divided into 15 clusters at the lower hierarchical level (Figure 4D), we observed that the cells from ACC, CLL, STAD, or UCEC were unbiasedly distributed in the different clusters, still showing their treatment heterogeneity. Interestingly, the cells from some cancers such as HNSC, KIRC, and SCLC were enriched in more than 2 clusters with distinctly different drug response. For instance, HNSC cells were enriched in Clusters L12, L4 and L5; KIRC cells were enriched in Clusters L15 and L4; SCLC cells were enriched in Clusters L13, L3, and L7. Interestingly, the SCLC cells in Cluster L13 shared a more similar profile with hematological tumors including DLBC, LAML, and LCML. To investigate the underlying mechanism of such treatment heterogeneity in SCLC, we applied DESeq2 analysis to call the differentially expressed genes between

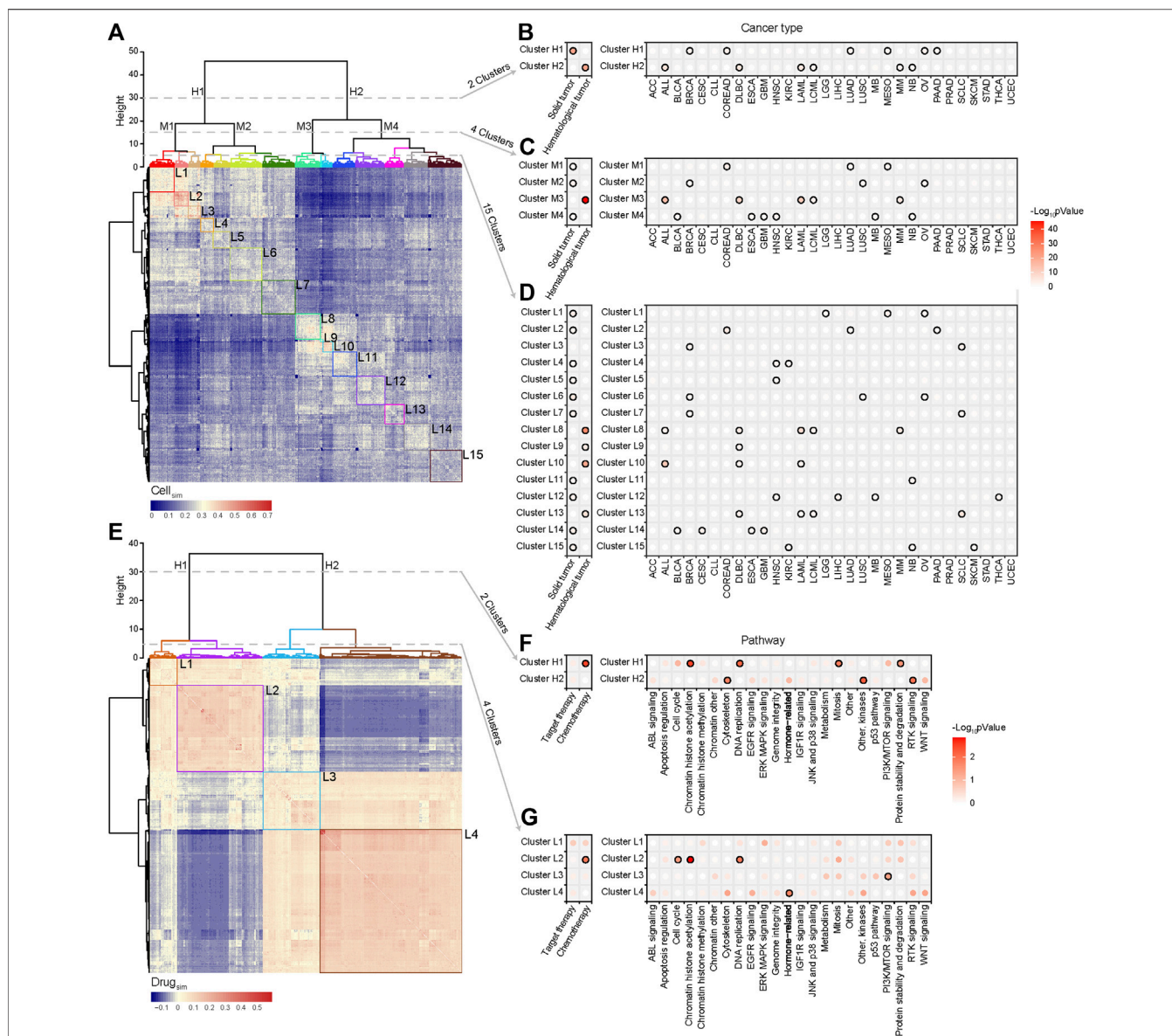


FIGURE 4 | Cell similarity and drug similarity based on pharmacological subtypes. **(A)** Hierarchical clustering of a cell similarity matrix which was generated based on drug response of cells to drugs, with the red color indicating high similarity, and the blue color indicating low similarity between two cell lines. Clusters were formed by partitioning of the dendrogram at a high, median and low height respectively, which indicates the distance between the clusters. **(B)** Two clusters (H1, H2) were formed by cutting the dendrogram at a height of 30. Points with black border represent statistical significance ($p < 0.05$) determined by Fisher's exact test. Almost all histological subtypes of blood cancer were grouped together. **(C)** Four clusters (M1, M2, M3 and M4) were formed by cutting the dendrogram at a height of 15. **(D)** Fifteen