

## Article

# A Transformative Technology Linking Patient's mRNA Expression Profile to Anticancer Drug Efficacy

Chen Yeh <sup>1,\*</sup> , Shu-Ti Lin <sup>1</sup> and Hung-Chih Lai <sup>2</sup><sup>1</sup> OncoDxRx, Los Angeles, CA 91006, USA<sup>2</sup> Division of Hematology and Oncology, Department of Internal Medicine, Shin-Kong Wu Ho-Su Memorial Hospital, Taipei 111, Taiwan

\* Correspondence: cyeh.oncodxrx@gmail.com

**Simple Summary:** Innovative therapies matter only if they can reach and benefit patients. Of the 2 million new cases of cancer diagnosed in the U.S. last year, between 70% and 80% were estimated to be non-responders to precision medicine. In these cases, treatment options are limited. Even among the responders, their tumors will frequently develop drug resistance against targeted drugs. When the progressive disease occurs following standard therapy, oncologists enter into an educational guess cycle where therapeutic outcomes become unpredictable and patients' benefits are unsecured. Our mission here is to provide every patient with more genetics-tailored and effective treatment options. Our transformative gene-to-drug technology works by testing tumor gene activities from a patient's own blood to guide therapies that are most likely to benefit the patient non-responders with refractory or relapsed disease. We found this unprecedented technology can help match patients with more FDA-approved drugs and significantly improve patient outcomes. This one-of-a-kind innovation revolutionizes gene-to-drug mapping power and opens a new dimension of how cancer drugs can be better matched to patients.

**Abstract:** As precision medicine such as targeted therapy and immunotherapy often have limited accessibility, low response rate, and evolved resistance, it is urgent to develop simple, low-cost, and quick-turnaround personalized diagnostic technologies for drug response prediction with high sensitivity, speed, and accuracy. The major challenges of drug response prediction strategies employing digital database modeling are the scarcity of labeled clinical data, applicability only to a few classes of drugs, and losing the resolution at the individual patient level. Although these challenges have been partially addressed by large-scale cancer cell line datasets and more patient-relevant cell-based systems, the integration of different data types and data translation from pre-clinical to clinical utilities are still far-fetched. To overcome the current limitations of precision medicine with a clinically proven drug response prediction assay, we have developed an innovative and proprietary technology based on in vitro patient testing and in silico data analytics. First, a patient-derived gene expression signature was established via the transcriptomic profiling of cell-free mRNA (cfmRNA) from the patient's blood. Second, a gene-to-drug data fusion and overlaying mechanism to transfer data were performed. Finally, a semi-supervised method was used for the database searching, matching, annotation, and ranking of drug efficacies from a pool of ~700 approved, investigational, or clinical trial drug candidates. A personalized drug response report can be delivered to inform clinical decisions within a week. The PGA (patient-derived gene expression-informed anticancer drug efficacy) test has significantly improved patient outcomes when compared to the treatment plans without PGA support. The implementation of PGA, which combines patient-unique cfmRNA fingerprints with drug mapping power, has the potential to identify treatment options when patients are no longer responding to therapy and when standard-of-care is exhausted.



**Citation:** Yeh, C.; Lin, S.-T.; Lai, H.-C. A Transformative Technology Linking Patient's mRNA Expression Profile to Anticancer Drug Efficacy. *Onco* **2024**, *4*, 143–162. <https://doi.org/10.3390/onco4030012>

Academic Editor: Constantin N. Baxevanis

Received: 17 June 2024

Revised: 12 July 2024

Accepted: 12 July 2024

Published: 14 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** precision medicine; drug response prediction; cell-free mRNA; transcriptomic profiling; gene expression signature

## 1. Introduction

The deciphering of human genome sequence has expedited the genetic data-driven revolution toward precision medicine, which has delivered earlier diagnoses, more targeted and personalized treatment, and the real-time monitoring of therapeutic efficacies. Precision medicine promises improved health outcomes by providing the right therapy to the right patient, at the “first” time without delay [1]. The current standard of care is to utilize actionable genomic data to tailor therapy for each individual cancer patient. However, the reality of today’s precision medicine is that only 5–10% of cancer patients experience a clinical benefit from treatments matched to tumor DNA mutations via biomarker testing [2–4]. Although there are many factors underlying this modest success rate, improved drug response prediction will significantly benefit more patients, especially those non-responders to targeted therapy or immunotherapy [4–6].

Mega pre-clinical databases, for example, Genomics of Drug Sensitivity in Cancer (GDSC) [7,8] and Cancer Cell Line Encyclopedia (CCLE) [9] provide a full spectrum of the genomic profiles of somatic mutation, copy number aberration, structural variant, transcriptomic, and methylomic data, together with in vitro dose–response information to a large number of targeted and chemotherapy drugs. More importantly, there are clinical datasets which register the responses of real-life patients to monotherapy or combination therapy, e.g., The Cancer Genome Atlas (TCGA) and ClinicalTrials.gov. Nevertheless, the pre-clinical datasets enable drug response modeling, training, and prediction, in particular for many drugs, from various types of pre-clinical systems to patients [8,9]. Working on pre-clinical big data, current in silico analyses are usually aimed at building computational deep learning, deep neural network methods to predict drug response [10,11]. However, it remains challenging to integrate and interpret the diverse and large number of high-dimensional multiomic data points in a clinically relevant manner. Further, the complex cellular signaling networks that regulate the anticancer drug response are largely overlooked in digital computation and simulation, thereby losing the translatability to real-world patients [12,13]. A computational approach should be trained on relevant and standardized clinical data to achieve translatability; unfortunately, the available clinical datasets such as TCGA do not have sufficient patient records with drug response information.

Extensive studies have suggested that gene expression data are the most effective data type for drug response prediction [11,14,15]. Although gene expression profiles provide a machine learning model with deeper insight into the same sample and promise a better characterization of biological processes, this approach has several limitations. First, it will miss much-needed resolution at the individual patient level, which may limit its ability to predict personalized drug response. Second, the non-real-time gene expression patterns will misrepresent the dynamics of input genes and data. Third, sample-specific gene expression data have not been deployed yet. The digital modeling was used to show gene-relatedness in the context of cancer, but not for sample-specific prediction tasks (e.g., drug response prediction). Therefore, it is not appropriate to use the most important input gene sets for each sample.

In the clinical application of drug response prediction, our goal is to predict which drugs will most likely benefit the patient based on the patient’s own gene expression signature. Since clinical gene–drug datasets are largely unavailable, many modeling studies have focused on large pre-clinical pharmacogenomics datasets such as cancer cell lines as a surrogate to patients. A majority of the digital computation approaches are trained on cell line datasets and then tested on patient datasets [16,17]. However, cell lines even with the same genetic alterations often do not recapitulate a patient’s drug response due to the lack of an immune system and a tumor microenvironment (TME) [18]. Moreover, in

cell lines, the drug response is often measured by the IC50 or AUC (Area Under Curve), whereas in patients, it is often based on changes in the size of the tumor and measured by metrics such as response evaluation criteria in solid tumors (RECIST) [19]. Therefore, drug response prediction is a regression problem in cell lines but a classification problem in patients. Discrepancies in both the input and output pharmacogenomics datasets must be re-aligned and resolved, with an urgent need for a translational technology to bridge this gap.

In this study, we developed an innovative liquid biopsy cell-free mRNA (cfmRNA) based technology, called patient-derived gene expression-informed anticancer drug efficacy (PGA), for predicting cancer drug responses. It applied cfmRNA profiling to measure gene expression and established a cancer type-specific, patient-unique gene expression signature. The signature was then used to digitally query, search, match, categorize, and rank drug efficacies from a library of more than 700 anticancer drugs to identify the most effective drugs for the patient. Importantly, PGA was further prospectively and clinically tested on gene expression data from a real-life group of patients with refractory or relapsed non-small cell lung cancer (NSCLC) to identify potentially effective drugs. Our results demonstrated that the first-ever PGA platform, combining *in vitro* patient testing with *in silico* data computation, enabled us to analyze each patient's cfmRNA data in real-time to better match them with tailored treatments and drug combinations. These findings underscored the clinical utility of PGA and contributed to the advancement of drug response prediction.

## 2. Materials and Methods

### 2.1. Sample Processing and RNA Isolation

All the paired tissue and blood samples were purchased from iSpecimen (Lexington, MA, USA), Discovery Life Sciences (Huntsville, AL, USA), or Precision for Medicine (Carlsbad, CA, USA). Ten milliliters of EDTA whole blood from each patient was collected and spun at  $1100\times g$  for 10 min within one hour of collection to separate plasma. The plasma samples were double spun at  $16,000\times g$  for 10 min and aliquoted to cryogenic vials, and stored at  $-80\text{ }^{\circ}\text{C}$  until analyzed. Archival formalin-fixed, paraffin-embedded (FFPE) tissue samples were used to extract tumor RNA. All the tumor tissue samples were pathologically examined to have  $>50\%$  tumor nuclei and the median tumor content was 77.5%. All the plasma samples for cell-free mRNA (cfmRNA) extraction were subjected to one freeze–thaw cycle. In total, 400  $\mu\text{L}$  of double-spun plasma was used to extract cfmRNA using the MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit. For the FFPE tumor and adjacent normal tissues, RNA was extracted from 5 consecutive 10  $\mu\text{m}$  thick sections by MagMAX™ FFPE DNA/RNA Ultra Kit (Applied Biosystems, Foster City, CA, USA). Both plasma and tissue RNA extractions were performed on the KingFisher™ Duo Prime Purification System (Thermo Fisher Scientific, Waltham, MA, USA). The isolated RNA was quantified using the Qubit RNA HS Assay Kit and Qubit 2.0 Fluorometer (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. For quality control, the size distribution of the extracted RNA fragments was assessed using the RNA 6000 Pico kit on a 2100 Bioanalyzer Lab-on-a-Chip platform (Agilent Technologies, Santa Clara, CA, USA), and expressed as the percentage of fragments greater than 200 base pairs (DV200).

