



Article

RNA Sequencing in Comparison to Immunohistochemistry for Measuring Cancer Biomarkers in Breast Cancer and Lung Cancer Specimens

Maxim Sorokin ^{1,2,3}, Kirill Ignatev ⁴, Elena Poddubskaya ^{1,5}, Uliana Vladimirova ³, Nurshat Gaifullin ⁶, Dmitriy Lantsov ⁷, Andrew Garazha ², Daria Allina ¹, Maria Suntsova ¹, Victoria Barbara ⁸ and Anton Buzdin ^{1,2,3,9,*}

¹ Institute of Personalized Medicine, I.M. Sechenov First Moscow State Medical University, 119048 Moscow, Russia; sorokin@oncobox.com (M.S.); podd-elena@ya.ru (E.P.); allina_dasha@mail.ru (D.A.); suntsova86@mail.ru (M.S.)

² Omicsway Corp., Walnut, CA 91789, USA; garazha@oncobox.com

³ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 117997 Moscow, Russia; uliana.bagina@gmail.com

⁴ Karelia Republic Oncological Hospital, 185000 Petrozavodsk, Russia; kirillignatev@bk.ru

⁵ Vitamed Oncological Clinical Center, 121309 Moscow, Russia

⁶ Faculty of Fundamental Medicine, Lomonosov Moscow State University, 119991 Moscow, Russia; gaifulin@rambler.ru

⁷ Kaluga Regional Oncological Hospital, 248007 Kaluga, Russia; lantsov@mail.ru

⁸ Oncological Dispensary of the Republic of Karelia, 185002 Petrozavodsk, Russia; evika9@rambler.ru

⁹ Moscow Institute of Physics and Technology, 141701 Moscow, Russia

* Correspondence: buzdin@oncobox.com; Tel.: +1-626-765-7785

Received: 29 March 2020; Accepted: 7 May 2020; Published: 9 May 2020



Abstract: RNA sequencing is considered the gold standard for high-throughput profiling of gene expression at the transcriptional level. Its increasing importance in cancer research and molecular diagnostics is reflected in the growing number of its mentions in scientific literature and clinical trial reports. However, the use of different reagents and protocols for RNA sequencing often produces incompatible results. Recently, we published the Oncobox Atlas of RNA sequencing profiles for normal human tissues obtained from healthy donors killed in road accidents. This is a database of molecular profiles obtained using uniform protocol and reagents settings that can be broadly used in biomedicine for data normalization in pathology, including cancer. Here, we publish new original 39 breast cancer (BC) and 19 lung cancer (LC) RNA sequencing profiles obtained for formalin-fixed paraffin-embedded (FFPE) tissue samples, fully compatible with the Oncobox Atlas. We performed the first correlation study of RNA sequencing and immunohistochemistry-measured expression profiles for the clinically actionable biomarker genes in FFPE cancer tissue samples. We demonstrated high (Spearman's rho 0.65–0.798) and statistically significant ($p < 0.00004$) correlations between the RNA sequencing (Oncobox protocol) and immunohistochemical measurements for *HER2/ERBB2*, *ER/ESR1* and *PGR* genes in BC, and for *PDL1* gene in LC; AUC: 0.963 for HER2, 0.921 for ESR1, 0.912 for PGR, and 0.922 for PDL1. To our knowledge, this is the first validation that total RNA sequencing of archived FFPE materials provides a reliable estimation of marker protein levels. These results show that in the future, RNA sequencing can complement immunohistochemistry for reliable measurements of the expression biomarkers in FFPE cancer samples.

Keywords: transcriptomics; RNA sequencing; immunohistochemistry; molecular diagnostics; biomarkers detection; clinical oncology; targeted therapies; personalized medicine; lung cancer; breast cancer; trastuzumab; NCT03521245; bioinformatics

1. Introduction

Both mRNA and protein levels can be used for interrogating gene expression in cancer tissues, both types of analysis having their advantages and limitations [1]. The protein level more closely reflects the cancer phenotype because these are proteins that execute major intracellular molecular functions. However, the mRNA and protein levels for known genes strongly correlate in the mammalian cells, so that mRNA levels explain ~84% of the variance in protein expression [2]. This has been also confirmed in different organisms by strong correlations between the mRNA and ribosomal footprinting or quantitative proteomics data (r range 0.59–0.89) [3–5].

Accurate, inexpensive, and reproducible high-throughput methods of quantitative proteomics are still under development [6]. However, there are many practical ways of measuring the expression of single proteins in tumor tissues, like immunohistochemistry, which has become a commonly used technique in clinical laboratory diagnostics [7]. Cancer transcriptomics provide direct analysis of RNA concentrations in tumor biosamples [8]. Transcriptomics have an advantage of being an approach unparalleled in terms of generation of high-throughput gene expression data due to the use of robust and relatively non-expensive experimental protocols applicable for the analysis of minute amounts of fresh or fixed cancer biomaterials [9]. The analysis of expression levels per single gene using such an approach is becoming a relatively cheap and easy task.

RNA sequencing is considered a gold standard approach in modern transcriptomics [10,11]. Various RNA sequencing platforms have been used for gene expression profiling in human cancers, including Illumina [12], Ion Torrent/Proton [12], and Oxford Nanopore [13]. They utilize different equipment and physical principles for detecting output signals, but also various library preparation protocols, including different enzymes and numbers of PCR cycles [14]. This diversity results in dramatic batch effects and incompatibility of the outputs obtained using different platforms, reagents, and kits [15,16], which is why experimental gene expression profiles are generally compared among those obtained using the same platform [16]. In most cancer biology applications, gene expression in the tumor is compared with the normal samples. Case-to-normal gene expression ratios can be evaluated per se [17,18]. Alternatively, pools of differentially regulated genes can be analyzed systemically and systematically, e.g., by assessing enrichments of Gene Ontology (GO) terms [19,20] or interrogating activation of molecular pathways [21–23].

However, effort should be made to compare only compatible data. Several collections of RNA sequencing profiles have been published for normal human tissues. Ideally, they should represent tissues from healthy donors screened in a single assay with the same equipment and reagents. The biggest published dataset GTEx [24] (11,688 samples), however, lacks publicly available data on the donors' exact age and requires complicated registration steps that cannot be performed by many researchers. There are also some basal contamination issues reported recently for GTEx [25]. Other relevant databases are freely accessible and include age information: The Cancer Genome Atlas, TCGA [26] (625 samples), ENCODE [27] polyA RNA-seq (41 samples), and ENCODE total RNA-seq (92 samples). Unfortunately, they also lack some of the previously mentioned features. In TCGA, the norms are adjacent to surgically removed tumors [28], but they can be not physiologically normal because of multiple pathologic effects tumors exert on the neighboring tissues, like inflammation [29], altered vascularization [30], and growth factors/cytokine balances [31]. In ENCODE, datasets were generated for the autopsy normal tissues using different library preparation methods, but they only include 1–4 samples per tissue type (including both male and female donors) and cannot form statistically significant reference groups in most of the cases. Finally, we recently published another

atlas of normal tissue expression profiles termed the Oncobox Atlas of Normal Tissue Expression (ANTE) [32]. It has statistically significant reference groups for 20 human tissues/organs, and represents 142 solid tissue samples from human healthy donors killed in road accidents and 17 blood samples from healthy volunteers. The expressions were profiled in the experiments using the same reagents and protocols.

However, very different RNA sequencing results can be obtained, depending on the source of clinical biomaterials. For fresh tissue specimens, high-integrity RNAs may be isolated, resulting in longer sequencing reads. For formalin-fixed paraffin embedded (FFPE) tissue samples, significantly degraded RNA preparations can be obtained, typically resulting in 25–50 bp single end reads [32]. While the read length depends on sequencing strategy and short reads could theoretically be obtained also from fresh-frozen tissues, storage in FFPE can potentially alter gene body coverage. This may lead to lower coverage for either 3' or 5' end of a gene [33]. Still, previous studies comparing FFPE vs. fresh-frozen samples obtained from the same tissues demonstrated lower (yet, comparable) gene body coverage for both storage techniques [34].

RNA reads are mapped to genes, while excluding ambiguous mapping entries, and the relative gene expression is then calculated. Working with degraded RNAs is problematic for the analysis of fused oncogenes because of too short reads that cannot be confidently mapped to fusion sites [35]. This is also the case for the analysis of differential alternative splice sites, because FFPE RNAseq results in lower percent of split-mapped reads when compared to RNAseq of fresh-frozen tissues [34]. However, degraded RNAs from FFPE specimens can provide high-quality expression profiles that cluster together with the samples from high-integrity RNAs of the same tissue, as shown by the ANTE project [32].

Here, we publish new original clinically and immunohistochemistry-annotated 39 breast cancer (BC) and 19 lung cancer (LC) RNA sequencing profiles, fully compatible with the Oncobox Atlas of Normal Tissues (ANTE). We performed the first correlation study of RNA sequencing and immunohistochemistry-measured expression profiles for the clinically actionable biomarker genes in FFPE cancer tissue samples. For *HER2/ERBB2*, *ER/ESR1*, and *PGR* genes in BC and for *PDL1* gene in LC, we demonstrated high and statistically significant correlations between the RNA sequencing (Oncobox protocol) and immunohistochemical measurements.

These results show that RNA sequencing, at least if the Oncobox Atlas protocol for library preparation, data mapping, and normalization is followed, in the future, can complement immunohistochemistry for reliable measurements of the expression cancer biomarkers in FFPE samples. In addition to the FFPE data, we also observed a good correlation between RNA sequencing data and immunohistochemistry for the freshly frozen BC samples from the TCGA project database [36] with known *HER2*, *ER*, and *PGR* statuses.