clusters (L1-15) were formed by cutting the dendrogram at a height of 5. Cancers with the same tissue of origin and organ did not cluster together. **(E)** Hierarchical clustering of the drug similarity profiles with the red color indicating a high similarity and the blue color indicating a low similarity. Clusters were formed by partitioning of the dendrogram at a high and low height respectively. **(F)** Two clusters (H1, H2) were formed by cutting the dendrogram at a height of 30. Points with black border represents statistical significance ($p < 0.05$) determined by Fisher's exact test. Chemotherapy related drugs were grouped together. **(G)** Four clusters (L1, L2, L3, and L4) were formed by cutting the dendrogram at a height of 5. Cytoskeleton, EGFR signalling, Hormone-related and WNT signalling pathways cluster together.

Cluster L13 and Cluster L3, and further identified the differential functions via gene set enrichment analysis (GSEA) (Subramanian et al., 2005). The results showed that the functions like DNA replication, RNA transport et al. were significantly upregulated in Cluster L13 cells, while in Cluster L3 cells the functions like immune response, antigen processing and presentation et al. were mainly upregulated (Supplementary Figure S5A). Moreover,

MYC, as an oncogenic transcription factor of cell growth and proliferation via enhancing the cell cycle regulated genes (Oster et al., 2002), was significantly upregulated in Cluster L13 than in L3, with a log2-fold change of 4.28 (Supplementary Figure S5B). These suggested that SCLC cell lines in Cluster L13 were higher Myc-dependent than those in Cluster L3, characterized with increased cell proliferation and evasion of immune response.