### 2.2. Reverse Transcriptase Quantitative PCR (RT-qPCR)

Double-stranded cDNA was synthesized from 1  $\mu\text{g}$  of total RNA using NEBNext RNA First Strand Synthesis Module and NEBNext Ultra™ II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instruction. For targeted plasma transcriptomic profiling, 9 TaqMan Gene Expression Arrays (Applied Biosystems, Foster City, CA, USA) covering 9 major signaling pathways of about 750 cancer-associated genes were employed. They were TaqMan array immune response #4414073, cell surface markers #4418754, DNA repair

mechanism #4418773, DNA methylation #4414127, transcription factors #4418784, p53 signaling #4414168, MAPK pathways #4414093, Molecular Mechanisms of Cancer #4418806, and tumor metastasis #4418743. Real-time qRT-PCR amplification and detection were performed with TaqMan Gene Expression Assay reagents in a QuantStudio 12K Flex System (Applied Biosystems, Foster City, CA, USA) using standard settings and cycling parameters. The 20  $\mu$ L reactions were carried out containing 10  $\mu$ L TaqMan Fast Advanced Master Mix and 10 ng of a cDNA template per well in a 96-well format.

The expression level of each individual gene in all the specimens was normalized to a reference RNA pool (made by pooling equal amounts of total RNA from each of the tumor-free specimens) used as a calibrator. In every qPCR run, ribosomal 18S RNA was included as the internal control. The relative changes in gene expression were determined by the  $\Delta\Delta$ Ct method using the Sequence Detection System (SDS) 2.1 software (Applied Biosystems, Foster City, CA, USA). The  $\Delta\Delta$ Ct approach quantifies the expression of the target gene normalized to an endogenous reference and relative to a calibrator, allowing direct comparison with the RNA-Seq data which the outputs are relative counts of transcripts.

### 2.3. FFPE Tissue RNA Sequencing

The sequencing library construction was performed using cDNA fragments of 250–300 bp in length. The pooled libraries were sequenced on an Illumina HiSeq platform (Illumina, San Diego, CA, USA) with 125 bp/150 bp paired-end reads generated. Raw sequencing data were smoothed by Cutadapt 4.0 to filter out the reads that were of low quality, low-read, and those containing adapters and sequencing artifacts. The clean reads were then aligned to the reference genome and counted the number of reads mapped to each gene using TopHat v2.0.12 and HTSeq v0.6.1, respectively. To correct for larger transcripts having higher read counts, all the sequencing data were normalized and expressed as Transcripts Per Million (TPM) or Reads/Fragments Per Kilo-base per Million mapped reads (RPKM/FPKM), which was calculated to determine relative gene expression levels [20]. Differential expression was digitally analyzed by the DESeq R package v1.18.0. Top high-confidence differentially expressed genes were identified by comparing the expression levels of all the transcripts in the TUMOR groups with those in the NORMAL group using the cutoff of  $|\log_2(\text{fold change})| > 1$  and an adjusted  $p$ -value  $< 0.05$ .

### 2.4. Single-Cell Gene Expression Profiling by RNA-Seq

The tumor tissues were digested with a human tumor dissociation kit (Miltenyi Biotec, Gaithersburg, MD, USA) following the manufacturer's protocol. If more than 5% of the dead cells in cell suspensions were indicated by trypan blue staining, the dead cells were filtered out using a dead cell removal kit (Miltenyi Biotec, Gaithersburg, MD, USA). Single-cell suspension was immediately used for RNA sequencing or frozen in the cryopreservation solution (90% fetal bovine serum and 10% dimethyl sulfoxide) at  $-80$  °C with cell concentration within 100–2000/ $\mu$ L.

The capturing and barcoding of single cells were conducted using the 10X Chromium platform (10X Genomics, Pleasanton, CA, USA). RNA-seq libraries were constructed following the instructions from the Chromium Single Cell 3' Reagent v3 Kits. Each lane of a 10X chip was loaded with approximately 5000 cells. The cells were partitioned into single-cell gel beads in emulsions (GEMs) inside the Chromium instrument, where full-length cDNA synthesis occurred. Following reverse transcription and cleanup, the cDNA from barcoded single-cell RNAs were amplified, and the 3' gene expression libraries were constructed. The cDNA pool corresponding to an insertion size of  $\sim$ 350–400 bp was then enriched. Sequencing libraries were quantified using Agilent Bioanalyzer High Sensitivity DNA chips (Agilent Technologies, Santa Clara, CA, USA) and pooled together to obtain similar numbers of reads from each single cell before sequencing on the NovaSeq 6000 S4 (Illumina, San Diego, CA, USA).

### 2.5. Single-Cell Spatial Transcriptomics Analysis

Single-cell RNA-Seq reads were mapped to the human reference genome (GRCh38) using Cell Ranger v1.1.0 pipeline with default settings. The resulting gene–cell matrices were subsequently imported into the Seurat (v3.1.5) R toolkit for further quality control and downstream analysis [21,22]. To search for clinically relevant genes co-localized and coexpressed with the selected PGA Lung biomarkers for drug efficacy prediction, we identified cell clusters using Seurat graph-based clustering methods at increasing resolutions to identify major cell types within a single-cell RNA-Seq dataset. We used marker genes—EGFR, KRAS, BRAF, MET, HER2, ALK, ROS1, and RET—to identify lung tumor cells; CD8, CD25, CD69, PD-1, CTLA-4 and B cell markers for immune cell population; KI67 and PCNA for highly proliferative cells; and SOX2, OCT4, KLF4, and MYC for cancer stem cells. Mesenchymal, stromal, vascular endothelial, and other cell-of-interest clusters can be further annotated based on canonical markers for further dimensionality reduction using the FindClusters function in the Seurat package. Cell clusters were identified at a resolution of 0.3 and annotated based on prior knowledge. We recognized a cluster with a minimum of 5% of the cells from the total cell population. If the marker genes are not our primary interest, we would leave the original clusters unchanged.

### 2.6. Correlation between Gene Expression and Drug Efficacy

A method of gene pathway/network generalization for drug response prediction is needed to take both pre-clinical and clinical samples during deployment. Therefore, we selected datasets and developed gene-drug correlation based on cancer cell lines, archived tumors, single-cell transcriptomics, and real-world patients. We employed the following resources for gene pathway generalization: Cancer Cell Line Encyclopedia (CCLE), The Genomics of Drug Sensitivity in Cancer (GDSCv1/2), The Cancer Therapeutics Response Portal (CTRPv2), EMBL-EBI Single Cell Expression Atlas, cBioPortal for Cancer Genomics, CREAMMIST database, Cancer Treatment Response Gene Signature database (CTR-DB), and The Cancer Genomic Atlas (TCGA). All the datasets were downloaded from the ORCESTR platform [23]. We focused on bridging cell line datasets to patient tumors because they are the missing link for translation from the pre-clinical to the clinical stage [24].

Correlation at the gene level: We applied two computational analytics to qualify the correlation between cell lines and the corresponding TCGA cohorts: (i) Spearman's correlation coefficient ( $\rho$ ) between every cancer cell line and its corresponding TCGA cohort was determined at the gene level. First, the TPM values of each transcript in a TCGA cohort were averaged per cohort. Then, for each TCGA cohort, Spearman's  $\rho$  was calculated between the averaged TPM values and those of the disease-matched cell lines across 20,053 protein-coding genes. (ii) The recapitulation of the TCGA cohort overexpressed gene profiles in cell lines was aligned and measured by gene set enrichment analysis (GSEA). We reasoned that the pattern of overactive genes that have an upregulated expression in a TCGA cohort can be considered as the cohort signature, and these overexpression profiles should be reflected by cell line models. For each cell line, we calculated with  $\log_2$  the transformation of the fold change in every gene expression relative to the disease baseline expression. In the end, the GSEA of the TCGA cohort overexpressed gene profiles was used for an association analysis against the data from the gene  $\log_2$  fold changes in cell lines. The final correlation outputs were reported as the normalized enrichment score (NES), with a positive value indicating a high correlation between a cell line and a disease-matched TCGA cohort.

Correlation at the pathway level: The activity of a total of 14 cancer-related pathways was interrogated using PROGENy, a pathway–response signature-based approach that is capable of deep data mining to obtain cancer-related pathway responsive genes [25], together with the CytoSig program which analyzes 43 cytokines gene expression profiles [26]. Both results were presented as z-scores to indicate the relative activities, with a  $p$ -value  $< 0.05$  considered as significant.

In the gene-to-drug mapping, the latent representations of the patient's tumor gene expression profile must then be aligned with each drug's latent representation through cell line data links. Low-rank multimodal fusion (LMF) was employed as the fusion method, and the output from this fusion was then passed to the final module which predicted the drug efficacy. LMF is a technique for combining multiple modalities in a neural network such that the latent representations of different features are forced to "interact" with each other [27]. LMF has shown higher performance than other fusion methods. It is especially important for modeling biology since it is known that various biomolecules in the cell interact with each other and thus must also be allowed to interact when modeling biology in silico.

### 2.7. Statistical Analysis

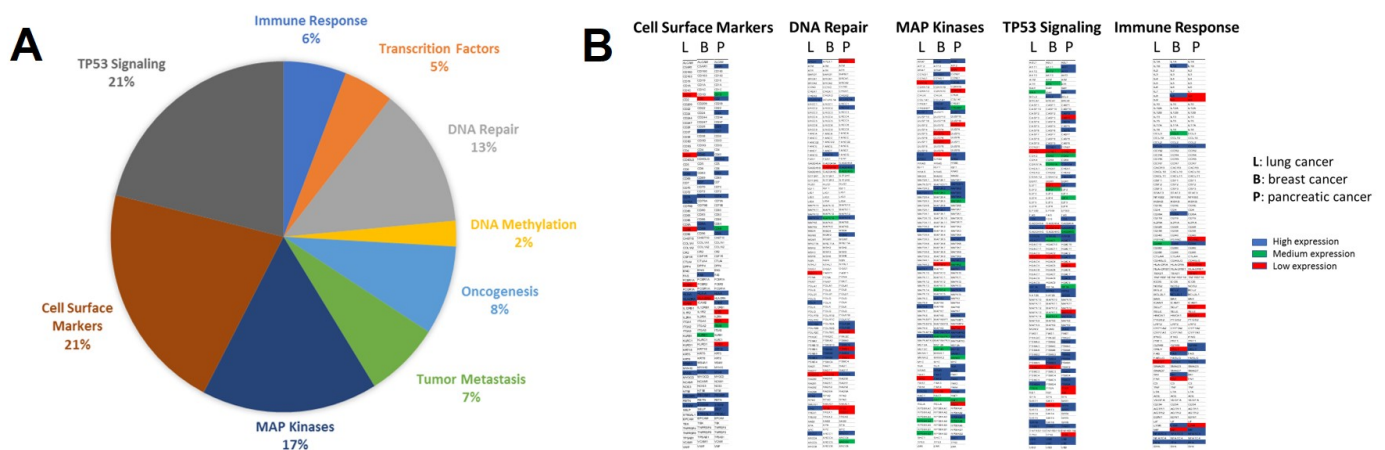
Data on patient outcomes were collected and analyzed from the date of treatment with or without PGA guidance to the time of the death or the date on which data were censored. All the statistical analyses were performed using the SPSS software v20.0 (SPSS Inc., Chicago, IL, USA). Progression-free survival (PFS) and overall survival (OS) were plotted with the Kaplan–Meier estimator and analyzed with the bilateral log-rank test. The combined effects of key variables in both PFS and OS were determined in multivariate analysis using Cox proportional hazards regression models.  $p < 0.05$  was considered a statistically significant difference.