2. Materials and Methods

2.1. BC Biosamples

All experimental biosamples of tumor tissues were formalin-fixed and embedded into paraffin blocks (FFPE). All biosamples were evaluated by a pathologist to confirm the tumor tissue origin and only the specimens with the content of tumor cells greater than 50% were investigated further. Of them, 16 breast cancer (BC) tissue samples were obtained from the Karelia Republic Oncological Hospital, Petrozavodsk, Russia, and 23 samples from Vitamed Oncological Clinical Center, Moscow, Russia. There were 30 primary tumors, 3 lymph node metastases, 2 scar metastases, 2 liver metastases, 1 brain metastasis, and 1 ovary metastasis. All the BC patients were women and the mean age was 51.9 years old (range 27–78 y.o.). Clinical annotation of the BC biosamples investigated is summarized in Table 1.

Table 1. Clinical and molecular annotation of the breast cancer biosamples.

Sample ID	Primary Tumor or Metastasis	Age	Stage	HER2 Score	ER Score	PR Score	Coverage (mln Mapped Reads)	RIN
BC-1	primary	39	T2N3aM0, IIC	3	0	0	9.42	2.1
BC-10	primary	48	T2N0M0, II	3	0	0	6.70	1
BC-12	primary	60	T2N0M0, IIA	3	0	0	5.12	1
BC-13	primary	69	T2N3M0, III C	3	8	4	9.03	1
BC-14	primary	49	T2N2M0, IIIA	3	0	0	6.11	2.4
BC-17	primary	59	T4N2M0	3	7	2	3.96	2.5
BC-18	lymph node metastasis	47	T3N1M0, IIIA	3	0	0	6.62	2.3
BC-19	primary	48	T1N0M0, I	3	5	5	9.07	1.1
BC-20	lymph node metastasis	51	T2N0M0, II	3	0	0	10.22	2.3
BC-21	primary	49	T1N3M0, IIIC	3	0	0	9.34	2.3
BC-22	primary	47	T2N0M0, II	3	6	5	10.52	2
BC-23	primary	46	T2N2M0, IIIA	3	7	6	8.39	2.1
BC-24	primary	57	T2N0M0, IIA	3	6	4	11.21	1
BC-27	primary	44	T2N0M0	3	0	0	13.82	2.2
BC-28	ovary metastasis	53	T2N0M0, IIA	0	7	4	4.65	3.7
BC-29	primary	65	T4N3M1,IV	3	0	0	12.56	2.2
BC-3	primary	55	T2N1M0, IIIa	3	0	0	6.84	1
BC-4	primary	58	T2N1M0, IIB	3	0	0	7.17	1
BC-46	liver metastasis	27	T2N2M0	0	8	8	15.07	3.3
BC-48	relapse in the scar	36	T3N1M0	1	0	0	20.54	NA
BC-49	primary	54	T1N2M0	0	2	8	10.54	2
BC-50	primary	51	T2N0M0	0	0	0	8.49	2.6
BC-51	primary	38	T2N1M0	0	0	0	8.68	3
BC-52	primary	78	T1N2M0	1	4	8	11.92	1.7
BC-53	primary	50	T2N0M0	1	0	8	8.06	1.9
BC-54	primary	50	T2N0M0	0	0	0	7.30	1.8
BC-55	primary	71	T2N3M0	1	8	8	9.32	3.3
BC-56	primary	60	T1N1M1	0	0	8	12.66	2.4
BC-57	primary	55	T3N2M0	1	0	0	13.77	2.8
BC-58	lymph node metastasis	55	T1N0M0	0	7	7	14.24	2.1
BC-59	scar metastasis	61	T1N1M0	0	3	1	16.88	1.2
BC-60	primary	33	T2N1M0	2	0	0	10.03	1.8
BC-61	liver metastasis	38	T2N2M0	0	8	8	5.42	3
BC-62	brain metastasis	44	T2N0M0	0	0	0	10.99	3
BC-63	primary	66	T4N2M0	0	0	0	10.11	3.7
BC-64	primary	60	T3N3M0	1	0	0	12.71	3.8
BC-65	primary	42	T2N0M0	0	0	0	9.92	2.6
BC-66	primary	55	T3N1M0	3	3	3	8.96	3.1
BC-9	primary	57	T1N1M0, IIB	3	8	5	6.88	1

RIN—RNA integrity number, mln—million, NA—not assessed.

2.2. LC Biosamples

Nineteen lung cancer (LC) samples were obtained from the Vitamed Oncological Clinical Center, Moscow, Russia ($n = 6$) and from Kaluga Regional Oncological Hospital, Kaluga, Russia ($n = 13$). There were nine lung adenocarcinomas, seven squamous cell carcinomas, one adeno-squamous cell carcinoma, one small cell carcinoma, and one was unidentified. The patients were 17 men and 2 women, aged from 57 to 79 with the mean age of 67 years.

We collected information about the patients' sex, age, diagnosis, and clinical history. Informed written consents to participate in the study and to include the results in this report were obtained from all patients. The consent procedure and the design of the study were approved by the ethical committees of both the Karelia Republic Oncological Hospital, Petrozavodsk, Russia and the Vitamed Oncological Clinical Center, Moscow, Russia. Clinical annotation of the LC biosamples investigated is summarized in Table 2.

2.3. Preparation of Libraries and RNA Sequencing

To isolate RNA, 10 μ M-thick paraffin slices were trimmed from each FFPE tissue block using a microtome. Eight paraffin slices were used for RNA extraction. RNA was extracted from FFPE slices using a QIAGEN RNeasy FFPE Kit following the manufacturer's protocol. RNA 6000 Nano or Qubit RNA Assay kits were used to measure RNA concentration. RNA Integrity Number (RIN) was measured using Agilent 2100 bio-analyzer. For depletion of ribosomal RNA and library construction, KAPA RNA Hyper with an rRNA erase kit (HMR only) was used. Different adaptors were used for multiplexing samples in one sequencing run. Library concentrations and quality were measured using

the Qubit dsDNA HS Assay kit (Life Technologies) and Agilent TapeStation (Agilent). Single-end RNA sequencing, 50 bp read length, for ~30 million raw reads per sample, was performed at Omicslab LLC, Moscow and at the Department of Pathology and Laboratory Medicine, University of California Los Angeles, using the Illumina HiSeq 3000 System. A data quality check was performed using the Illumina Sequencing Analysis Viewer and de-multiplexing was performed with Illumina Bcl2fastq2 v 2.17 software. Sequencing data were deposited to NCBI Sequencing Read Archive (SRA) under accession ID PRJNA565016 and PRJNA578290.

Table 2. Clinical and molecular annotation of the lung cancer biosamples.

ID	Histology	Age	Stage	Sex	Percent of PDL1 Positive Cells	Coverage (mln Mapped Reads)	RIN
LuC_16	squamous cell carcinoma	75	T3N2M1, IV	male	1%–50%	11.54	2.4
LuC_18	squamous cell carcinoma	63	T2N1M0	male	0	15.45	3
LuC_19	squamous cell carcinoma	65	T2N0M0	male	>50%	12.57	3
LuC_30	Unidentified	79	T2NXM0	male	>50%	11.01	4.9
LuC_31	adenocarcinoma	66	T3N2M0	male	1%–50%	10.27	4.5
LuC_32	adeno-squamous cell carcinoma	70	T2N1M0	male	>50%	12.14	2.7
LuC_33	squamous cell carcinoma	57	T3N0M0	male	0	14.12	3.8
LuC_42	adenocarcinoma	67	T1N1M0	male	1%–50%	11.9	1.4
LuC_23	adenocarcinoma	60	T2N0M0	male	0	12.06	3.2
LuC_24	adenocarcinoma	67	T2N0M0	male	0	10.77	3.8
LuC_26	small cell carcinoma	65	T3N2M0, IIIa	male	1%–50%	5.71	1.1
LuC_28	adenocarcinoma	76	T2N0M0	male	0	12.37	1.8
LuC_29	squamous cell carcinoma	65	T2N0M0	male	0	16.58	2.4
LuC_34	adenocarcinoma	62	pT1bN0M0	female	0	11.82	2.3
LuC_35	squamous cell carcinoma	75	T3N0M0	male	>50%	12.28	3.2
LuC_36	adenocarcinoma	57	pT2N0M0	male	1%–50%	11.3	2.6
LuC_37	squamous cell carcinoma	68	T3N1M0	male	0	11.93	2.3
LuC_38	adenocarcinoma	68	pT2aN2M0	male	1%–50%	15.38	3.5
LuC_39	adenocarcinoma	68	pT2pNXpM1	female	0	8.58	

RIN—RNA integrity number, mln—million, NA—not assessed.

2.4. Processing of RNA Sequencing Data

RNA sequencing FASTQ files were processed with STAR aligner [37] in “GeneCounts” mode with the Ensembl human transcriptome annotation (build version GRCh38 and transcript annotation GRCh38.89). The STAR output contained expression levels for 58,233 individual genes. Ensembl gene IDs were converted to Human Gene Nomenclature Committee (HGNC, <https://www.genenames.org/>, database version from 13 July 2017) gene symbols. In total, expression level was calculated for 36,596 genes with the corresponding HGNC identifiers.

2.5. Data Clustering

Log-transformed DESeq2 [38] normalized counts were used for hierarchical clustering analysis. The analysis was performed using R “ward.D” method. The dendrogram was visualized using a custom R script.

2.6. Immunohistochemistry

Immunohistochemistry assay for BC samples for HER2, ESR1, and PGR proteins was performed at the Clinical Diagnostic Laboratory of the Oncology Center of the Republic of Karelia, Russia, using antibody kits (Roche Diagnostics, Indianapolis, IN, USA) to identify the respective statuses of the tumors. For HER2, the output statuses were: (i) baseline staining (0), (ii) “+” (1), (iii) “++” (2), and (iv) “+++” (3). The “++” and “+++” statuses were confirmed using ISH DNA Probe Cocktail assay (Roche Diagnostics, USA). For ESR1 and PGR status, 0–8 grades were used.