Redefining the Similarity of Drugs Based on Pharmacological Subtypes

The similarity of pairwise drugs was also defined as the summarized similarity of drug response across the same cells which were tested for both drugs. For two drugs, if there are quite a few of cell pairs which are in the sensitive (or resistant) pharmacological subtypes of one drug are also in the sensitive (or resistant) subtypes of the other drug, they gain a high similarity. Otherwise, they gain a low similarity. We then generated a drug similarity profile across 367 drugs, and similarly, a hierarchical clustering algorithm was used to group drugs based on drug similarity (Figure 4E). The results showed that at the high level, drugs were divided into two clusters, demonstrating two different similarity patterns. The first cluster (Cluster H1) mostly included chemotherapy drugs such as those effecting in DNA replication and mitosis, and the second cluster was mixed with target and chemotherapy drugs, specifically enriched by the drugs on RTK signaling, other kinases, and cytoskeleton (Figure 4F). When the drugs were divided into 4 clusters (Figures 4E,G), we observed that the drugs enriched in L2 included chemotherapy drugs effecting in cell cycle, chromatin histone acetylation, and DNA replication, the drugs enriched in L3 were functioning in PI3K/mTOR signaling pathway, and the drugs enriched in L4 were hormone related ones. Moreover, L4, accounting for a massive part of drugs, also included the drugs targeting the different types of signaling pathways such as ABL, EGFR, Jak, and P38, RTK, and WNT signalling. These drugs did not reach the significant level but provided the obvious association.

An Access to Pharmacological Subtypes of Drugs

To enable knowledge sharing and reuse, we developed a database (<http://www.hywanglab.cn/dtdb/>), Supplementary Figure S6), which provided the comprehensive knowledge of pharmacological subtypes of drugs in this study, for supporting the further exploration of therapeutic response-based classification of tumors. The web interface of the database is user-friendly and allows users to search, browse and download data. A record of the database includes the following information: the pharmacological subtypes of a drug, the drug sensitivity values of each subtype, the distribution of histological subtypes in each pharmacological subtype, genomic and expression alterations connected to subtypes. In addition, the users can acquire the other information of a drug, a gene, or a specific cancer type in terms of our analysis. For example, given a drug, the other drugs with the most similar drug response across cancer cells are provided. Similarly, given a cancer cell line, the other cell lines with the most similar drug response across drugs are provided. Given a histological cancer type, such as SCLC, the pharmacological subtypes with the distinct treatment response were depicted, as well as the underlying mechanism of such treatment heterogeneity. The resources are expected to support the researches of precision medicine from a purely therapy-oriented perspective.

DISCUSSION

In this study, we integrated drug response information to pharmacologically reclassify tumor subtypes, as well as identify intrinsic concordant molecular mechanisms. Unlike the many studies that have aimed to unravel cancer heterogeneity by defining subtypes within the same tissues, our study aimed to systematically uncover that cancers across disparate tissues of origin can belong to one pharmacological subtype, benefiting from similar anticancer therapies, because they share a common molecular mechanism of oncogenesis. Besides, we also developed the new measures to redefine cell similarity and drug similarity from the therapeutic concordance, which provided a new point of view to study cancer heterogeneity. The similarity of cells further depicted that cells from the different origin of tissue could share the similar responses of drugs; likewise, that cells from the same origin of tissue could have distinct drug responses, thus indicating the new subtypes.

Our analysis identified, for instance, that although SKCM and COREAD are histologically different, both belonged to one pharmacological subtype, the MEK1/2 inhibitor C1 subtype, and were similarly sensitive to MEK1/2 inhibitors (Figures 1A–D). By connecting genomic alterations with pharmacological subtypes, mutations in the oncogene BRAF or KRAS were found to be overwhelmingly more frequent in the C1 group (Figure 2H). BRAF and KRAS are the kinases upstream of MEK1/2 that transmits the signals down through MEK1/2 without other major signalling branches (McCubrey et al., 2007; Simanshu et al., 2017). Our results illustrated that MEK inhibitors were able to effectively block aberrantly activated signals from BRAF or KRAS mutations. Further investigation by cBioportal (<https://www.cbioportal.org/>) of TCGA patients also revealed that approximately 60% of THCA (thyroid carcinomas) and 50% of SKCM (melanomas) harboured BRAF mutations, and 65% of PAAD (pancreas) and 40% COREAD (colorectal carcinomas) possessed KRAS mutations, a rate that was strikingly higher than that of other cancers (Supplementary Figures S7A,B). Given the high mutant frequency of BRAF in THCA and SKCM, and KRAS in PAAD and COREAD, such 4 cancer types were significantly enriched in C1 pharmacological subtype of MEK1/2 inhibitor Refametinib (THCA, $p = 1.90e-2$; SKCM, $p = 1.23e-14$; PAAD, $p = 3.01e-4$; COREAD, $p = 3.84e-5$; Hypergeometric Distribution Test) (Supplementary Table S2). In contrast, mutations in the oncogenes RB1 was found to be less frequent in the C1 group than in other groups. RB1 is a negative regulator of the cell cycle, with its active hypophosphorylated form binding the transcription factor E2F1 (Chen et al., 2009). RB1 mutations lead to the ectopic activation of the cell cycle, which cannot be controlled using MEK inhibitors. This may be because RB1, located downstream of MEK1/2, gain ectopic activation independent of upstream stimulation. In another instance, therapeutic concordance was identified across histologically disparate blood cancers in regard to methotrexate, a chemotherapy drug specially acting during DNA and RNA synthesis (Figure 3H). This is well known that Methotrexate is used in haematological malignancies but less so in solid cancers (Kozmiński et al., 2020). Functional analysis of the sensitively