## 3. Results

### 3.1. Cancer Type-Specific and Patient-Derived Gene Expression Profiles

Although RNA-Seq and microarrays were standard methods benchmarked for differential gene expression and prediction model development, the challenge of quantifying low-abundance, short half-life cfRNA species is compounded by the time-consuming and labor-intensive workflows. The requirement of high-quality and sufficient quantity of cfRNA also imposes a technical barrier on top of the interference from abundant globin mRNA and ribosomal RNA (rRNA). Although globin and rRNA removal strategies have mitigated some of these issues, they require a large amount of total cfRNA input and thus are not practical in clinical settings. To overcome these limitations, we applied multiplex RT-qPCR-based targeted transcriptomic profiling, followed by the quantitative analysis of cfRNA abundance by  $\Delta\Delta C_t$ , the difference in  $C_t$  values between reference gene (18S) and target gene, and normalized to the control samples.

Circulating cfRNA was isolated from the pooled plasma samples of patient cohorts with lung, pancreatic, or breast cancer. The targeted transcriptomic profiling of ~750 well-established cancer-associated genes was performed. These genes were selected by their key roles in major cancer signaling pathways: immune response (IR), cell surface markers (CSMs), transcription factors (TFs), DNA repair (DR), DNA methylation (DM), oncogenesis (ONC), tumor metastasis (TM), TP53 signaling (TS), MAP kinases (MKs). The distribution of detected cfRNA species from the lung cancer cohort was demonstrated in Figure 1A. We identified that the same percentage of genes belonged to cell surface markers and the TP53 signaling pathway (21%), 17% of the detected transcripts were members of the MAP kinase family, 13% were involved in DNA repair, 8% correlated with oncogenesis, 7% associated with tumor metastasis, 6% involved in immune response, 5% are transcription factors, and 2% related to DNA methylation.



**Figure 1.** Plasma cfRNA profiling by cancer type, functional cluster, and expression level. (A) The pie chart displayed the distribution of the various functional classes of cfRNA in lung cancer; (B) representative gene expression heatmaps showing high-, medium-, and low-expressing transcripts involved in different pathways from different cancer types.

Figure 1B illustrates a global cfRNA expression and functional landscape in lung, breast, and pancreatic cancers. The circulating cell-free transcriptome composition of TP53 signaling and MAP kinases was particularly dominant in these three cancer types. For the quantification of cfRNA expression levels, we categorized transcripts into 4 classes based on  $\Delta Ct$  values—high, medium, low, and not detected. Those genes with  $\Delta Ct$  values between 0 and 15 were classified as “high expression” (blue);  $\Delta Ct$  values between 15 and 20 were interpreted as “medium expression” (green); and  $\Delta Ct$  values of 20–30 were considered as “low expression” (red) after normalization with the control samples. The genes with  $\Delta Ct$  values > 30 were considered “not detected” and not color-coded. From the representative cfRNA expression heatmaps, one can easily identify the differentially expressed genes of specific functional clusters in particular cancer type. For example, ERCC2, MDM2, POLR2B, and PSMB10 genes involved in the DNA repair pathway are highly expressed in pancreatic cancer; whereas FANCG is a breast cancer-specific gene; and POLH and RPA2 are strongly expressed in lung cancer. Among those cell surface markers, C5AR1, CD24, CD28, and SELP genes are highly expressed in pancreatic cancer; whereas CD7, CD8A, and FAS overexpression are breast cancer-specific; and CD79 and MS4A1 are strongly expressed as lung cancer-specific genes.

Here, we have established first-ever, highly distinct cfRNA expression profiles, and functional clusters specific for lung, breast, and pancreatic cancers. As shown in the heatmaps, pancreatic cancer was characterized by the highest heterogeneity of gene expression as a wide spectrum of genes were activated by its transcriptional machinery. In contrast, lung cancer has relatively low cfRNA heterogeneity and fewer specific genes identified in the circulating cfRNA transcriptome. Multiple deregulated pathways are responsible for transducing mechanical and growth stimuli into the continuous activation of specific gene expression (i.e., always ON and never turn OFF) during cancer development and association with drug susceptibility. In this groundbreaking work, we have established an unprecedented functional cfRNA database which will guide (i) the illustration of a comprehensive landscape of cfRNA in circulation, (ii) the classification of cfRNA species by their functions, (iii) the identification of differentially expressed cfRNA in a particular cancer type, and (iv) the establishment of specific cfRNA expression signatures for different cancer types. The cfRNA expression profiles identified in this study represented the functional genomic fingerprints in circulation for specific cancer types, thus offering the exciting opportunity of personalized drug efficacy prediction.

### 3.2. Selection and Validation of PGA Lung cfRNA Biomarkers

Genomic features are frequently regarded as the state-of-the-art markers for drug response prediction. Numerous studies, on the other hand, have shown that the measurement of gene expression is a potent and still under-utilized method for identifying cell vulnerabilities, with superior performance over genomic features in genetic and compound response prediction [28–30]. The advantage of expression-based profiles over DNA-based alterations held consistently across multiple experimental platforms, models, and databases [31].

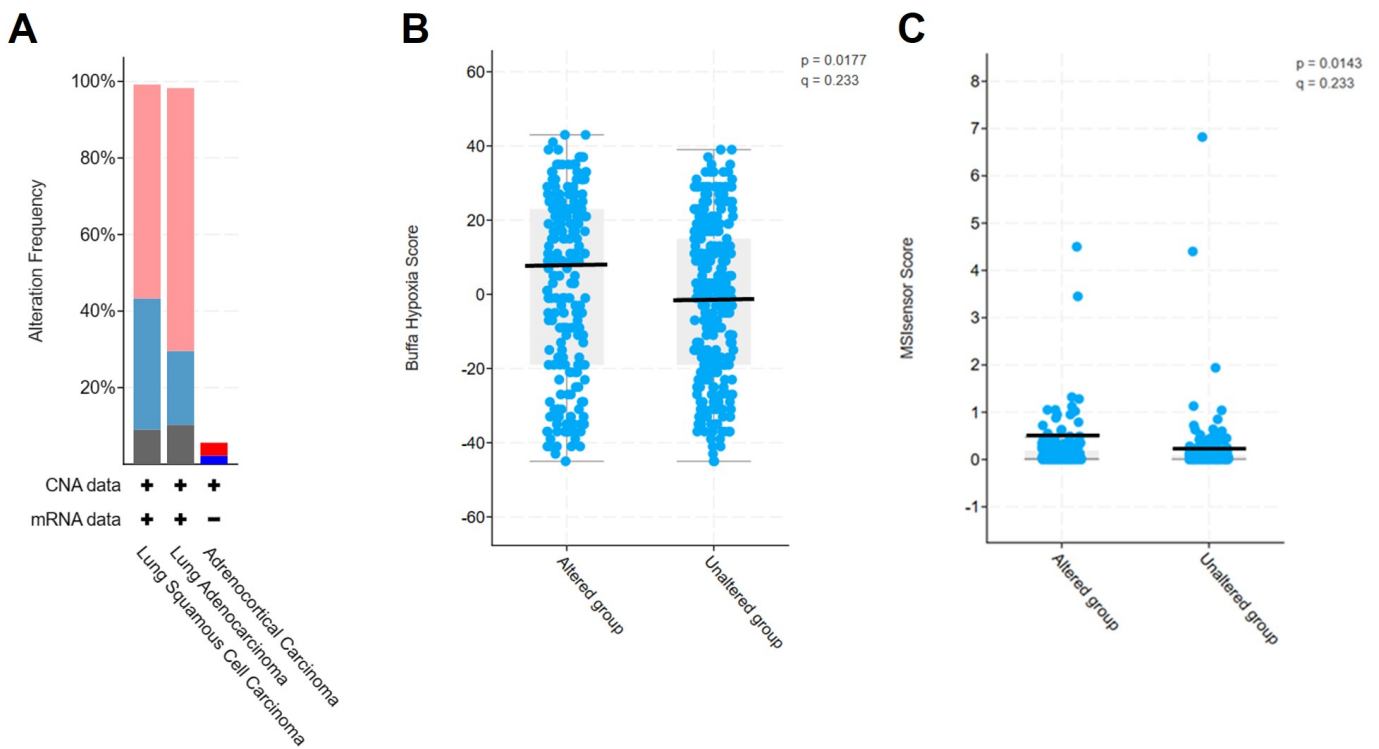
Most importantly, contrary to the common perception in the literature, the most accurate expression-based models depended on only a few features, suggesting that a full RNA-Seq profile of tumors is not necessary to gain powerful prediction for precision therapy [31]. Since many cell vulnerabilities can be identified with just one or two expression features, cost- and time-efficient technologies such as RT-qPCR with identified biomarkers could offer unmatched benefits. Specifically, genes exhibiting bimodal expression and covering important cancer-associated pathways can be used to robustly predict drug response across datasets [32]. These bimodal predictive biomarkers have a high potential for clinical translatability given the clear separation they would provide between patient responder and non-responder cohorts, and the practicality of measuring a few genes for treatment planning using various targeted assays instead of whole-transcriptome sequencing.

Consistently, another line of evidence, essentially by guided trial and error, has demonstrated that it is possible to “reprogram” cell type by manipulating only a handful of genes [33]. It has been delicately proven that the forced expression of only four genes SOX2, OCT4, KLF4, and MYC was able to turn adult cells back into pluripotent or embryonic-like stem cells [34]. Overall, it was estimated that 10–200 meta-analytic genes are required to provide optimal downstream performance and make available replicable marker lists for the 85 BICCN cell types [33]. Even modern precision medicine supports this notion that a single hotspot mutation in a single gene, e.g., EGFR, KRAS, BRAF, ABL1, and JAK2, is sufficient to predict effective targeted therapy.