LC biosamples were profiled at Unim Laboratory, Moscow (<http://new.unim.su>) for PDL1 protein expression. Hematoxylin-Eosin and PD-L1(ZR3) antibody (Sigma-Aldrich, USA) staining was used to assess the tumor statuses. The following output measures were used: (i) no cell membrane staining in biosample or staining of up to 1% of cells, (ii) staining of 1%–50% of cells, (iii) staining of 50%–100% of cells.

2.7. Literature Gene Expression Data

To compare freshly frozen tissue RNA sequencing and IHC data, we extracted all BC gene expression profiles with IHC-confirmed receptor status from The Cancer Genome Atlas project (TCGA), using the R “TCGAbiolinks” package [36]. In total, we analyzed 634 samples with confirmed HER2 status, 924 samples with confirmed ESR1 status, and 922 samples with confirmed PGR status. Identifiers of samples included in the analysis are given in Supplementary Table S1.

2.8. Statistical Analysis

Statistical analysis was performed using R software. Area under the receiver-operator curve was calculated using ROCR package. For the ROC-AUC analysis of the experimental data we used threshold >2 for separating ESR1-positive and PGR-positive cases from corresponding negative groups, according to [39]; HER2 “+++” were considered as HER2-positive, according to [40]; and tumors with more than 1% of cells stained with PD-L1 were considered as PD-L1-positive, according to [41]. Spearman’s correlation coefficient was used to test the significance of the correlation. Trendlines and 95% confidence intervals were built using `stat_smooth` function of `ggplot2` package. The log rank test was used for survival analysis.

3. Results

3.1. RNA Sequencing Data

In this study, we investigated correlations between gene expression profiles established for formalin-fixed paraffin-embedded (FFPE) tissue biosamples, using RNA sequencing data and immunohistochemistry (IHC) staining. To this end, we experimentally profiled 39 breast cancer (BC) and 19 lung cancer (LC) FFPE tissue samples, using RNA sequencing; original data were deposited to NCBI sequencing read archive under accession number PRJNA565016.

We used the same protocol as for generating the Oncobox Atlas of RNA sequencing profiles of normal human tissues derived from healthy donors [32]. We found that application of the coverage threshold of 2.5 million mapped reads resulted in tissue specific clustering, whereas for the profiles with lower number of mapped reads, we observed biased clustering. In this study, we used the same sequencing and data processing and filtering protocol. All the current 39 breast cancer and 19 lung cancer RNA sequencing profiles passed the 2.5 million threshold (Tables 1 and 2) and were analyzed further. The number of uniquely mapped reads appeared to be ranged from 3.96 to 20.54, which is common for sequencing of the RNA derived from FFPE [9,33].

The samples investigated were stored as FFPE tissue blocks in the Clinical Diagnostic Laboratory for 1–79 months before extraction of RNA (Figure 1). They had RNA integrity number (RIN) values ranging from 1 to 4.9, where lower RIN generally corresponded to more degraded RNA (Figure 1). We found significant correlation between the time from paraffinization to RNA extraction in days and the values of RIN (Spearman’s $\rho = -0.496$ (p -value = 0.00012); Figure 1A). However, low RIN and samples’ age turned out not to be an informative marker of the insufficient number of gene-mapped reads, and all samples with $1 \leq \text{RIN} \leq 2$ passed the coverage threshold as well (Figure 1B,C). All tumor gene expression profiles investigated were clustering mostly on a tissue-specific basis, thus confirming that they are of quality sufficient for analysis (Figure 2).

We then assessed reproducibility of gene expression profiles by performing RNA sequencing for four different slices from the same FFPE tissue block (LC specimen LuC-18, see Table 2). The resulting four replicate samples were blinded and separately sent for sequencing. For all replicates, we observed high pairwise correlation coefficients (Spearman’s ρ 0.96) between gene expression values (Figure 3). We, therefore, concluded that the RNA sequencing profiles obtained were reproducible for this sample.

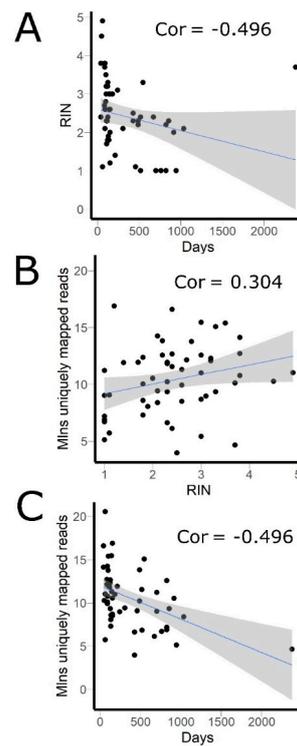


Figure 1. Effect of time interval between paraffinization and analysis (in days) on the quality of the sample. (A) RIN vs. time between paraffinization and analysis (Days): Spearman’s rho = -0.496 (p -value = 0.00012). (B) RIN vs. number of uniquely mapped reads per sample: Spearman’s rho = 0.304 (p -value = 0.022). (C) Time between paraffinization and analysis (Days) vs. number of uniquely mapped reads per sample: Spearman’s rho = -0.496 (p -value = 0.0001). Grey zone indicates 95% confidence interval for the trendlines. RIN—RNA integrity number, mln—million, cor—correlation coefficient.

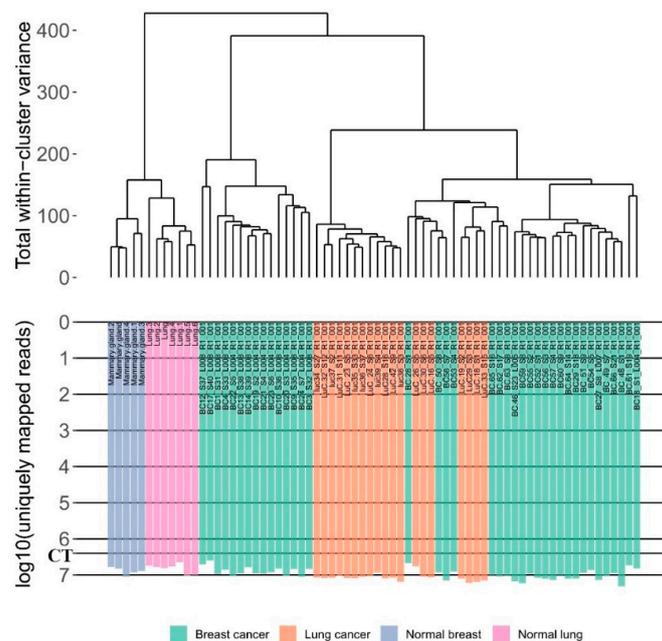


Figure 2. The hierarchical clustering dendrogram of experimental RNA sequencing profiles of breast and lung cancer and corresponding normal tissues from the ANTE database. Gene expression data were used to calculate Euclidean distances between the samples. The color markers indicate tissue types. The lower scale indicates the number of uniquely mapped reads. ‘CT’ denotes the coverage threshold of 2.5 million uniquely mapped reads.

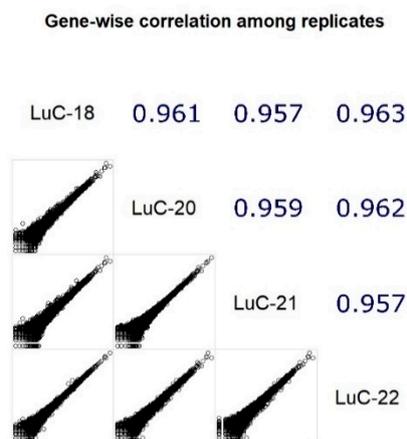


Figure 3. Correlation plot for four technical replicates (different slices from the same FFPE block) obtained from lung cancer tissue specimen. The samples were sequenced and processed separately. Top diagonal shows correlation coefficients (Spearman's rho). Bottom diagonal shows pairwise plots for gene expression values.

3.2. Comparison of RNA Sequencing and Immunohistochemistry Staining Results

We then compared expressions of clinically actionable biomarker genes measured by RNA sequencing using Oncobox protocol (same as used previously to generate the Oncobox Atlas of Normal Tissues Expression [32]) and by immunohistochemistry (IHC). For the 39 BC specimens, HER2 (ERBB2), ER (ESR1), and PR (PGR) protein levels were measured by IHC. For 19 LC specimens, PDL1 protein levels were measured by IHC. Only clinically approved protocols and reagent sets were used for the IHC measurements. We then compared these results with the corresponding gene expression values obtained from RNA sequencing data. We found that the gene expression values were highly congruent with the IHC-measured protein levels for all four genes under investigation. The highest correlations were observed for *PDL1* expression in LC (Spearman's rho = 0.797, $p = 0.00004$), *HER2* expression in BC (Spearman's rho = 0.798, $p = 6.9 \times 10^{-10}$), and *ESR1* expression in BC (Spearman's rho = 0.777, $p = 3.8 \times 10^{-9}$), while correlation with *PGR* in BC was lower yet still highly statistically significant (Spearman's rho = 0.653, $p = 4.9 \times 10^{-6}$; Figure 4).

In order to determine minimal numbers of uniquely mapped reads per sample required for statistically significant correlations between IHC and RNA sequencing data, we simulated samples with decreased coverage by randomly selecting reads from each sequencing experiment. Simulated coverage was in the range between 500 and 3,500,000 mapped reads. For each value of simulated coverage, we then calculated correlation coefficient and p-value for RNA sequencing-based gene expression vs. IHC status. We found that 3.5 million of uniquely mapped reads per sample was enough to obtain significant correlation for all biomarkers investigated, but the thresholds for minimal number of uniquely mapped reads varied for different biomarkers. Reducing the coverage to as low as to 100,000 mapped reads was enough for reliable estimation of *HER2* and *ESR1* levels in breast cancer tissues, while not less than a million mapped reads was required for *PGR* (Figure 5). We had 19 lung cancer samples, which can be the reason for greater variability observed for *PDL1* correlations across simulations. However, all correlation coefficients were significant in cases with more than 2,500,000 total coverage by gene-mapped reads (Figure 5).