correlated genes showed that haematological tumors were ectopically activated in cell cycle than solid tumors (**Figures 3I,J**). The molecular basis of such extraordinary activation of the cell cycle in haematological tumors could explain their good response to chemotherapy drugs, which work mainly by inhibiting mitosis and cell division. This demonstrates pharmacological subtypes benefiting from similar anticancer therapies can be due to common molecular mechanisms related to tissue-specific gene expression and pathways.

We also investigated the treatment heterogeneity for six types of histological cancers, including LUAD, BRCA, SKCM, SCLC, COREAD, and HNSC. High treatment heterogeneity in LUAD was observed compared with other types of cancers, suggesting that anticancer therapies for LUAD should be more complicated than those for other types of cancers (**Figures 1F–H**). Moreover, the increased number of subtypes of EGFR inhibitors in LUAD, and of drugs targeting in ERK/MAPK signalling in SKCM, indicated there was high treatment heterogeneity when EGFR inhibitors were applied into the therapies of LUAD patients and when ERK/MAPK signalling targeting drugs were applied into SKCM patients (**Figure 1H**). We also observed the low treatment heterogeneity for HNSC (**Figures 1F,G**). However, our findings were only based on the current information provided by the CGP dataset. The low treatment heterogeneity could be due to the fact that active drugs in this histological cancer are potentially still missing. Drug development in future may increase the treatment heterogeneity as well.

Our analysis showed that genes whose genomic/expression alterations frequently connect to anticancer drugs had a higher likelihood of being cancer genes. At genomic level, the top 10 genes were connected to more than 10 drugs, 30% of which were annotated as cancer censor genes. At expression level, each of genes connected to more drugs and the top genes had higher cancerous percentages. This indicated that some genes currently not identified as cancer genes could be causally implicated in cancer. In our analysis, KCTD1_PCAT18, TNFRSF9, and LTB, as the top three most connected genes, were worth for further investigation. KCTD1 (potassium channel tetramerization domain containing 1) was reported to regulate the Wnt/ β -catenin pathway (Li et al., 2014; Hu et al., 2020), and PCAT18 (prostate cancer-associated transcript 18, lncRNA), was found to promote the progression of colorectal or gastric cancer through miR-759 or miR-135b (Hu et al., 2020; Zhang et al., 2020), thus KCTD1_PCAT18 fusion may resulting in the development of cancer. TNFRSF9 (tumor necrosis factor receptor superfamily member 9), also known as 4-1BB and CD137, is an immune co-stimulatory receptor. TNFRSF9 is expressed on activated immune cells including natural killer (NK) cells, effector T cells and antigen presenting cells, among them dendritic cells, macrophages, and B cells (Fröhlich et al., 2020). LTB (Lymphotoxin Beta) is a type II membrane protein of the TNF family. Giuseppina et al. found that a decrease in lymphotoxin- β production by tumor cells was associated with a loss of follicular dendritic cell phenotype and diffuse growth of follicular lymphomas (Pepe et al., 2018). Recently, the gene was found to be associated with immune infiltration of breast and endometrial cancer tumors (Ding et al., 2020; Terkelsen et al., 2020). The analysis using the TCGA dataset showed that