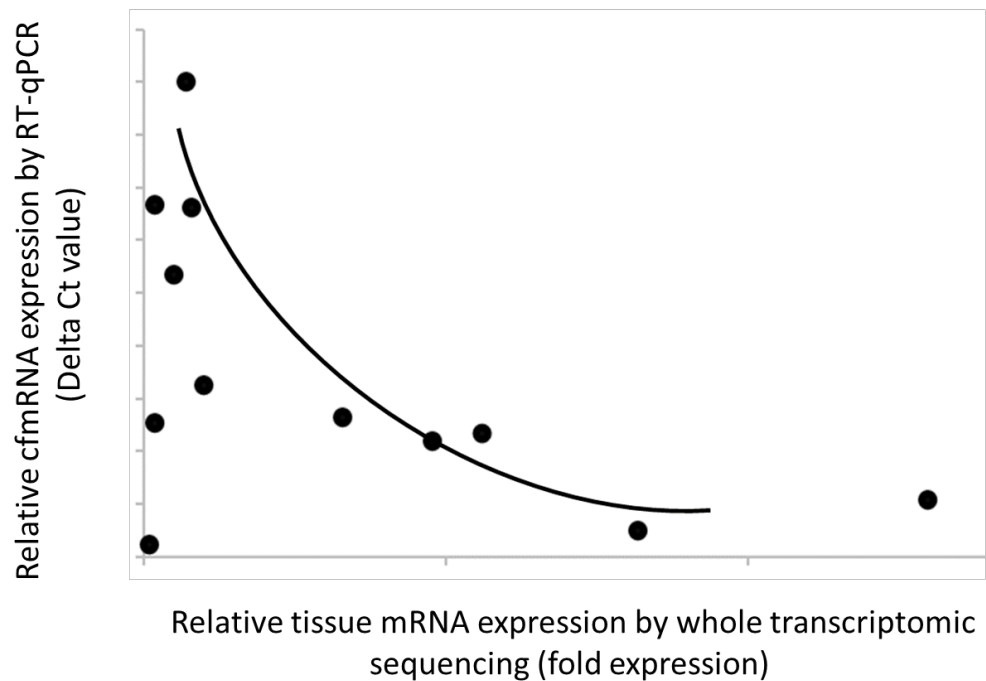
We thus set up to select dozens of lung cancer-specific cfRNA biomarkers based on four criteria: (i) tumor-specific, highly expressed, and readily detectable biomarkers; (ii) biomarkers involved in nine major cancer functional clusters, directly affecting more than 10,000 genes; (iii) biomarkers that were retrospectively verified in tumor tissues as overactive; and (iv) biomarkers associated with drug efficacy, e.g., cell death, proliferation, survival, hypoxia, and microsatellite instability (MSI). The selected biomarkers were next interrogated through the TCGA database to assess their expression in patient tumors. As expected, these PGA Lung biomarkers were found to be overexpressed in 60–70% of the 1145 lung cancer patient tumors (Figure 2A). Further, the overexpression of these selected biomarkers significantly correlated with hypoxia ( $p = 0.0177$ , Figure 2B) and MSI scores ( $p = 0.0143$ , Figure 2C) in the TCGA PanCancer Database of 510 LUAD samples.

In parallel, the transcriptome-wide characterization of lung tumor tissues was conducted using RNA-Seq technology. Of 17,780 detected and annotated transcripts, 5185 (29%) displayed at least 1.2-fold higher expression over noncancer samples. Within those high-confidence, top-performance overactive genes, we have identified lung cancer-specific transcripts that are recurrently detected in both plasma and tissue. These selected PGA Lung biomarkers met our set criteria: (i) they were undetectable in noncancer or other cancer plasma, (ii) they were overexpressed in the cancer group compared to the noncancer group, and (iii) they were detected in more than one cancer sample in the lung cancer cohort. Our results demonstrated a strong correlation between the cfRNA levels in plasma and mRNA expression in tissue, suggesting that these biomarkers with relatively high expression in tumor tissue could enhance cancer detection by the non-invasive liquid biopsy technology (Figure 3). Overall, our data validated the clinical relevance of the selected PGA Lung biomarkers for informed drug efficacy.





**Figure 2.** Validation of the selected PGA Lung biomarkers for drug efficacy prediction. (A) The overexpression of PGA Lung biomarkers in most lung tumor tissues from the TCGA database (1145 samples). Significant association of PGA Lung biomarkers with hypoxia (B) and MSI scores (C) in the TCGA PanCancer database (510 LUAD samples).

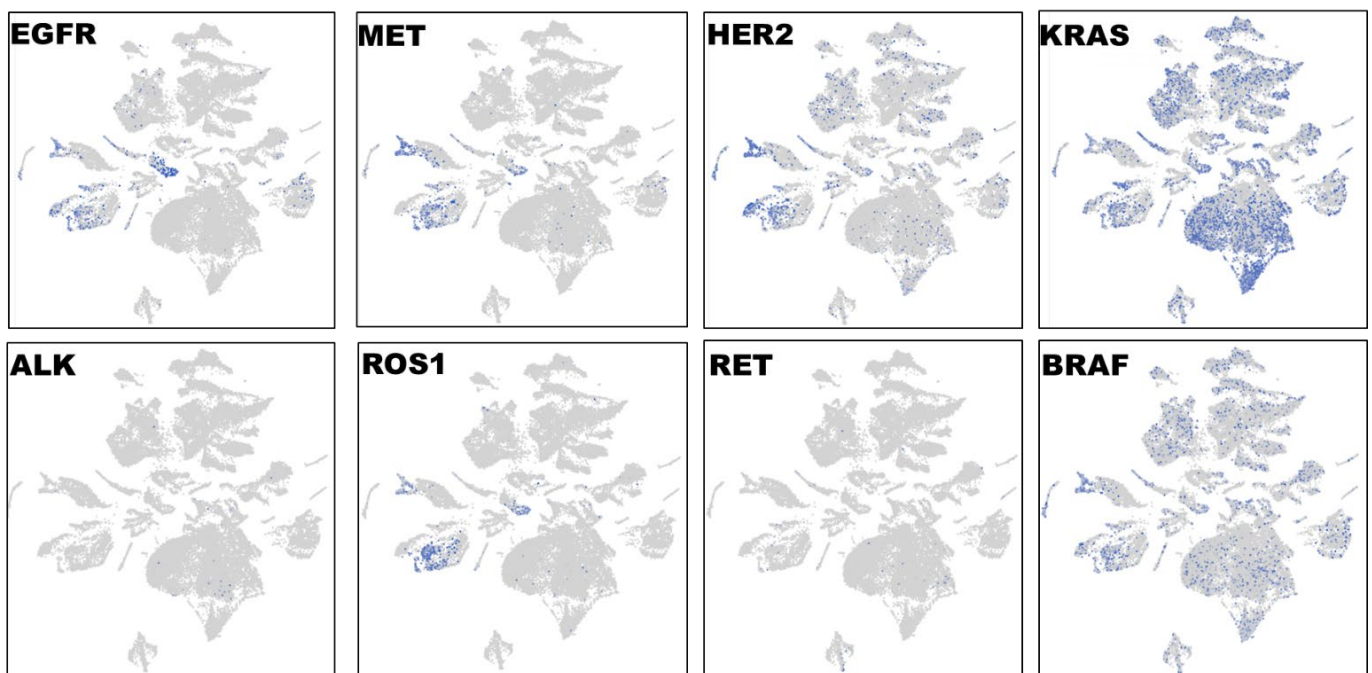


**Figure 3.** Strong correlation of the PGA Lung biomarker expression levels between the plasma and tissue samples. The relative cfRNA levels were expressed as delta Ct values, whereas the tissue mRNA expression was normalized as fold expression. The data showed a positive correlation between the cfRNA and tissue mRNA expression (i.e., an inverse relationship between delta Ct and fold expression).

### 3.3. Single-Cell Spatial Transcriptomics Analyses

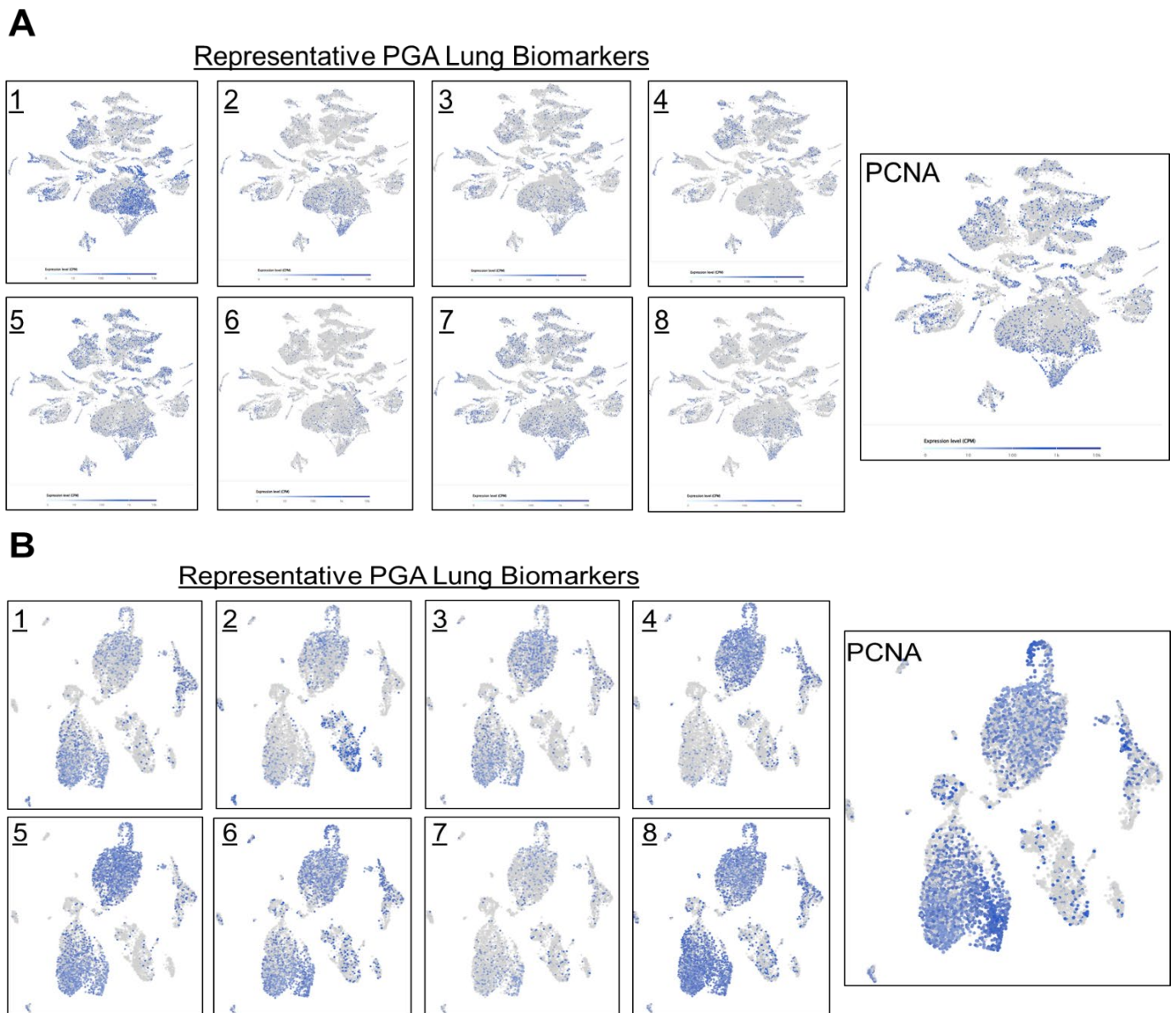
Unraveling quantitative information on gene expression changes in situ can provide valuable insights into genetic interaction, sub-population classification, cell lineage evolution, and the tumor microenvironment. Spatial transcriptomics, an emerging technique that utilizes spatially barcoded, complementary DNA probes for full-transcriptome capture on tissue sections can be added to RNA-seq data to transform our understanding of tissue functional organization and extracellular and intracellular interactions in situ. The analysis of single-cell RNA expression in their spatial context provides critical insight into tumors, immune cells, and their microenvironment. This also helps to decipher the subcellular co-localization and coexpression of target RNA biomarkers, leading to an unprecedented resolution for drug efficacy prediction.