To explain variability of minimal required coverage for different biomarkers, we calculated percentiles based on raw counts of each marker gene in every sample. We found that *HER2* was highly expressed at mRNA level even in IHC-negative breast cancer samples and was always in top 10% of most highly expressed genes (Figure 6). *ESR1* was in top 40% and *PGR* and *PDL1* in top 50% of the most strongly expressed genes. Therefore, higher mRNA abundance may be connected with the smaller coverage required for reliable estimation of gene expression, and vice versa.

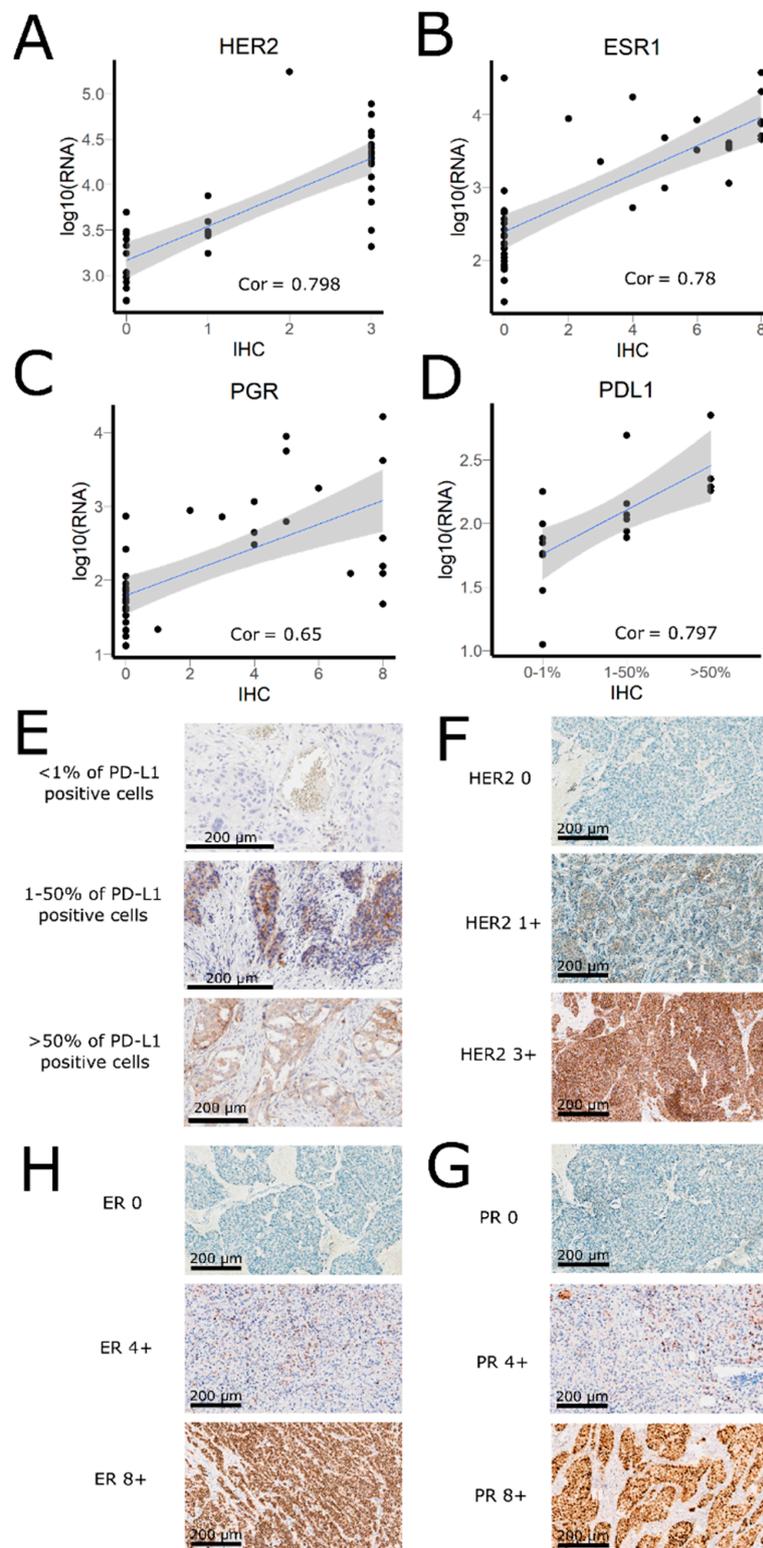


Figure 4. IHC results vs. mRNA level measured by NGS RNA sequencing: (A) HER2: correlation coefficient (Spearman's rho) = 0.798 (p -value = 6.9×10^{-10}); (B) ESR1: correlation coefficient (Spearman's rho) = 0.777 (p -value = 3.8×10^{-9}); (C) PGR: correlation coefficient (Spearman's rho) = 0.653 (p -value = 4.9×10^{-6}); (D) PD-L1: correlation coefficient (Spearman's rho) = 0.797 (p -value = 4.4×10^{-5}). Grey zone indicates 95% confidence interval for the trendlines; (E) PD-L1 IHC staining examples. (F) HER2 IHC staining examples; (H) ER (ESR) IHC staining examples; (H) PR (PGR) IHC staining examples. Cor—correlation coefficient.

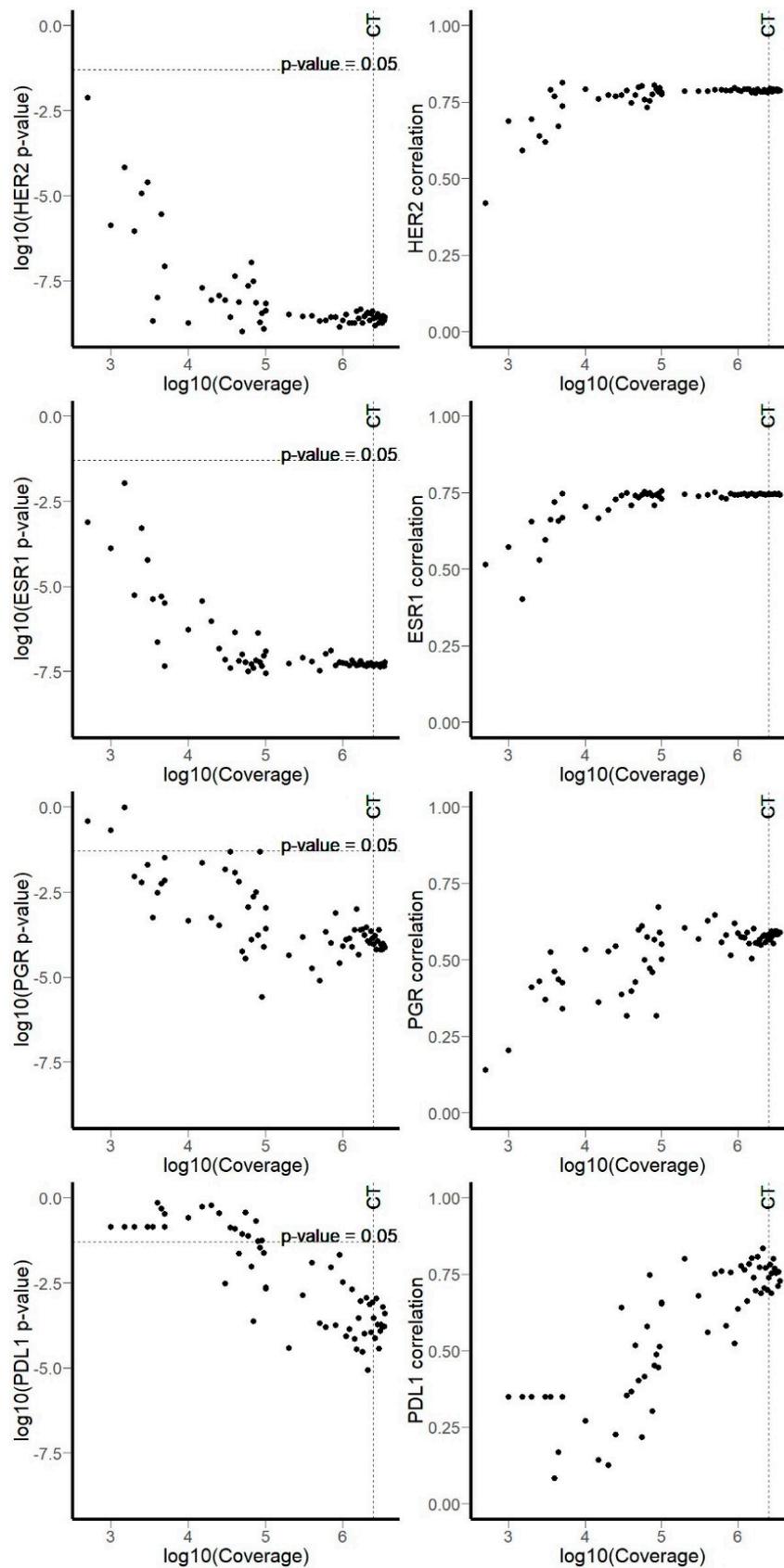


Figure 5. Computational simulation of gene-mapped reads coverage using random reads permutations. Left panels: p -value for Spearman’s rho vs. coverage. Right panels: Spearman’s rho vs. coverage. “CT” indicates coverage threshold of 2,500,000 reads.