TNFRSF9 was altered in approximately 5% of cases of Cholangiocarcinoma, adrenocortical carcinoma (ACC) and lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), with deep deletion being the dominant alteration (**Supplementary Figure S8A**). In addition, more than 14% of DLBC patients had altered LTB genes, with mutation and deletion being the major alterations (**Supplementary Figure S8B**). Therefore, the importance of the roles of these genes in cancers could be underappreciated.

Our analysis also showed that some drugs connected to many known cancer genes (**Figure 2G**). This suggested that these drugs are unspecific or the genes are important for many different pathways. So these drugs may not be ideal from a therapeutic perspective. Moreover, genes whose alterations were sensitively connected with drug resistance were significantly more frequent than those resistant connected (**Figures 2F, 3F**). A sensitive connection meant that genomic alterations or high expression indicated increased drug sensitivity. In this way, genes with a sensitive connection could be indicators of therapeutic efficiency for connected drugs. Genes with resistant connections could be potential biomarkers for developing new therapies to overcome the resistance of connected drugs. Our analysis suggests that most of the genes whose expressions are altered during cancer initiation and development (driver genes or not) are located in ectopically activated pathways that could be controlled by anticancer drugs. Only a proportion of altered genes are outside of the cellular pathways targeted by anticancer drugs, and their alterations may maintain the growth signals of cancer cells when anticancer drugs are used. Therefore, these genes may be potential therapeutic biomarkers for overcoming drug resistance to the connected drugs.

Based on pharmacological subtypes, the cell similarity and the drug similarity could be re-defined (**Figure 4**). The cells from the same origin of tissue can be dispersed in the different pharmacological subtypes, showing the completely different response to the same drug, because of their intrinsically molecular heterogeneity. For instance, HNSC, KIRC, and SCLC cells were separated into more than two clusters with distinct drug response (**Figure 4D**). Specifically, the SCLC cells with more malignant signatures, in particular, higher expression levels of MYC, upregulated pathways involving DNA replication and cell cycle and downregulated pathways on immune response, were far away from the other SCLC cells and close to hematological tumors, instead. Interestingly, by connecting to expression alterations, our analysis further confirmed that SCLC cells more like hematological tumors had the characteristic of the ectopic activation of cell cycle, which led to their sensitive to a chemotherapy Methotrexate (**Figures 3H–J**). Our pharmacological analysis has recognized a new subtype of SCLC cells, with the high expression of an oncogenic transcription factor MYC as a marker, can benefit from chemotherapy than other SCLC subtypes. This finding was consistent with the latest studies (Bian et al., 2020; Ireland et al., 2020). Therefore, the subtypes identified by re-classifying the same origin of tissue using pharmacological data are worth to be further investigated. In addition, our analysis of the drug similarity unraveled that the drugs were categorized into two major groups, wherein the chemotherapeutic drugs tended to share the similar pattern (**Figures 4E,F**).

There are several limitations to this study. We measured the similarity of cells/drugs on the scale of all drugs/cells, so the similarity at the local space might be underestimated. Moreover, there is lack of experimental validation for our findings. To overcome these limitations, further proof of concept studies that reclassifying tumors based on therapeutic responses independent of histology subtypes is warranted.

In summary, this study demonstrates how pharmacogenomic data can be used to systematically classify cancers in terms of response to various compounds and provides us with a purely therapy-oriented perspective to view tumor classifications independent of histology subtypes. Moreover, the knowledge of pharmacological subtypes of 367 drugs are available via our website, providing the resources for precision medicine in the perspective of therapeutic response-based re-classification of tumors.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancerrxgene.org/downloads/bulk_download.