To characterize the phenotypic and functional interaction of tumor cells and their microenvironment in lung cancer, we first performed single-cell RNA-Seq with spatial transcriptomics followed by graph-based clustering analyses to distinguish EGFR-expressing tumor cells in three lung carcinomas in a total of 32,341 cells (Figure 4). Distinct from other tumor clusters, EGFR-expressing tumor cells only made up a fraction (less than 30%) of the entire tumor population. Interestingly, EGFR+ staining was highly overlapping with MET, HER2, and ROS1 expression, whereas very few ALK- and RET-expressing cells were detected, and distinct from the EGFR+ cells. The high expression level of KRAS was detected in over 50% of the tumor population, and its distribution was more consistent with BRAF expression. The single-cell spatial analyses have distinguished diverse cell types in lung cancer: EGFR-/MET-/HER2-/ROS1-expressing, KRAS-/BRAF-positive, and ALK+ or RET+ cells. Our novel observations here were somehow surprising in terms of the high and complex heterogeneity at the single-cell level, and will provide potential guidance on target-tailored therapy strategy in lung cancer.



**Figure 4.** Single-cell RNA-Seq spatial transcriptomic analysis in lung carcinoma tissues (32,341 cells). The visualization of tumor cells expressing the key lung cancer driver genes EGFR, KRAS, BRAF, MET, HER2, ALK, ROS1, or RET. The expression patterns of EGFR, MET, HER2, and ROS1 were highly overlapped, and these EGFR-/MET-/HER2-/ROS1-coexpressing cells only constituted a small fraction of the entire tumor population. By contrast, the expression profiles of KRAS and BRAF were similar and distributed across the entire section.

Next, we took a closer look at the expression distribution of the selected PGA Lung biomarkers to infer their roles in regulating drug responses. Most of the selected PGA Lung biomarkers were similar in expression patterns, resembled KRAS/BRAF, and they may constitute a relatively homogeneous population (Figure 5A). We also discovered a concurrent expression of PCNA in this population, indicative of highly proliferative activities. PCNA is well documented as an important prognostic predictor of cancer. Its expression has been found to be significantly elevated in various malignant tumors. PCNA expression thus can reflect cell dynamics and represent the proliferative potentials of cells, and can be used as a marker for chemotherapy efficacy [35].



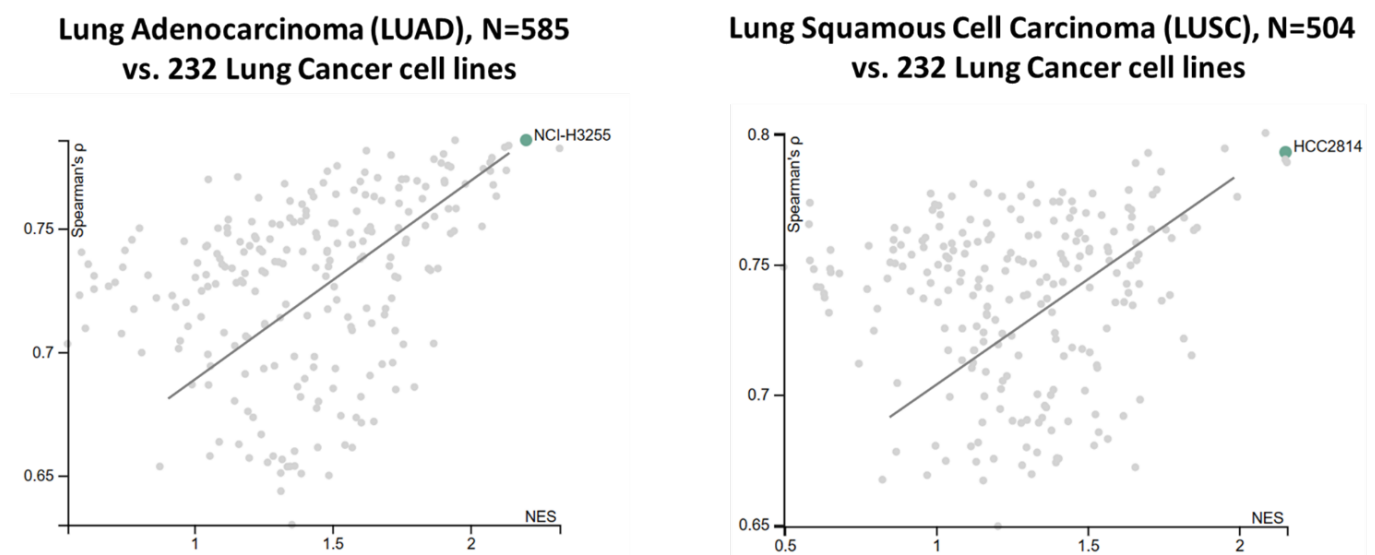
**Figure 5.** Single-cell RNA-Seq spatial transcriptomic analysis of PGA Lung biomarkers in (A) lung carcinoma tissues (32,341 cells) and (B) dissociated tumor cells from the pleural effusion of lung adenocarcinoma patients (7511 cells). The visualization of tumor cells expressing the representative PGA Lung biomarkers 1–8. The expression patterns of these PGA Lung genes were highly similar and distributed across the entire section, resembling those of KRAS and BRAF. Most significantly, the population of tumor cells expressing PGA Lung biomarkers was found to be PCNA-positive, indicative of high proliferation potential.

The immune cells that existed in other clusters were also examined by the following markers: CD4, CD8, CD25, CD69, CD19, CD20, PD-1, and CTLA-4. However, these immune cells were not in close proximity to EGFR-/MET-/HER2-/ROS1-expressing cells suggesting that they might not be the infiltrated immune cells (Supplementary Data; Figure S1). It will be of great interest to assess markers closely related to the pro-invasive or immunosuppressive tumor microenvironment to predict immunotherapy response.

To confirm what we have observed in lung carcinoma tissues, we set out to conduct the single-cell spatial transcriptomics of tumor cells obtained from the pleural effusion of lung adenocarcinoma patients with a total of 7511 cells (Figure 5B). As expected and consistent with the tumor tissue results, dissociated tumor cells in the pleural effusion showed similar expression distribution among the selected PGA Lung biomarkers, and most also coexpressed PCNA. Together, we have demonstrated the coexpression of the selected PGA Lung biomarkers with PCNA in two different sample types from different lung cancer patients.

### 3.4. From Patient's Gene Expression Signature to Drug Efficacy Prediction

Cancer cell lines with pharmacological, genomic, and transcriptomic characteristics are the most important resource available today for drug response studies. These datasets can be pooled, analyzed, and trained from the Cancer Cell Line Encyclopedia (CCLE), The Genomics of Drug Sensitivity in Cancer (GDSCv1/2), and The Cancer Therapeutics Response Portal (CTRPv2). A total of 232 lung cancer cell lines were digitally mapped and analyzed for their correlation with the corresponding TCGA lung cancer cohorts (a total of 1089 patient tumors). Since most of the patient tumors also harbor immune and other normal cells, the TCGA samples with a tumor content lower than 70% were excluded from the analysis. The correlation between the cell lines and the corresponding TCGA cohort was determined using Spearman's correlation coefficient ( $\rho$ ) and the normalized enrichment score (NES). A positive value indicated high consistency between a cell line and a disease-matched TCGA cohort. Overall, we found strong genetic similarity between lung cancer cell lines and lung cancer patient tumors (Figure 6). These cell lines faithfully recapitulate gene expression profiles and major cancer pathway activities in tumors, many of these associated with drug sensitivity/resistance.



**Figure 6.** Strong functional genomics similarity between the TCGA lung tumors and lung cancer cell lines. The Spearman correlation and normalized enrichment score (NES) were derived from the expression patterns of overactive genes and the activities of the cancer-related pathways.

We next took advantage of the high-degree representation of tumor functional activities in cancer cell lines for the pharmacogenomic prediction of drug sensitivity. Publicly available gene expression datasets for a large cohort of cell lines (CCLE), single cells (EMBL), and primary tumors (TCGA) were retrospectively pooled and merged to identify clinically relevant features called cancer consensus modules (CCMs). Prospectively collected cancer type-specific, patient-derived gene expression signatures were then used to align, filter, homogenize, and map with CCM. The resultant datasets were applied to predict *in vivo* drug efficacies (Figure 7). We have identified significant gene-drug interactions for the majority of 700+ anticancer drugs (approved, investigational, or clinical trial) via PGA. A pathway-centric approach highlighted the power of drug efficacy prediction by those PGA Lung biomarkers involved in cancer pathways. For example, MEK and PARP inhibitors have been identified by the PGA test to be effective for a number of refractory or recurrent lung cancer patients. Together, we have discovered and established a translational linkage demonstrating that lung cancer patient-derived gene expression signatures can be mapped onto molecularly annotated human cancer cell lines and correlated with sensitivity to more than 700 anticancer drugs. Our data fusion and mapping analytics ensured accurate translation from functional genotypes to cellular phenotypes and identified effective therapeutics to benefit lung cancer patients.