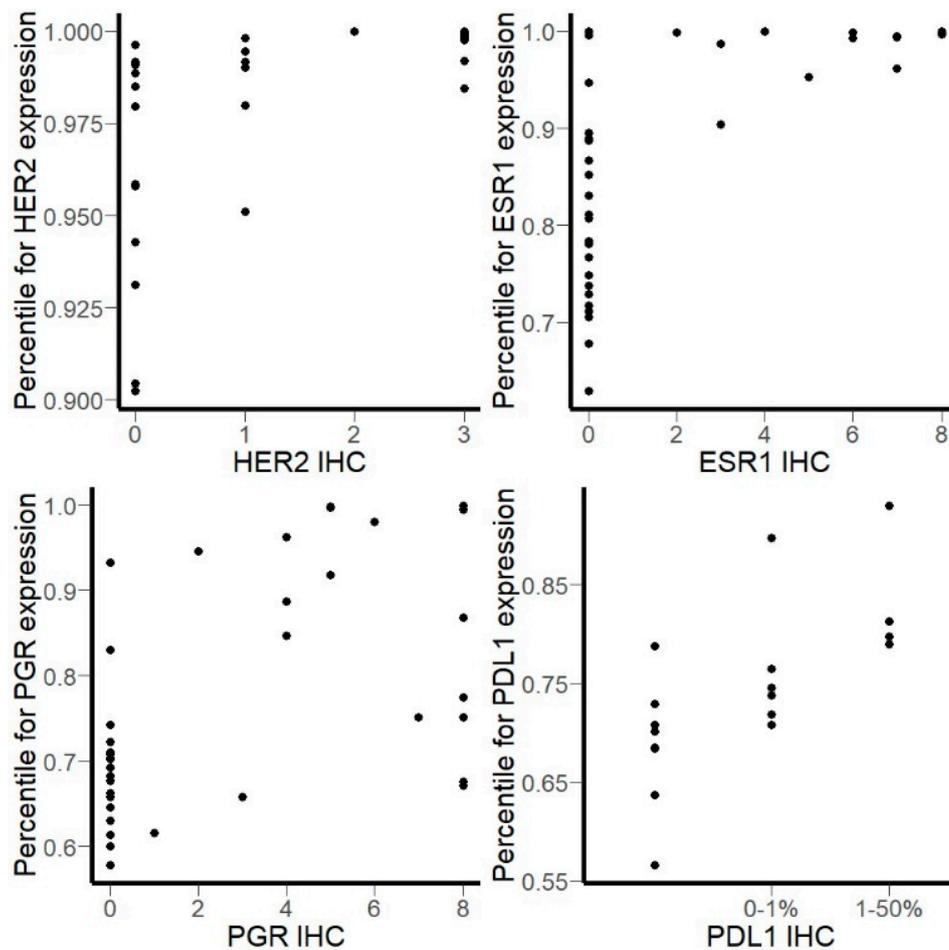


Figure 6. Percentile of gene counts for marker genes versus IHC (immunohistochemistry) score in breast and lung cancer samples.

3.3. Correlation of HER2, ER, and PGR Statuses Measured by RNA Sequencing and IHC for Freshly Frozen Tumor Samples

To estimate the ability of RNA sequencing data from fresh-frozen tissue samples to predict IHC status, we extracted from The Cancer Genome Atlas project [36] all BC data with receptor status confirmed by IHC. We used binary classification (IHC negative/positive) for this analysis because only ~20% of TCGA BC samples were annotated with exact IHC scores for ESR1 and PGR. In total, we analyzed 634 samples with confirmed HER2 status, 924 samples with confirmed ESR1 status, and 922 samples with confirmed PGR status (Figure 7). We calculated area under the receiver-operator curve (AUC) value so that RNA sequencing data could be used to classify samples by the IHC status. AUC is the universal characteristic of biomarker robustness determined by its sensitivity and specificity [42]. This statistical approach is applicable to a wide range of different types of biomarkers in oncology [43–48]. AUC positively correlates with the quality of a biomarker and varies from 0.5 to 1. The standard discrimination threshold is 0.7 and the parameters with greater AUC are considered as high-quality biomarkers, and vice versa [49]. We obtained the following AUC values for TCGA data: 0.818 for HER2, 0.959 for ESR1, and 0.923 for PGR (Table 3). We then applied the same approach to our experimental FFPE data and obtained the following AUC: 0.963 for HER2, 0.921 for ESR1, 0.912 for PGR, and 0.922 for PDL1 (Table 3).

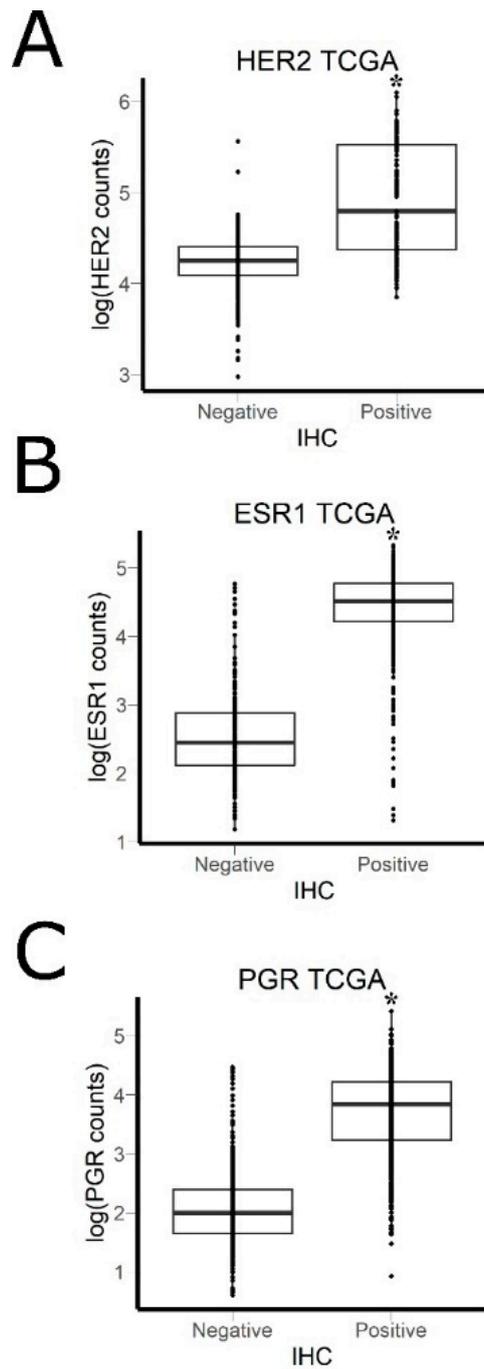


Figure 7. IHC results vs. mRNA level measured by NGS RNA sequencing in The Cancer Genome Atlas (TCGA) data: **(A)** HER2: area under the receiver-operator curve (AUC = 0.82); **(B)** ESR1: area under the receiver-operator curve (AUC = 0.96); **(C)** PGR: area under the receiver-operator curve (AUC = 0.92). * p -value < 2.2×10^{-16} (Wilcoxon rank-sum test).

Table 3. Area under the receiver-operator curve (AUC) for predicting IHC status using RNA sequencing data.

Protein	Experimental Dataset	The Cancer Genome Atlas
HER2	0.963	0.818
ESR1	0.921	0.959
PGR	0.912	0.923
PDL1	0.922	Not available

3.4. Correlation of HER2, ER, and PGR Expression Measured by RNA Sequencing versus Quantitative Proteomics

To investigate whether mRNA levels measured by RNA sequencing may serve as reliable markers of protein abundance, we analyzed quantitative proteomic profiles from The Clinical Proteomic Tumor Analysis Consortium (CPTAC) database [50,51]. The corresponding gene expression profiles for the same biosamples were extracted from the TCGA project database. For *HER2*, *ESR1*, and *PGR* analysis, we were able to identify 102 breast cancer samples with matched transcriptomic and proteomic profiles. Unfortunately, lung cancer samples were not annotated with expression of *PDL1* on protein level in CPTAC database. We, therefore, correlated mRNA and protein levels for the remaining biomarkers (Figure 8). The correlation coefficients for different biomarkers tested varied between 0.62 and 0.81.

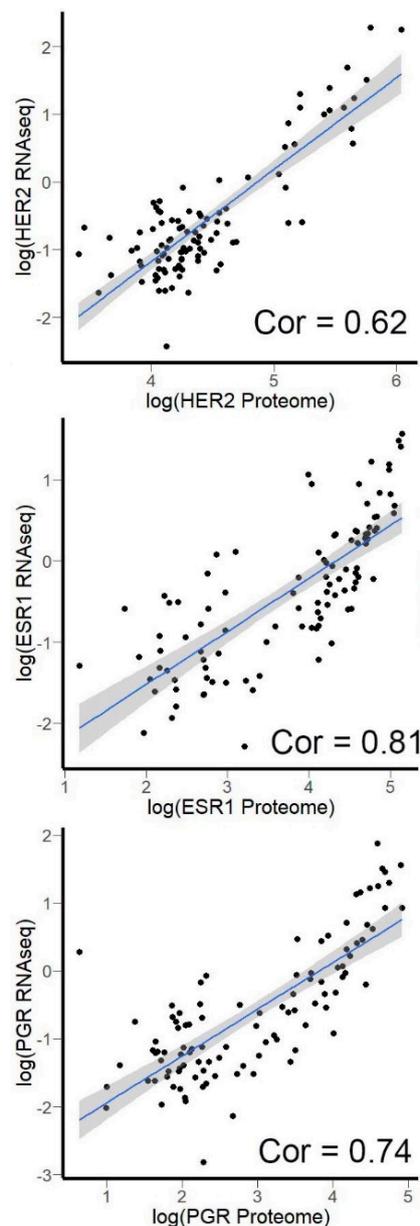


Figure 8. Proteomic results vs. mRNA level measured by NGS RNA sequencing in CPTAC data: (A) HER2: correlation coefficient (Spearman's rho) = 0.62 (p -value $< 2.2 \times 10^{-16}$); (B) ESR1: correlation coefficient (Spearman's rho) = 0.81 (p -value $< 2.2 \times 10^{-16}$); (C) PGR: correlation coefficient (Spearman's rho) = 0.74 (p -value $< 2.2 \times 10^{-16}$); Grey zone indicates 95% confidence interval for the trendlines. Cor—correlation coefficient.