REFERENCES

- Arozarena, I., and Wellbrock, C. (2019). Phenotype Plasticity as Enabler of Melanoma Progression and Therapy Resistance. *Nat. Rev. Cancer* 19, 377–391. doi:10.1038/s41568-019-0154-4.3
- Bedard, P. L., Hansen, A. R., Ratain, M. J., and Siu, L. L. (2013). Tumour Heterogeneity in the Clinic. *Nature* 501, 355–364. doi:10.1038/nature12627.1
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.18
- Bian, X., Wang, X., Zhang, Q., Ma, L., Cao, G., Xu, A., et al. (2020). The MYC Paralog-PARP1 Axis as a Potential Therapeutic Target in MYC Paralog-Activated Small Cell Lung Cancer. *Front. Oncol.* 10, 565820. doi:10.3389/fonc.2020.565820.34
- Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45, 1113–1120. doi:10.1038/ng.2764.16
- Chen, H.-Z., Tsai, S.-Y., and Leone, G. (2009). Emerging Roles of E2Fs in Cancer: an Exit from Cell Cycle Control. *Nat. Rev. Cancer* 9, 785–797. doi:10.1038/nrc2696.25
- Ding, H., Fan, G.-L., Yi, Y.-X., Zhang, W., Xiong, X.-X., and Mahgoub, O. K. (2020). Prognostic Implications of Immune-Related Genes' (IRGs) Signature Models in Cervical Cancer and Endometrial Cancer. *Front. Genet.* 11, 725. doi:10.3389/fgene.2020.00725.33
- Ding, X., He, M., Chan, A. W. H., Song, Q. X., Sze, S. C., Chen, H., et al. (2019). Genomic and Epigenomic Features of Primary and Recurrent Hepatocellular Carcinomas. *Gastroenterology* 157, 1630–1645. doi:10.1053/j.gastro.2019.09.005.2
- Drilon, A., Laetsch, T. W., Kummar, S., DuBois, S. G., Lassen, U. N., Demetri, G. D., et al. (2018). Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. *N. Engl. J. Med.* 378, 731–739. doi:10.1056/NEJMoa1714448.15
- Fröhlich, A., Loick, S., Bawden, E. G., Fietz, S., Dietrich, J., Diekmann, E., et al. (2020). Comprehensive Analysis of Tumor Necrosis Factor Receptor TNFRSF9 (4-1BB) DNA Methylation with Regard to Molecular and Clinicopathological Features, Immune Infiltrates, and Response Prediction to Immunotherapy in Melanoma. *EBioMedicine* 52, 102647. doi:10.1016/j.ebiom.2020.102647.30
- Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., et al. (2012). Systematic Identification of Genomic Markers of Drug Sensitivity in Cancer Cells. *Nature* 483, 570–575. doi:10.1038/nature11005.17

AUTHOR CONTRIBUTIONS

HW conceived the hypothesis. YX, JZ, and HW designed and performed the data analysis. ZC, WL, YD, XC, and TW collected and preprocessed the data. HW, YX, JZ, JK, and FZ wrote and revised the manuscript.

FUNDING

This work was supported by grants from the National Key Research and Development Program (2017YFC0908500 to HW) and the National Natural Science Foundation of China (31771469 and 31571363 to HW; 81772465 to FZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.773101/full#supplementary-material>