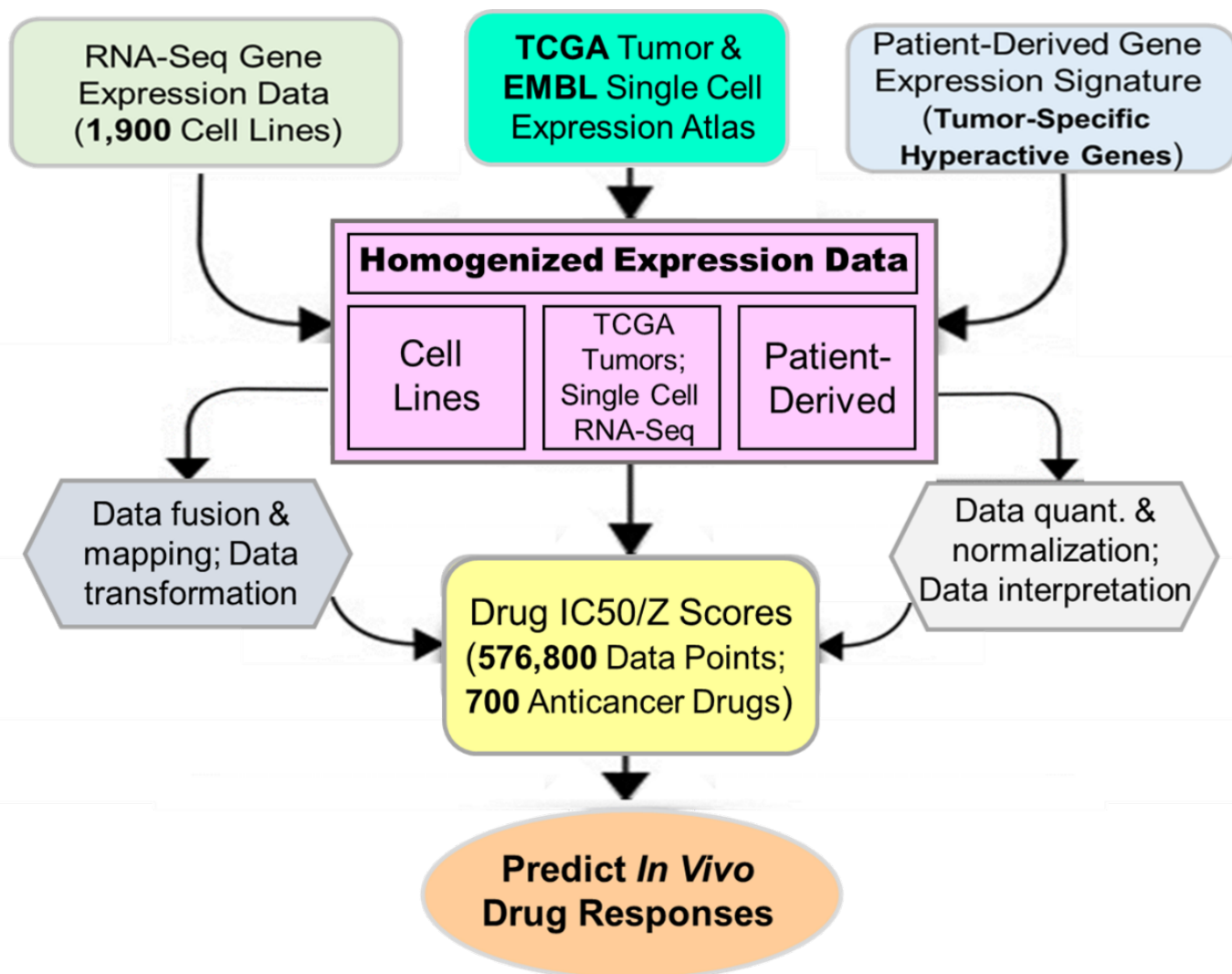
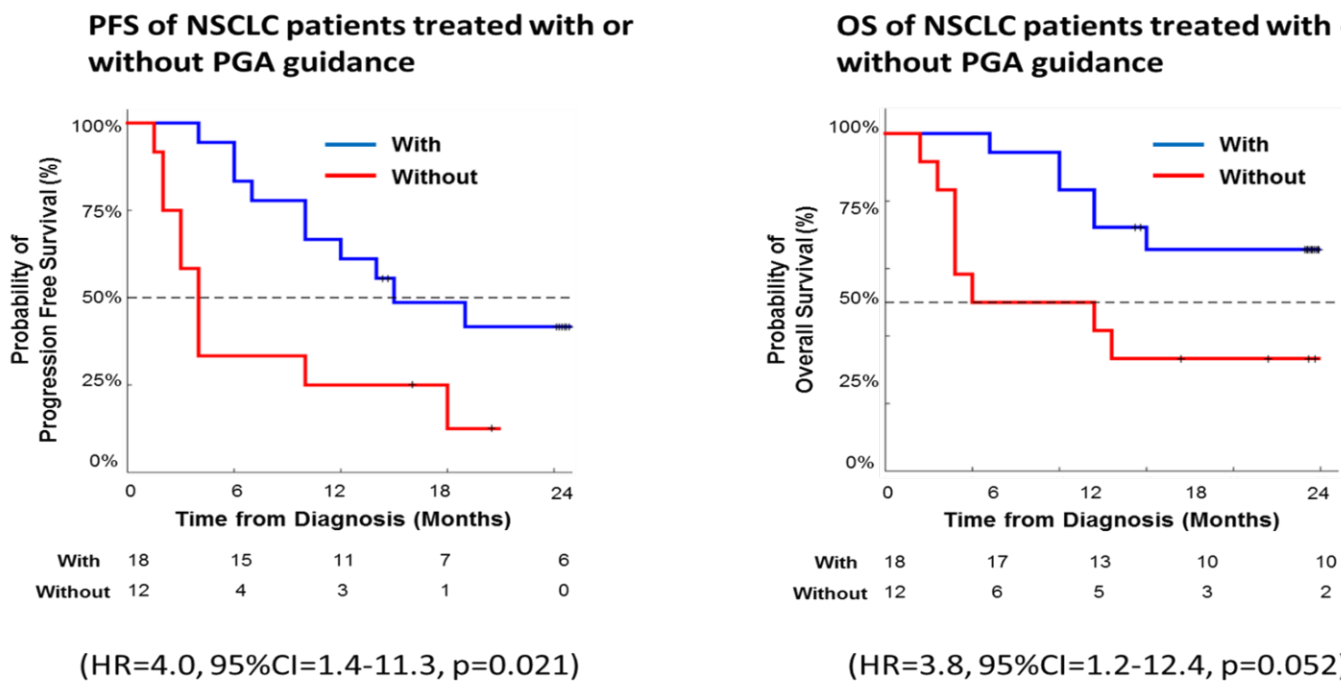


Figure 7. Overview of in silico data fusion, annotation, mapping, and analyses in the PGA Lung test.

### 3.5. Clinical Utility and Validity of the PGA Lung Test

As a proof of principle, we further evaluated PGA clinical validity on a small cohort of 30 patients with recurrent or progressive lung cancer. To ensure the cross-group comparison of the trial, we divided patients into two groups, each with the indicated numbers of age-, gender- and stage-matched subjects. In the placebo group of 12 patients, clinicians treated these patients according to current medical guidelines without the PGA test, while in the experimental group of 18 patients, patients went through the PGA test and clinicians treated these patients with PGA’s drug efficacy information. Tumor response following treatment was evaluated by a standard-of-care computed tomography scan based on the response evaluation criteria in solid tumors (RECIST). The Kaplan–Meier method and a log-rank test were used to analyze the univariate discrimination of progression-free survival (PFS) and overall survival (OS) with demographic, baseline clinical information and toxicity data. The Kaplan–Meier curve is a non-parametric statistic used to estimate the survival function from lifetime data. In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment. In clinical trials or community trials, a plot of the Kaplan–Meier estimator is a series of declining horizontal steps which, with a large enough sample size, approaches the true survival function for that population. The log-rank test is used to test whether the difference between survival times between two groups is statistically different or not. It is widely used in clinical trials to establish the efficacy of a new treatment in comparison with a control treatment when the measurement is the time to event.

In our pilot trial, the Kaplan–Meier survival analysis revealed significantly longer PFS and OS among the PGA-guided patients compared with the patients without PGA support (PFS: hazard ratio, 4.0; 95% CI, 1.4–11.3;  $p = 0.021$ ; OS: hazard ratio, 3.8; 95% CI, 1.2–12.4;  $p = 0.052$ ) (Figure 8). Thus, the real-world data here demonstrated PGA’s clinical utility and validity with a significant effect on long-term survival in our cohort of lung cancer patients.



**Figure 8.** Kaplan–Meier analysis of progression-free survival (PFS) and overall survival (OS) for the treatment of real-world lung cancer patients with or without the support from the PGA Lung test.

#### 4. Discussion

The hallmark of precision medicine is the ability to obtain early genetic evidence of whether the medicine is working and use that to inform clinical decisions. Unfortunately, all too often, the actual benefits of precision therapy to patients are short-lived because tumors are heterogeneous and drug resistance emerges quickly. The harsh reality is that only 20–30% of cancer patients are eligible, and in the qualified population, about one-fourth actually respond to targeted treatment. As a result, a small fraction of cancer patients (5–10%) experience a clinical benefit from treatments matched to tumor DNA mutations (via biomarker testing). Finding reliable and interpretable biomarkers that can predict non-responder patients' response to anticancer drugs thus remains a huge unmet clinical need. In this study, we have invented and employed cutting-edge functional genomics to translate a patient's genetic profile into a drug response to benefit more patients. The PGA classifier was designed to categorize patient's cfRNA expression data into "responder" or "not responder"—in reality, PGA does not provide a binary answer but instead generates drug efficacy prediction (or drug response prediction) to more than 700 anticancer drugs.