4. Discussion

Immunohistochemistry (IHC) remains a method of choice for detecting expression of cancer biomarkers in most of clinical laboratories around the world [52–54]. However, RNA sequencing can be considered an even more accurate instrument for measuring the expression of biomarker genes, as this is the case for *PDL1* gene, whose expression positively correlates with patient's response to anti-PD1/PDL1 immunotherapy [55]. It was previously shown that RNA sequencing of the same biosamples from FFPE materials and matched fresh-frozen tissues produce highly concordant expression profiles for breast [54,55] and ovarian cancers [56]. In addition, RNA sequencing generated coherent biological signals for the same FFPE samples when compared with targeted NanoString [57] or TaqMan PCR assays [58] for several biomarker gene products. That RNA sequencing can help accurately measure PDL1 has been reported previously for fourteen ovarian cancer tissue specimens [56], as well as the congruence of its concentration profiles obtained using IHC, qRT-PCR, and RNA sequencing, for both fresh-frozen and FFPE cancer tissue materials [56]. Another investigation of 437 samples from patients with non-small cell lung cancer revealed high correlation between PDL1 levels measured using IHC and qRT-PCR [57]. Recently, detection of expression of mRNA in cancer cells was thoroughly investigated using optical and electrochemical biosensors [58–60]; however, despite significant progress, these promising methods have not been introduced into wide laboratory practice yet.

Previous studies also investigated the possibility of using RNA sequencing data for predicting the IHC status of five conventional breast cancer biomarkers: ESR, HER2, PGR, Ki67, and Nottingham histologic grade (NHG) [61]. The authors observed good concordance between protein status determined by IHC and the level of corresponding gene expression determined by RNA sequencing. However, the main limitation of the study by Brueffer et al. is the use of fresh-frozen or RNA later preserved tissue [61]. Conroy et al. used FFPE samples for estimating PD-L1 level in various cancer types following targeted RNA sequencing approach, which was limited by the rather small number of genes analyzed in the experiment [55]. In our study, FFPE tissue blocks were investigated using total RNA sequencing. Such an approach allowed reliable estimation of cancer biomarkers and additionally provided gene expression data on a larger scale.

Multiple layers of gene expression regulation, including post-transcriptional, translational, and post-translational, contribute to the proteomic landscape of the cell [62], and thus, may cause inconsistencies between results of RNA sequencing and immunohistochemistry measuring cancer biomarkers. However, ours and previous studies reported a high degree of concordance between these methods, at least for clinically relevant genes, thus providing evidence that RNA sequencing may complement IHC for measuring cancer biomarkers [55,61]. Although, it potentially may not be true for genes heavily regulated by post-transcriptional or post-translational modifications, and therefore, correlation between RNA sequencing and IHC should be independently validated for other biomarkers.

Here, we investigated correlations between the IHC- and RNA sequencing-measured expression profiles for four clinically actionable biomarker genes in 39 BC and 19 LC cancer biosamples. Among them, positive ESR1 and PGR status is crucial for the use of hormone therapy to treat BC, and HER2 status of 2 or 3 is an indication for targeted anti-Her2 therapeutic antibodies prescription in BC, e.g., trastuzumab [40]. In turn, PDL1 status is an important biomarker for immunotherapy in several cancer types, including lung cancer, where PDL1-positive staining of membranes of more than 50% of cancer cells serves as the key indication for prescription of PD1-specific immune check-point inhibitors, e.g., pembrolizumab, nivolumab, and atezolizumab [63]. We found that the results of RNA sequencing strongly correlate with the results obtained by IHC methods in different clinical laboratories. This suggests that, theoretically, RNA sequencing can be used to select the optimal treatment strategy for FFPE cancer tissue samples as an alternative or as an addition to IHC. In addition to simply profiling few clinical biomarker genes, RNA sequencing data enable identification of differentially expressed drug target genes [64] and measuring molecular pathway activation [21,23]. Among the others, this allows patient-oriented personalized ranking of cancer drugs with known molecular specificities [21,23,65].

However, different RNA sequencing platforms and protocols often generate incompatible results, and it is important for data reproducibility to define the experimental procedure and analytic pipeline used to obtain the results. This is especially important for comparisons with the expression in normal reference tissues [66]. To obtain and analyze RNA sequencing data, we followed strictly the procedure previously published for generating the atlas of human normal tissue transcriptomes (i.e., the Oncobox Atlas of Normal Tissue Expression) [32]. This made these two datasets fully compatible in terms of further data analysis. We also show that this experimental and analytic procedure ensures obtaining high-quality transcriptomic data that strongly correlate with the gene expression values measured by IHC. In addition to the FFPE data, we also observed a good correlation between the RNA sequencing data and the results of immunohistochemistry for the freshly frozen BC samples with known HER2, ER, and PGR statuses from the TCGA project database [36].

However, RNA sequencing provides expression levels for all genes, thus revealing much more information about a tumor, which could be applied synergistically in clinical practice. For example, high-throughput gene expression analyses were used during WINTHER [67] and Oncobox [68] clinical trials. Both trials used gene expression profiling of tumor biopsy for personalized prescription of targeted drugs to patients with advanced tumors. In addition, high-throughput gene expression profiling was previously used to select successful therapies for solid tumor patients, as described in several previous reports [69–72]. Moreover, nowadays, RNA sequencing can be performed for approximately 250 USD per sample, and this price tends to decrease further [73]. At the same time, single immunohistochemical staining procedure can cost up to 220 USD per sample [74]. Thus, with the rapid emergence of new biomarkers and their introduction into clinical practice, RNA sequencing can potentially become an at least equally useful and cost-effective solution.

Supplementary Materials: The following is available online at <http://www.mdpi.com/2227-9059/8/5/114/s1>, Table S1: TCGA BC data with IHC confirmed receptor status.

Author Contributions: M.S. (Maxim Sorokin), M.S. (Maria Suntsova), K.I., E.P., N.G., A.G., V.B. and A.B. contributed to the conception and design of the study. K.I., V.B., N.G., E.P., D.L., U.V. isolated and prepared the tissue samples. D.A. performed histological investigations. M.S. (Maria Suntsova), D.A., U.V. performed molecular analyses. M.S. (Maxim Sorokin), M.S. (Maria Suntsova), N.G., V.B., A.G., E.P., and A.B. analyzed the data. M.S. (Maxim Sorokin), M.S. (Maria Suntsova), A.B. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the OmicsWay research program in oncology and by Russian Science Foundation grant 18-15-00061 (M.S. (Maria Suntsova), M.S. (Maxim Sorokin), U.V. and A.B.).

Conflicts of Interest: M.S. (Maxim Sorokin), A.G. and A.B. have a financial relationship with OmicsWay Corp. OmicsWay Corp. was not involved in the study design, collection, analysis, and interpretation of data. The remaining authors have nothing to disclose.

References

1. Ruggles, K.V.; Krug, K.; Wang, X.; Clauser, K.R.; Wang, J.; Payne, S.H.; Fenyö, D.; Zhang, B.; Mani, D.R. Methods, Tools and Current Perspectives in Proteogenomics. *Mol. Cell. Proteom.* **2017**, *16*, 959–981. [[CrossRef](#)] [[PubMed](#)]
2. Li, J.J.; Biggin, M.D. Gene expression. Statistics requantitates the central dogma. *Science* **2015**, *347*, 1066–1067. [[CrossRef](#)] [[PubMed](#)]
3. de Klerk, E.; Fokkema, I.F.A.C.; Thiadens, K.A.M.H.; Goeman, J.J.; Palmblad, M.; den Dunnen, J.T.; von Lindern, M.; 't Hoen, P.A.C. Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res.* **2015**, *43*, 4408. [[CrossRef](#)] [[PubMed](#)]
4. Barry, K.C.; Ingolia, N.T.; Vance, R.E. Global analysis of gene expression reveals mRNA superinduction is required for the inducible immune response to a bacterial pathogen. *Elife* **2017**, *6*. [[CrossRef](#)]
5. Dunn, J.G.; Foo, C.K.; Belletier, N.G.; Gavis, E.R.; Weissman, J.S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* **2013**, *2*, e01179. [[CrossRef](#)] [[PubMed](#)]
6. Zhu, Y.; Piehowski, P.D.; Kelly, R.T.; Qian, W.-J. Nanoproteomics comes of age. *Expert Rev. Proteom.* **2018**, *15*, 865–871. [[CrossRef](#)]