- Gay, C. M., Stewart, C. A., Park, E. M., Diao, L., Groves, S. M., Heeke, S., et al. (2021). Patterns of Transcription Factor Programs and Immune Pathway Activation Define Four Major Subtypes of SCLC with Distinct Therapeutic Vulnerabilities. *Cancer Cell* 39, 346–360. doi:10.1016/j.ccell.2020.12.014.5
- Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., et al. (2014). Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell* 158, 929–944. doi:10.1016/j.cell.2014.06.049.14
- Hu, L., Chen, L., Yang, L., Ye, Z., Huang, W., Li, X., et al. (2020). KCTD1 Mutants in Scalp-ear-nipple S-syndrome and AP-2 α P59A in Char S-syndrome Reciprocally A-brogate T-her I-nteractions, but C-an R-egulate Wnt/ β -catenin S-signaling. *Mol. Med. Rep.* 22, 3895–3903. doi:10.3892/mmr.2020.11457.27
- Ireland, A. S., Micinski, A. M., Kastner, D. W., Guo, B., Wait, S. J., Spainhower, K. B., et al. (2020). MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. *Cancer Cell* 38, 60–78. doi:10.1016/j.ccell.2020.05.001.35
- Ko, T. K., Javed, A., Lee, K. L., Pathiraja, T. N., Liu, X., Malik, S., et al. (2020). An Integrative Model of Pathway Convergence in Genetically Heterogeneous Blast Crisis Chronic Myeloid Leukemia. *Blood* 135, 2337–2353. doi:10.1182/blood.2020004834.4
- Koźmiński, P., Halik, P. K., Chesori, R., and Gniazdowska, E. (2020). Overview of Dual-Acting Drug Methotrexate in Different Neurological Diseases, Autoimmune Pathologies and Cancers. *Int. J. Mol. Sci.* 21. doi:10.3390/ijms21103483.26
- Li, X., Chen, C., Wang, F., Huang, W., Liang, Z., Xiao, Y., et al. (2014). KCTD1 Suppresses Canonical Wnt Signaling Pathway by Enhancing β -catenin Degradation. *PLoS one* 9, e94343. doi:10.1371/journal.pone.0094343.28
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8.20
- Lu, H. M., Li, S., Black, M. H., Lee, S., Hoiness, R., Wu, S., et al. (2019). Association of Breast and Ovarian Cancers with Predisposition Genes Identified by Large-Scale Sequencing. *Jama Oncol.* 5, 51–57. doi:10.1001/jamaoncol.2018.2956
- Marcoux, N., Gettinger, S. N., O'Kane, G., Arbour, K. C., Neal, J. W., Husain, H., et al. (2019). EGFR-mutant Adenocarcinomas that Transform to Small-Cell Lung Cancer and Other Neuroendocrine Carcinomas: Clinical Outcomes. *Jco.* 37, 278, 285. doi:10.1200/Jco.18.01585.7
- McCubrey, J. A., Steelman, L. S., Chappell, W. H., Abrams, S. L., Wong, E. W. T., Chang, F., et al. (2007). Roles of the Raf/MEK/ERK Pathway in Cell Growth, Malignant Transformation and Drug Resistance. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1773, 1263–1284. doi:10.1016/j.bbamcr.2006.10.001.23