Gene expression profiling is an innovative functional genomics for identifying tumor vulnerabilities, with superior performance over genomic features in both genetic and drug response prediction. Studies have shown the advantage of expression-based features over DNA-based alterations held consistently across multiple experimental platforms using different perturbation technologies. It has been suggested that the expression of gene panels, such as pathway clusters or transcription factor classes, are more robust and reliable predictors than the expression of individual genes [36,37]. Moreover, it was able to "reprogram" cell type by manipulating only a handful of genes, and it was estimated that 10-200 genes are sufficient to robustly determine a cell's type. Based on these findings, we have conducted plasma cfRNA profiling to identify cancer type-specific, patient-unique gene expression signatures for drug efficacy prediction. We have selected dozens of tumor-overexpressed biomarkers involved in nine cancer pathways, i.e., immune response, cell surface markers, DNA repair, DNA methylation, oncogenesis, tumor metastasis, transcription factors, TP53 signaling, and MAPK pathways, to be broadly representative and ensured the capture of tumor and non-tumor signals. These selected PGA Lung biomarkers are capable of directly affecting more than 10,000 genes, and their over-activation has been retrospectively verified in tumor tissues and TCGA cohorts. Most significantly, these biomarkers were implicated in drug response, e.g., cell growth, survival, death, hypoxia, and microsatellite instability (MSI).

We further profiled the cell subtypes expressing PGA Lung biomarkers and their spatial distribution in lung tumor tissues as well as dissociated tumor cells from the pleural effusion by single-cell RNA-Seq and spatial transcriptome. As a result, we created an atlas of PGA Lung biomarker-expressing cells in lung cancer. We defined these cell clusters using representative PGA Lung biomarkers, key lung cancer driver genes, and immune cell markers, and identified their spatial distribution. We found that the EGFR-/MET-/HER2-/ROS1-expressing tumor cells constituted only a small fraction of the tumor population, while the KRAS-/BRAF-positive cell clusters were distributed over the entire tumor section. Most cells expressing PGA Lung biomarkers were also KRAS+/BRAF+, suggesting this relatively homogeneous population could be a more effective target for therapeutic intervention than the EGFR-/MET-/HER2-/ROS1-positive cells. Interestingly, immune T- and B-cells were found to be in distinct cell clusters and distant from the EGFR-/MET-/HER2-/ROS1-expressing tumor cells. The PGA Lung biomarker-expressing cells were also enriched with PCNA, indicative of high proliferation potential. The spatial patterns were reproducible in tumor cells from the pleural effusion. Therefore, our data identified, for the first time, the PGA Lung-/KRAS-/PCNA-coexpressing cells as the dominant and representative subtype in lung tumors which will serve as an important cell atlas in illustrating the complex transcriptomics and potential therapeutic targets for lung cancer. Moreover, treatment strategies targeting EGFR/MET/HER2/ROS1 may not be sufficient.

Tumor cell subtypes, immune cell proximity, and gene expression in individual cell types in lung cancer could partly explain the failure of targeted therapy and immunotherapy. The single-cell spatial transcriptome also revealed that the relatively homogeneous coexpression of PGA Lung biomarkers in the same population, instead of heterogeneous expression in different cell clusters, would make PGA Lung assay more accurate and consistent for drug efficacy prediction. Overall, the spatial atlas of the transcriptional profiles of PGA Lung biomarkers in the tumor cell subtypes further validated their predictive power for drug efficacy.

To date, multiple cellular and molecular changes in lung cancer, including those mutations in the driver genes and interaction between tumor and immune cells, are thought to contribute to the pathological state. However, it is a daunting and risky task to extrapolate pharmacogenomics datasets from cells directly to humans. A number of computational drug response predictions are trained and learned on pre-clinical datasets and subsequently “humanized” to bridge pre-clinical models and human tumors. Most approaches applied molecular profiles and drug screens from the large-scale databases of pre-clinical models with advanced machine learning and training, e.g., transfer learning or deep neural network learning, to correct for differences between pre-clinical models and human tumors [10,12,15]. Although promising, these digital approaches either do not take into account the real-time, real-world patient data and dynamic tumor evolution or only model these differences as a technical batch effect, leading to “one-size-fits-all” generalized software packages. To reach accuracies that are acceptable for clinical applications, existing databases and technologies just cannot provide the right source code for cell-to-tumor translation. In this study, we have correlated gene overexpression patterns and pathway activities of more than 200 lung cancer cell lines with the corresponding TCGA tumor cohorts. Our results of cell-tumor comparisons demonstrated substantial similarities in the gene-pathway functional profiles across pre-clinical and clinical barriers. Our work established the first-ever molecular algorithm for data fusion, translation, and extrapolation combining in vitro patient testing and in silico analytics, providing a quantum leap for drug efficacy prediction in lung cancer. In the long term, PGA technology could serve as a powerful tool to advance our understanding of the molecular mechanisms in cancer that mediate vulnerability or drug sensitivity.

Our analysis of >1000 patient tumor samples, ~40,000 single cells, and the subsequent superimposing of consensus genomic features onto cell lines exemplifies how gene expression signatures can be used to reliably predict drug efficacy at the individual patient level, and maximizes the clinical utility of the PGA Lung test reported. The majority of cancer consensus modules (CCMs) identified from the TCGA tumors and single-cell transcriptomics are captured within a large number of lung cancer cell lines and often at a similar extent to those observed in patient cohorts. Pharmacological datasets in cancer cell lines also offer an unbiased and plug-and-play resource for potential leverage on drug efficacy.

We introduced the PGA Lung test to integrate pre-clinical and clinical data in a semi-supervised way. Our approach functionally aligned cell-to-tumor similarity matrices and extracted relevant CCM for mapping drug efficacy. By performing a functional gene-pathway alignment instead of a direct database comparison, CCM limited the effect of sample selection bias and filtered out variables. Although we restricted ourselves to dozens of PGA Lung biomarkers, deploying CCM that incorporate patient-derived gene expression signatures specifically tailored for personalized drug efficacy prediction is a potentially revolutionary avenue. The identified and defined CCM was present in real-world patients at a frequency that would make PGA Lung testing in a clinical setting feasible. We have found that more than 90% of the primary tumor samples harbor at least one CCM associated with increased drug response. Hence, prioritizing molecular diagnostics that deliver real-time gene expression profiles could be the most cost- and time-effective means to stratify patients for cancer treatment.



Today, the vast majority of cancer patients have no detectable biomarkers for precision medicine. Therefore, expanding our arsenal of accurate theranostics would pave the way for personalized medicine by identifying the most effective drug for each patient. The PGA Lung test was able to predict drug efficacies for patients, either as monotherapy or combination therapy. We convincingly demonstrated that its performance was substantially better than an educated guess for a number of therapies of high clinical importance, such as platinum-based chemotherapies, gemcitabine, and paclitaxel. PGA Lung assay is versatile, generalizable, scalable, and can be implemented to provide guidance in alternative treatment options (e.g., drug repurposing) for patients with refractory or relapsed disease or when the standard-of-care treatments are exhausted.

PGA Lung technology still has room for improvement. First, a few CCMs are not well represented by a single cell line or not at all, and coverage by individual patients is variable. As we are in an era of precision oncology, where many drugs are active in small molecularly defined subgroups of patients, the broadness of CCM for different tumor genotypes can be further improved. As the pre-clinical and clinical databases keep expanding, they will make CCM encompassing the molecular diversity of cancer a realistic possibility. Second, our ability to validate some pharmacogenomic associations was restricted by the limited number of overlapping cell lines and drugs between these studies. The consistency between datasets is not perfect, and efforts toward standardization to reduce methodological and biological differences across the different studies are likely to improve future CCM representation between datasets. Third, we focused on cfRNA expression. The integration of other genomic features—for example, mutations, copy number, methylation, and chromatin accessibility—may help refine drug efficacy prediction by providing additional signals. Finally, we do assume the functional clustering from CCM follows the same biological dogma in pre-clinical models and human tumors. This assumption, albeit reasonable, might be debatable.

Real-life patient-derived gene expression profiling opens new paths to understanding how cancer drugs can be better matched to patients. The breakthrough PGA technology enables us to analyze each patient's molecular portrait to better match them with tailored treatments. PGA Lung test also sheds light on the complex relationships between gene activity within tumors and how different treatments will affect them.

## 5. Conclusions

PGA-based drug efficacy predictions, for the first time, revealed a clinically strong relationship between drugs and gene pathways in the context of treatment response. This groundbreaking technology connected a systematic drug efficacy prediction pipeline with layered *in vitro* and *in silico* analyses involving plasma cfRNA profiling, cancer type-specific biomarkers, individualized gene expression signatures, and anticancer drug database, which are the most important prerequisite for the clinical implementation of the PGA Lung platform.

Owing to the explicit use of cfRNA biomarkers, PGA Lung highlights the underpinning biological mechanisms contributing to drug efficacy. The plasma gene expression-based prediction approach allowed us to capture novel signals from a non-tumor environment, immune cell communication, and interaction in real-time. This can enable drug efficacy prediction at cellular resolution from both tumor and non-tumor tissues, thus providing a high degree of specificity much more so than using tumor DNA sequencing data alone.

The number of tumor mutations can sometimes help doctors identify the patients most likely to benefit from targeted therapy but unfortunately, most cancer patients (70–80%) carry no actionable mutations and do not respond to targeted therapy. Even in those responders, drug resistance will inevitably develop. Treatment options for progressive disease continue to dwindle as mortality rates are rising. The one-of-a-kind PGA Lung technology is able to nominate existing drugs for further consideration to meet the unmet

demands of enabling personalized treatments for “non-responder” patients based on tumor molecular profiles, thereby fulfilling the precision medicine promise.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/onco4030012/s1>, Figure S1: Distinct immune cell clusters in tumor microenvironment by single cell RNA-Seq spatial transcriptomic analysis in lung carcinoma tissues (32,341 cells).

**Author Contributions:** Conceptualization, C.Y.; data curation, S.-T.L. and H.-C.L.; formal analysis, C.Y.; funding acquisition, C.Y.; investigation, C.Y., S.-T.L. and H.-C.L.; methodology, C.Y.; project administration, S.-T.L.; resources, C.Y.; software, S.-T.L.; supervision, C.Y.; validation, S.-T.L. and H.-C.L.; visualization, C.Y.; writing—original draft, C.Y.; writing—review and editing, C.Y., S.-T.L. and H.-C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by OncoDxRx under the project Onco-cfmrna 00363521.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use Guideline for Good Clinical Practice (ICH-GCP), and approved by the Institutional Research Board of Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan, IRB 20180804Rv2.