7. Painter, J.T.; Clayton, N.P.; Herbert, R.A. Useful immunohistochemical markers of tumor differentiation. *Toxicol. Pathol.* **2010**, *38*, 131–141. [[CrossRef](#)]
8. Ma, L.; Liang, Z.; Zhou, H.; Qu, L. Applications of RNA Indexes for Precision Oncology in Breast Cancer. *Genom. Proteom. Bioinform.* **2018**, *16*, 108–119. [[CrossRef](#)]
9. Bossel Ben-Moshe, N.; Gilad, S.; Perry, G.; Benjamin, S.; Balint-Lahat, N.; Pavlovsky, A.; Halperin, S.; Markus, B.; Yosepovich, A.; Barshack, I.; et al. mRNA-seq whole transcriptome profiling of fresh frozen versus archived fixed tissues. *BMC Genom.* **2018**, *19*, 419. [[CrossRef](#)]
10. Nault, R.; Fader, K.A.; Zacharewski, T. RNA-Seq versus oligonucleotide array assessment of dose-dependent TCDD-elicited hepatic gene expression in mice. *BMC Genom.* **2015**, *16*, 373. [[CrossRef](#)]
11. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **2014**, *32*, 903–914. [[CrossRef](#)] [[PubMed](#)]
12. Lahens, N.F.; Ricciotti, E.; Smirnova, O.; Toorens, E.; Kim, E.J.; Baruzzo, G.; Hayer, K.E.; Ganguly, T.; Schug, J.; Grant, G.R. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genom.* **2017**, *18*, 602. [[CrossRef](#)] [[PubMed](#)]
13. Kono, N.; Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev. Growth Differ.* **2019**. [[CrossRef](#)] [[PubMed](#)]
14. Borisov, N.; Suntsova, M.; Sorokin, M.; Garazha, A.; Kovalchuk, O.; Aliper, A.; Ilnitskaya, E.; Lezhnina, K.; Korzinkin, M.; Tkachev, V.; et al. Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data. *Cell Cycle* **2017**, *16*. [[CrossRef](#)]
15. Buzdin, A.A.; Zhavoronkov, A.A.; Korzinkin, M.B.; Roumiantsev, S.A.; Aliper, A.M.; Venkova, L.S.; Smirnov, P.Y.; Borisov, N.M. The OncoFinder algorithm for minimizing the errors introduced by the high-throughput methods of transcriptome analysis. *Front. Mol. Biosci.* **2014**, *1*, 8. [[CrossRef](#)]
16. Borisov, N.; Shabalina, I.; Tkachev, V.; Sorokin, M.; Garazha, A.; Pulin, A.; Eremin, I.I.; Buzdin, A. Shambhala: A platform-agnostic data harmonizer for gene expression data. *BMC Bioinform.* **2019**, *20*, 66. [[CrossRef](#)]
17. Buzdin, A.; Sorokin, M.; Poddubskaya, E.; Borisov, N. High-Throughput Mutation Data Now Complement Transcriptomic Profiling: Advances in Molecular Pathway Activation Analysis Approach in Cancer Biology. *Cancer Inform.* **2019**, *18*, 1176935119838844. [[CrossRef](#)]
18. Tkachev, V.; Sorokin, M.; Mescheryakov, A.; Simonov, A.; Garazha, A.; Buzdin, A.; Muchnik, I.; Borisov, N. FLOating-Window Projective Separator (FloWPS): A Data Trimming Tool for Support Vector Machines (SVM) to Improve Robustness of the Classifier. *Front. Genet.* **2019**, *9*, 717. [[CrossRef](#)]
19. Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **2009**, *10*, 48. [[CrossRef](#)]
20. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)]
21. Buzdin, A.; Sorokin, M.; Garazha, A.; Sekacheva, M.; Kim, E.; Zhukov, N.; Wang, Y.; Li, X.; Kar, S.; Hartmann, C.; et al. Molecular pathway activation—New type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin. Cancer Biol.* **2018**, *53*. [[CrossRef](#)] [[PubMed](#)]
22. Aliper, A.M.; Korzinkin, M.B.; Kuzmina, N.B.; Zenin, A.A.; Venkova, L.S.; Smirnov, P.Y.; Zhavoronkov, A.A.; Buzdin, A.A.; Borisov, N.M. Mathematical Justification of Expression-Based Pathway Activation Scoring (PAS). *Methods Mol. Biol.* **2017**, *1613*, 31–51. [[PubMed](#)]
23. Buzdin, A.A.; Prassolov, V.; Zhavoronkov, A.A.; Borisov, N.M. Bioinformatics Meets Biomedicine: OncoFinder, a Quantitative Approach for Interrogating Molecular Pathways Using Gene Expression Data. *Methods Mol. Biol.* **2017**, *1613*, 53–83. [[PubMed](#)]
24. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)] [[PubMed](#)]
25. Nieuwenhuis, T.O.; Yang, S.; Verma, R.X.; Pillalamarri, V.; Arking, D.E.; Rosenberg, A.Z.; McCall, M.N.; Halushka, M.K. Basal Contamination of Sequencing: Lessons from the GTEx dataset. *bioRxiv* **2020**, 602367.
26. Cancer Genome Atlas Research Network, J.N.; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120.

27. Davis, C.A.; Hitz, B.C.; Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Gabdank, I.; Hilton, J.A.; Jain, K.; Baymuradov, U.K.; Narayanan, A.K.; et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **2018**, *46*, D794–D801. [[CrossRef](#)]
28. Huang, X.; Stern, D.F.; Zhao, H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival—Evidence from TCGA Pan-Cancer Data. *Sci. Rep.* **2016**, *6*, 20567. [[CrossRef](#)]
29. Casbas-Hernandez, P.; Sun, X.; Roman-Perez, E.; D’Arcy, M.; Sandhu, R.; Hishida, A.; McNaughton, K.K.; Yang, X.R.; Makowski, L.; Sherman, M.E.; et al. Tumor Intrinsic Subtype Is Reflected in Cancer-Adjacent Tissue. *Cancer Epidemiol. Biomark. Prev.* **2015**, *24*, 406–414. [[CrossRef](#)]
30. Zhao, Y.; Yu, P.; Wu, R.; Ge, Y.; Wu, J.; Zhu, J.; Jia, R. Renal cell carcinoma-adjacent tissues enhance mobilization and recruitment of endothelial progenitor cells to promote the invasion of the neoplasm. *Biomed. Pharmacother.* **2013**, *67*, 643–649. [[CrossRef](#)]
31. Jones, A.C.; Antillon, K.S.; Jenkins, S.M.; Janos, S.N.; Overton, H.N.; Shoshan, D.S.; Fischer, E.G.; Trujillo, K.A.; Bisoffi, M. Prostate Field Cancerization: Deregulated Expression of Macrophage Inhibitory Cytokine 1 (MIC-1) and Platelet Derived Growth Factor A (PDGF-A) in Tumor Adjacent Tissue. *PLoS ONE* **2015**, *10*, e0119314. [[CrossRef](#)] [[PubMed](#)]
32. Suntsova, M.; Gaifullin, N.; Allina, D.; Reshetun, A.; Li, X.; Mendeleeva, L.; Surin, V.; Sergeeva, A.; Spirin, P.; Prassolov, V.; et al. Atlas of RNA sequencing profiles for normal human tissues. *Sci. Data* **2019**, *6*, 36. [[CrossRef](#)] [[PubMed](#)]
33. Zhao, Y.; Mehta, M.; Walton, A.; Talsania, K.; Levin, Y.; Shetty, J.; Gillanders, E.M.; Tran, B.; Carrick, D.M. Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas. *PLoS ONE* **2019**, *14*, e0216050. [[CrossRef](#)] [[PubMed](#)]
34. Esteve-Codina, A.; Arpi, O.; Martinez-García, M.; Pineda, E.; Mallo, M.; Gut, M.; Carrato, C.; Rovira, A.; Lopez, R.; Tortosa, A.; et al. A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. *PLoS ONE* **2017**, *12*, e0170632. [[CrossRef](#)]
35. Scolnick, J.A.; Dimon, M.; Wang, I.-C.; Huelga, S.C.; Amorese, D.A. An Efficient Method for Identifying Gene Fusions by Targeted RNA Sequencing from Fresh Frozen and FFPE Samples. *PLoS ONE* **2015**, *10*, e0128916. [[CrossRef](#)]
36. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [[CrossRef](#)]
37. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
38. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
39. Harvey, J.M.; Clark, G.M.; Osborne, C.K.; Allred, D.C. Estrogen Receptor Status by Immunohistochemistry Is Superior to the Ligand-Binding Assay for Predicting Response to Adjuvant Endocrine Therapy in Breast Cancer. *J. Clin. Oncol.* **1999**, *17*, 1474–1481. [[CrossRef](#)]
40. Chen, Y.; Liu, L.; Ni, R.; Zhou, W. Advances in HER2 testing. In *Advances in Clinical Chemistry*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 91, pp. 123–162.
41. Verocq, C.; Decaestecker, C.; Rocq, L.; Clercq, S.D.; Verrellen, A.; Mekinda, Z.; Ocak, S.; Compère, C.; Stanciu-Pop, C.; Salmon, I.; et al. The daily practice reality of PD-L1 (CD274) evaluation in non-small cell lung cancer: A retrospective study. *Oncol. Lett.* **2020**, *19*, 3400. [[CrossRef](#)]
42. Green, D.; Swets, J. *Signal Detection Theory and Psychophysics*; Wiley: New York, NY, USA, 1966.
43. Chen, L.; Zhou, Y.; Tang, X.; Yang, C.; Tian, Y.; Xie, R.; Chen, T.; Yang, J.; Jing, M.; Chen, F.; et al. EGFR mutation decreases FDG uptake in non-small cell lung cancer via the NOX4/ROS/GLUT1 axis. *Int. J. Oncol.* **2019**, *54*, 370–380. [[CrossRef](#)] [[PubMed](#)]
44. Liu, T.; Cheng, G.; Kang, X.; Xi, Y.; Zhu, Y.; Wang, K.; Sun, C.; Ye, J.; Li, P.; Yin, H. Noninvasively evaluating the grading and IDH1 mutation status of diffuse gliomas by three-dimensional pseudo-continuous arterial spin labeling and diffusion-weighted imaging. *Neuroradiology* **2018**, *60*, 693–702. [[CrossRef](#)] [[PubMed](#)]
45. Tanioka, M.; Fan, C.; Parker, J.S.; Hoadley, K.A.; Hu, Z.; Li, Y.; Hyslop, T.M.; Pitcher, B.N.; Soloway, M.G.; Spears, P.A.; et al. Integrated Analysis of RNA and DNA from the Phase III Trial CALGB 40601 Identifies Predictors of Response to Trastuzumab-Based Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Clin. Cancer Res.* **2018**, *24*, 5292–5304. [[CrossRef](#)] [[PubMed](#)]