- Menzer, C., Menzies, A. M., Carlino, M. S., Reijers, I., Groen, E. J., Eigentler, T., et al. (2019). Targeted Therapy in Advanced Melanoma with Rare BRAF Mutations. *Jco* 37, 3142–3151. doi:10.1200/Jco.19.00489.6
- Oster, S. K., Ho, C. S. W., Soucie, E. L., and Penn, L. Z. (2002). The Myc Oncogene: Omplex. *Adv. Cancer Res.* 84, 81–154. doi:10.1016/s0065-230x(02)84004-0.22
- Pepe, G., Di Napoli, A., Cipitelli, C., Scarpino, S., Pillozzi, E., and Ruco, L. (2018). Reduced Lymphotoxin-Beta Production by Tumour Cells Is Associated with Loss of Follicular Dendritic Cell Phenotype and Diffuse Growth in Follicular Lymphoma. *J. Pathol. Clin. Res.* 4, 124–134. doi:10.1002/cjp.2.97.31
- Petralia, F., Tignor, N., Reva, B., Koptyra, M., Chowdhury, S., Rykunov, D., et al. (2020). Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* 183, 1962–1985. doi:10.1016/j.cell.2020.10.044.13
- Prat, A., Parker, J. S., Fan, C., and Perou, C. M. (2012). PAM50 Assay and the Three-Gene Model for Identifying the Major and Clinically Relevant Molecular Subtypes of Breast Cancer. *Breast Cancer Res. Treat.* 135, 301–306. doi:10.1007/s10549-012-2143-0.8
- Rueda, O. M., Sammut, S.-J., Seoane, J. A., Chin, S.-F., Caswell-Jin, J. L., Callari, M., et al. (2019). Dynamics of Breast-Cancer Relapse Reveal Late-Recurring ER-Positive Genomic Subgroups. *Nature.* 567, 399, 404. doi:10.1038/s41586-019-1007-8.10
- Schmitz, R., Wright, G. W., Huang, D. W., Johnson, C. A., Phelan, J. D., Wang, J. Q., et al. (2018). Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N. Engl. J. Med.* 378, 1396–1407. doi:10.1056/NEJMoa1801445.11
- Simanshu, D. K., Nissley, D. V., and McCormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. *Cell* 170, 17–33. doi:10.1016/j.cell.2017.06.009.24
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102.19
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 47, D941–d947. doi:10.1093/nar/gky1015.21
- Terkelsen, T., Russo, F., Gromov, P., Haakensen, V. D., Brunak, S., Gromova, I., et al. (2020). Secreted Breast Tumor Interstitial Fluid microRNAs and Their Target Genes Are Associated with Triple-Negative Breast Cancer, Tumor Grade, and Immune Infiltration. *Breast Cancer Res.* 22, 73. doi:10.1186/s13058-020-01295-6.32
- Waks, A. G., and Winer, E. P. (2019). Breast Cancer Treatment. *Jama* 321, 288–300. doi:10.1001/jama.2018.19323.9
- Zhang, X.-Z., Mao, H.-L., Zhang, S.-J., Sun, L., Zhang, W.-J., Chen, Q.-Z., et al. (2020). lncRNA PCAT18 Inhibits Proliferation, Migration and Invasion of Gastric Cancer Cells through miR-135b Suppression to Promote CLDN11 Expression. *Life Sci.* 249, 117478. doi:10.1016/j.lfs.2020.117478.29

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Xu, Zheng, Cai, Li, Köhler, Dai, Cheng, Wu, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

ACC Adrenocortical carcinoma	LUSC Lung squamous cell carcinoma
ALL Acute lymphoblastic leukemia	MB Medulloblastoma
BLCA Bladder urothelial carcinoma	MESO Mlistdefinitionesothelioma
BRCA Breast invasive carcinoma	MM Multiple myeloma
CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma	NB Neuroblastoma
CLL Chronic lymphocytic leukemia	OV Ovarian serous cystadenocarcinoma
COREAD Colon adenocarcinoma and rectum adenocarcinoma	PAAD Pancreatic adenocarcinoma
DLBC Lymphoid neoplasm diffuse large B-cell Lymphoma	PRAD Prostate adenocarcinoma
ESCA Esophageal carcinoma	SCLC Small cell lung cancer
GBM Glioblastoma multiforme	SKCM Skin cutaneous melanoma
HNSC Head and neck squamous cell carcinoma	STAD Stomach adenocarcinoma
KIRC Kidney renal clear cell carcinoma	THCA Thyroid carcinoma
LAML Acute myeloid leukemia	UCEC Uterine corpus endometrial carcinoma
LCML Chronic myelogenous leukemia	CGP The Cancer Genome Project
LGG Brain lower grade glioma	IC50 Half maximal inhibitory concentration
LIHC Liver hepatocellular carcinoma	TCGA The Cancer Genome Atlas Program
LUAD Lung adenocarcinoma	FDA U.S. Food and Drug Administration
	RTK Receptor tyrosine kinase
	MEK Mitogen-activated protein kinase kinase.