**Informed Consent Statement:** Written informed consent was obtained from all the subjects involved in the study for the use and publication of data (28 August 2018 version 2). All the experiments were carried out in accordance with the ICH-GCP in its last revised version.

**Data Availability Statement:** The data presented in this study are available within the article.

**Acknowledgments:** We would like to thank Daniel Lin and Sharon Yeh for their project management, experimental, and logistic support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Denny, J.C.; Collins, F.S. Precision medicine in 2030—Seven ways to transform healthcare. *Cell* **2021**, *184*, 1415–1419. [[CrossRef](#)] [[PubMed](#)]
2. Acanda De La Rocha, A.M.; Berlow, N.E.; Fader, M.; Coats, E.R.; Saghira, C.; Espinal, P.S.; Galano, J.; Khatib, Z.; Abdella, H.; Maher, O.M.; et al. Feasibility of functional precision medicine for guiding treatment of relapsed or refractory pediatric cancers. *Nat. Med.* **2024**, *30*, 990–1000. [[CrossRef](#)]
3. Cheng, M.L.; Berger, M.F.; Hyman, D.M.; Solit, D.B. Clinical tumor sequencing for precision oncology: Time for a universal strategy. *Nat. Rev. Cancer* **2018**, *18*, 527. [[CrossRef](#)]
4. Marquart, J.; Chen, E.Y.; Prasad, V. Estimation of the percentage of us patients with cancer who benefit from genome-driven oncology. *JAMA Oncol.* **2018**, *4*, 1093–1098. [[CrossRef](#)] [[PubMed](#)]
5. Gavan, S.P.; Thompson, A.J.; Payne, K. The economic case for precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **2018**, *3*, 1–9. [[CrossRef](#)] [[PubMed](#)]
6. Mishra, A.; Verma, M. Cancer biomarkers: Are we ready for the prime time? *Cancers* **2010**, *2*, 190–208. [[CrossRef](#)] [[PubMed](#)]
7. Garnett, M.J.; Edelman, E.J.; Heidorn, S.J.; Greenman, C.D.; Dastur, A.; Lau, K.W.; Greninger, P.; Thompson, I.R.; Luo, X.; Soares, J.; et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012**, *483*, 570–575. [[CrossRef](#)] [[PubMed](#)]
8. Iorio, F.; Knijnenburg, T.A.; Vis, D.J.; Bignell, G.R.; Menden, M.P.; Schubert, M.; Aben, N.; Gonçalves, E.; Barthorpe, S.; Lightfoot, H.; et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **2016**, *166*, 740–754. [[CrossRef](#)] [[PubMed](#)]
9. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Wilson, C.J.; Lehár, J.; Kryukov, G.V.; Sonkin, D.; et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483*, 603–607. [[CrossRef](#)] [[PubMed](#)]
10. Partin, A.; Brettin, T.S.; Zhu, Y.; Narykov, O.; Clyde, A.; Overbeek, J.; Stevens, R.L. Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Front. Med.* **2023**, *10*, 1086097. [[CrossRef](#)]
11. Chen, H.; King, F.J.; Zhou, B.; Wang, Y.; Canedy, C.J.; Hayashi, J.; Zhong, Y.; Chang, M.W.; Pache, L.; Wong, J.L.; et al. Drug target prediction through deep learning functional representation of gene signatures. *Nat. Commun.* **2024**, *15*, 1853. [[CrossRef](#)] [[PubMed](#)]

12. Taj, F.; Stein, L.D. MMDRP: Drug response prediction and biomarker discovery using multi-modal deep learning. *Bioinform. Adv.* **2024**, *4*, vbae010. [[CrossRef](#)] [[PubMed](#)]
13. He, D.; Liu, Q.; Wu, Y.; Xie, L. A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening. *Nat. Mach. Intell.* **2022**, *4*, 879–892. [[CrossRef](#)] [[PubMed](#)]
14. Chawla, S.; Rockstroh, A.; Lehman, M.; Ratther, E.; Jain, A.; Anand, A.; Gupta, A.; Bhattacharya, N.; Poonia, S.; Rai, P.; et al. Gene expression based inference of cancer drug sensitivity. *Nat. Commun.* **2022**, *13*, 5680. [[CrossRef](#)] [[PubMed](#)]
15. Park, A.; Lee, Y.; Nam, S. A performance evaluation of drug response prediction models for individual drugs. *Sci. Rep.* **2023**, *13*, 11911. [[CrossRef](#)] [[PubMed](#)]
16. Tang, Y.C.; Powell, R.T.; Gottlieb, A. Molecular pathways enhance drug response prediction using transfer learning from cell lines to tumors and patient-derived xenografts. *Sci. Rep.* **2022**, *12*, 16109. [[CrossRef](#)] [[PubMed](#)]
17. Partin, A.; Brettin, T.; Evrard, Y.A.; Zhu, Y.; Yoo, H.; Xia, F.; Jiang, S.; Clyde, A.; Shukla, M.; Fonstein, M.; et al. Learning curves for drug response prediction in cancer cell lines. *BMC Bioinform.* **2021**, *22*, 252. [[CrossRef](#)] [[PubMed](#)]
18. Mourragui, S.; Loog, M.; van de Wiel, M.A.; Reinders, M.J.; Wessels, L.F. Precise: A domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics* **2019**, *35*, i510–i519. [[CrossRef](#)] [[PubMed](#)]
19. Schwartz, L.H.; Litière, S.; de Vries, E.; Ford, R.; Gwyther, S.; Mandrekar, S.; Shankar, L.; Bogaerts, J.; Chen, A.; Dancey, J.; et al. Recist 1.1—Update and clarification: From the recist committee. *Eur. J. Cancer* **2016**, *62*, 132–137. [[CrossRef](#)] [[PubMed](#)]
20. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [[CrossRef](#)] [[PubMed](#)]
21. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R.; et al. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. [[CrossRef](#)] [[PubMed](#)]
22. Su, Z.; Ho, J.W.K.; Yau, R.C.H.; Lam, Y.L.; Shek, T.W.H.; Yeung, M.C.F.; Chen, H.; Oreffo, R.O.C.; Cheah, K.S.E.; Cheung, K.S.C. A single-cell atlas of conventional central chondrosarcoma reveals the role of endoplasmic reticulum stress in malignant transformation. *Commun. Biol.* **2024**, *7*, 124. [[CrossRef](#)] [[PubMed](#)]
23. Mammoliti, A.; Smirnov, P.; Nakano, M.; Safikhani, Z.; Beri, C.; Ho, G.; Haibe-Kains, B. ORCESTRA: A platform for orchestrating and sharing high-throughput pharmacogenomics analyses. *bioRxiv* **2020**. [[CrossRef](#)]
24. Jin, H.; Zhang, C.; Zwahlen, M.; von Feilitzen, K.; Karlsson, M.; Shi, M.; Yuan, M.; Song, X.; Li, X.; Yang, H. Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nat. Commun.* **2023**, *14*, 5417. [[CrossRef](#)] [[PubMed](#)]
25. Schubert, M.; Klinger, B.; Klünemann, M.; Sieber, A.; Uhlitz, F.; Sauer, S.; Garnett, M.J.; Blüthgen, N.; Saez-Rodriguez, J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **2018**, *9*, 20. [[CrossRef](#)] [[PubMed](#)]
26. Jiang, P.; Zhang, Y.; Ru, B.; Yang, Y.; Vu, T.; Paul, R.; Mirza, A.; Altan-Bonnet, G.; Liu, L.; Ruppin, E.; et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat. Methods* **2021**, *18*, 1181–1191. [[CrossRef](#)] [[PubMed](#)]
27. Liu, Z.; Shen, Y.; Lakshminarasimha, V.B.; Liang, P.P.; Zadeh, A.; Morency, L.-P. Efficient low-rank multimodal fusion with modality-specific factors. In Proceedings of the ACL 2018—56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; Volume 1, pp. 2247–2256.
28. Ali, M.; Aittokallio, T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* **2019**, *11*, 31–39. [[CrossRef](#)]
29. Ding, Z.; Zu, S.; Gu, J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics* **2016**, *32*, 2891–2895. [[CrossRef](#)] [[PubMed](#)]
30. Costello, J.C.; Heiser, L.M.; Georgii, E.; Gönen, M.; Menden, M.P.; Wang, N.J.; Bansal, M.; Ammad-ud-din, M.; Hintsanen, P.; Khan, S.A.; et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **2014**, *32*, 1202–1212. [[CrossRef](#)] [[PubMed](#)]
31. Dempster, J.M.; Krill-Burger, J.M.; McFarland, J.M.; Warren, A.; Boehm, J.S.; Vazquez, F.; Hahn, W.C.; Golub, T.R.; Tsherniak, A. Gene expression has more power for predicting in vitro cancer cell vulnerabilities than genomics. *bioRxiv* **2020**. [[CrossRef](#)]
32. Ba-Alawi, W.; Nair, S.K.; Li, B.; Mammoliti, A.; Smirnov, P.; Mer, A.S.; Penn, L.Z.; Haibe-Kains, B. Bimodal gene expression in patients with cancer provides interpretable biomarkers for drug sensitivity. *Cancer Res.* **2022**, *82*, 2378–2387. [[CrossRef](#)] [[PubMed](#)]
33. Fischer, S.; Gillis, J. How many markers are needed to robustly determine a cell’s type? *iScience* **2021**, *24*, 103292. [[CrossRef](#)] [[PubMed](#)]
34. Takahashi, K.; Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **2006**, *126*, 663–676. [[CrossRef](#)]
35. Ye, X.; Ling, B.; Xu, H.; Li, G.; Zhao, X.; Xu, J.; Liu, J.; Liu, L. Clinical significance of high expression of proliferating cell nuclear antigen in non-small cell lung cancer. *Medicine* **2020**, *99*, e19755. [[CrossRef](#)] [[PubMed](#)]

- 
36. Wang, X.; Sun, Z.; Zimmermann, M.T.; Bugrim, A.; Kocher, J.-P. Predict drug sensitivity of cancer cells with pathway activity inference. *BMC Med. Genom.* **2019**, *12*, 15. [[CrossRef](#)] [[PubMed](#)]
  37. Rydenfelt, M.; Wongchenko, M.; Klinger, B.; Yan, Y.; Blüthgen, N. The cancer cell proteome and transcriptome predicts sensitivity to targeted and cytotoxic drugs. *Life Sci. Alliance* **2019**, *2*, e201900445. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.