46. Zolotovskaia, M.A.; Sorokin, M.I.; Roumiantsev, S.A.; Borisov, N.M.; Buzdin, A.A. Pathway Instability Is an Effective New Mutation-Based Type of Cancer Biomarkers. *Front. Oncol.* **2018**, *8*, 658. [[CrossRef](#)]
47. Lezhnina, K.; Kovalchuk, O.; Zhavoronkov, A.A.; Korzinkin, M.B.; Zabolotneva, A.A.; Shegay, P.V.; Sokov, D.G.; Gaifullin, N.M.; Rusakov, I.G.; Aliper, A.M.; et al. Novel robust biomarkers for human bladder cancer based on activation of intracellular signaling pathways. *Oncotarget* **2014**, *5*, 9022–9032. [[CrossRef](#)]
48. Borisov, N.M.; Terekhanova, N.V.; Aliper, A.M.; Venkova, L.S.; Smirnov, P.Y.; Roumiantsev, S.; Korzinkin, M.B.; Zhavoronkov, A.A.; Buzdin, A.A. Signaling pathways activation profiles make better markers of cancer than expression of individual genes. *Oncotarget* **2014**, *5*, 10198–10205. [[CrossRef](#)]
49. Boyd, J.C. Mathematical tools for demonstrating the clinical usefulness of biochemical markers. *Scand. J. Clin. Lab. Investig. Suppl.* **1997**, *227*, 46–63. [[CrossRef](#)]
50. Kahles, A.; Lehmann, K.-V.; Toussaint, N.C.; Hüser, M.; Stark, S.G.; Sachsenberg, T.; Stegle, O.; Kohlbacher, O.; Sander, C.; Cancer Genome Atlas Research Network, G.; et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **2018**, *34*, 211–224.e6. [[CrossRef](#)]
51. Edwards, N.J.; Oberti, M.; Thangudu, R.R.; Cai, S.; McGarvey, P.B.; Jacob, S.; Madhavan, S.; Ketchum, K.A. The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* **2015**, *14*, 2707–2713. [[CrossRef](#)]
52. Kim, S.-J.; Kim, S.; Kim, D.-W.; Kim, M.; Keam, B.; Kim, T.M.; Lee, Y.; Koh, J.; Jeon, Y.K.; Heo, D.S. Alterations in PD-L1 Expression Associated with Acquisition of Resistance to ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Res. Treat.* **2018**. [[CrossRef](#)]
53. Smith, N.R.; Womack, C. A matrix approach to guide IHC-based tissue biomarker development in oncology drug discovery. *J. Pathol.* **2014**, *232*, 190–198. [[CrossRef](#)] [[PubMed](#)]
54. Adam, J.; Le Stang, N.; Rouquette, I.; Cazes, A.; Badoual, C.; Pinot-Roussel, H.; Tixier, L.; Danel, C.; Damiola, F.; Damotte, D.; et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. *Ann. Oncol.* **2018**, *29*, 953–958. [[CrossRef](#)] [[PubMed](#)]
55. Conroy, J.M.; Pabla, S.; Nesline, M.K.; Glenn, S.T.; Papanicolau-Sengos, A.; Burgher, B.; Andreas, J.; Giamo, V.; Wang, Y.; Lenzo, F.L.; et al. Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors. *J. Immunother. Cancer* **2019**, *7*, 18. [[CrossRef](#)]
56. Paluch, B.E.; Glenn, S.T.; Conroy, J.M.; Papanicolau-Sengos, A.; Bshara, W.; Omilian, A.R.; Brese, E.; Nesline, M.; Burgher, B.; Andreas, J.; et al. Robust detection of immune transcripts in FFPE samples using targeted RNA sequencing. *Oncotarget* **2017**, *8*, 3197–3205. [[CrossRef](#)] [[PubMed](#)]
57. Tsimafeyeu, I.; Imyanitov, E.; Zavalishina, L.; Raskin, G.; Povilaitite, P.; Savelov, N.; Kharitonova, E.; Rumyantsev, A.; Pugach, I.; Andreeva, Y.; et al. Agreement between PDL1 immunohistochemistry assays and polymerase chain reaction in non-small cell lung cancer: CLOVER comparison study. *Sci. Rep.* **2020**, *10*, 3928. [[CrossRef](#)]
58. Ratajczak, K.; Krazinski, B.E.; Kowalczyk, A.E.; Dworakowska, B.; Jakiela, S.; Stobiecka, M. Hairpin-Hairpin Molecular Beacon Interactions for Detection of Survivin mRNA in Malignant SW480 Cells. *ACS Appl. Mater. Interfaces* **2018**, *10*, 17028–17039. [[CrossRef](#)]
59. Stobiecka, M.; Ratajczak, K.; Jakiela, S. Toward early cancer detection: Focus on biosensing systems and biosensors for an anti-apoptotic protein survivin and survivin mRNA. *Biosens. Bioelectron.* **2019**, *137*, 58–71. [[CrossRef](#)]
60. Stobiecka, M.; Dworakowska, B.; Jakiela, S.; Lukasiak, A.; Chalupa, A.; Zembrzycki, K. Sensing of survivin mRNA in malignant astrocytes using graphene oxide nanocarrier-supported oligonucleotide molecular beacons. *Sens. Actuators B Chem.* **2016**, *235*, 136–145. [[CrossRef](#)]
61. Brueffer, C.; Vallon-Christersson, J.; Grabaut, D.; Ehinger, A.; Häkkinen, J.; Hegardt, C.; Malina, J.; Chen, Y.; Bendahl, P.-O.; Manjer, J.; et al. Clinical Value of RNA Sequencing–Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network—Breast Initiative. *JCO Precis. Oncol.* **2018**, 1–18. [[CrossRef](#)]
62. Liu, Y.; Beyer, A.; Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **2016**, *165*, 535–550. [[CrossRef](#)]
63. Passiglia, F.; Galvano, A.; Rizzo, S.; Incorvaia, L.; Listì, A.; Bazan, V.; Russo, A. Looking for the best immune-checkpoint inhibitor in pre-treated NSCLC patients: An indirect comparison between nivolumab, pembrolizumab and atezolizumab. *Int. J. Cancer* **2018**, *142*, 1277–1284. [[CrossRef](#)] [[PubMed](#)]

64. Lazar, V.; Rubin, E.; Depil, S.; Pawitan, Y.; Martini, J.-F.; Gomez-Navarro, J.; Yver, A.; Kan, Z.; Dry, J.R.; Kehren, J.; et al. A simplified interventional mapping system (SIMS) for the selection of combinations of targeted treatments in non-small cell lung cancer. *Oncotarget* **2015**, *6*, 14139–14152. [[CrossRef](#)] [[PubMed](#)]
65. Buzdin, A.; Sorokin, M.; Glusker, A.; Garazha, A.; Poddubskaya, E.; Shirokorad, V.; Naskhletashvili, D.; Kashintsev, K.; Sokov, D.; Suntsova, M.; et al. Activation of intracellular signaling pathways as a new type of biomarkers for selection of target anticancer drugs. *J. Clin. Oncol.* **2017**, *35*, e23142. [[CrossRef](#)]
66. Buzdin, A.; Sorokin, M.; Garazha, A.; Glusker, A.; Aleshin, A.; Poddubskaya, E.; Sekacheva, M.; Kim, E.; Gaifullin, N.; Giese, A.; et al. RNA sequencing for research and diagnostics in clinical oncology. *Semin. Cancer Biol.* **2019**. [[CrossRef](#)] [[PubMed](#)]
67. Rodon, J.; Soria, J.-C.; Berger, R.; Miller, W.H.; Rubin, E.; Kugel, A.; Tsimberidou, A.; Saintigny, P.; Ackerstein, A.; Braña, I.; et al. Genomic and transcriptomic profiling expands precision cancer medicine: The WINTHER trial. *Nat. Med.* **2019**, *25*, 751–758. [[CrossRef](#)]
68. Poddubskaya, E.; Buzdin, A.; Garazha, A.; Sorokin, M.; Glusker, A.; Aleshin, A.; Allina, D.; Moiseev, A.; Sekacheva, M.; Suntsova, M.; et al. Oncobox, gene expression-based second opinion system for predicting response to treatment in advanced solid tumors. *J. Clin. Oncol.* **2019**, *37*, e13143. [[CrossRef](#)]
69. Poddubskaya, E.; Bondarenko, A.; Boroda, A.; Zotova, E.; Glusker, A.; Sletina, S.; Makovskaia, L.; Kopylov, P.; Sekacheva, M.; Moiseev, A.; et al. Transcriptomics-Guided Personalized Prescription of Targeted Therapeutics for Metastatic ALK-Positive Lung Cancer Case Following Recurrence on ALK Inhibitors. *Front. Oncol.* **2019**, *9*, 1026. [[CrossRef](#)]
70. Poddubskaya, E.V.; Baranova, M.P.; Allina, D.O.; Sekacheva, M.I.; Makovskaia, L.A.; Kamashev, D.E.; Suntsova, M.V.; Barbara, V.S.; Kochergina-Nikitskaya, I.N.; Aleshin, A.A. Personalized prescription of imatinib in recurrent granulosa cell tumor of the ovary: Case report. *Cold Spring Harb. Mol. Case Stud.* **2019**, *5*, a003434. [[CrossRef](#)]
71. Poddubskaya, E.V.; Baranova, M.P.; Allina, D.O.; Smirnov, P.Y.; Albert, E.A.; Kirilchev, A.P.; Aleshin, A.A.; Sekacheva, M.I.; Suntsova, M. V Personalized prescription of tyrosine kinase inhibitors in unresectable metastatic cholangiocarcinoma. *Exp. Hematol. Oncol.* **2018**, *7*, 21. [[CrossRef](#)]
72. Moiseev, A.; Albert, E.; Lubarsky, D.; Schroeder, D.; Clark, J. Transcriptomic and Genomic Testing to Guide Individualized Treatment in Chemoresistant Gastric Cancer Case. *Biomedicines* **2020**, *8*, 67. [[CrossRef](#)]
73. BGI. Available online: <https://www.bgi.com/> (accessed on 29 March 2020).
74. Loong, H.H.; Wong, C.K.H.; Leung, L.K.S.; Dhankhar, P.; Insinga, R.P.; Chandwani, S.; Hsu, D.C.; Lee, M.Y.K.; Huang, M.; Pellissier, J.; et al. Cost Effectiveness of PD-L1-Based Test-and-Treat Strategy with Pembrolizumab as the First-Line Treatment for Metastatic NSCLC in Hong Kong. *Pharmacoeconomics Open* **2019**. [[CrossRef](#)] [[PubMed](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).