

## Article

# Mutation Enrichment and Transcriptomic Activation Signatures of 419 Molecular Pathways in Cancer

Marianna A. Zolotovskaia <sup>1,2,3,\*</sup>, Victor S. Tkachev <sup>4</sup>, Alexander P. Seryakov <sup>5</sup>, Denis V. Kuzmin <sup>3</sup>, Dmitry E. Kamashev <sup>6,7</sup>, Maxim I. Sorokin <sup>4,6,7</sup>, Sergey A. Roumiantsev <sup>2</sup> and Anton A. Buzdin <sup>1,3,6,7</sup>

<sup>1</sup> Oncobox Ltd., Skolkovo Innovation Center, 121205 Moscow, Russia; buzdin@oncobox.com

<sup>2</sup> Department of Oncology, Hematology and Radiotherapy of Pediatric Faculty, Pirogov Russian National Research Medical University, 117997 Moscow, Russia; s\_roumiantsev@mail.ru

<sup>3</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, 141701 Moscow, Russia; kuzmin.dv@mipt.ru

<sup>4</sup> Omicsway Corp., Office 6040, Walnut, CA 91789, USA; tkachev@oncobox.com (V.S.T.); sorokin@oncobox.com (M.I.S.)

<sup>5</sup> Medical Holding SM-Clinic, 105120 Moscow, Russia; alseryakov@yandex.ru

<sup>6</sup> The Laboratory of Clinical Bioinformatics, I.M. Sechenov First Moscow State Medical University, 119991 Moscow, Russia; dkamashev@gmail.com

<sup>7</sup> Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 117997 Moscow, Russia

\* Correspondence: zolotovskaya@oncobox.com

Received: 16 December 2019; Accepted: 20 January 2020; Published: 22 January 2020

**Abstract:** Carcinogenesis is linked with massive changes in regulation of gene networks. We used high throughput mutation and gene expression data to interrogate involvement of 278 signaling, 72 metabolic, 48 DNA repair and 47 cytoskeleton molecular pathways in cancer. Totally, we analyzed 4910 primary tumor samples with individual cancer RNA sequencing and whole exome sequencing profiles including ~1.3 million DNA mutations and representing thirteen cancer types. Gene expression in cancers was compared with the corresponding 655 normal tissue profiles. For the first time, we calculated mutation enrichment values and activation levels for these pathways. We found that pathway activation profiles were largely congruent among the different cancer types. However, we observed no correlation between mutation enrichment and expression changes both at the gene and at the pathway levels. Overall, positive median cancer-specific activation levels were seen in the DNA repair, versus similar slightly negative values in the other types of pathways. The DNA repair pathways also demonstrated the highest values of mutation enrichment. However, the signaling and cytoskeleton pathways had the biggest proportions of representatives among the outstandingly frequently mutated genes thus suggesting their initiator roles in carcinogenesis and the auxiliary/supporting roles for the other groups of molecular pathways.

**Keywords:** cancer; DNA mutation; intracellular molecular pathways; carcinogenesis; transcriptome; molecular pathway activation

## 1. Introduction

Cancer has complex pathogenesis [1] and molecular mechanisms underlying its development and progression still remain underexplored [2]. According to the somatic mutation theory, the key point in cancer transformation is DNA damage, e.g., due to aberrantly regulated or nonfunctional DNA repair pathways [3]. Cancer cells accumulate mutations several times faster than the normal cells [4]. It contributes to deregulation of molecular pathways including those responsible for apoptosis, cell growth, metabolism, motility and immunosuppression [5,6]. In turn, this leads to epigenetic changes that promote further malignization, e.g., through the repression of tumor suppressor genes [7].

However, the analysis of individual genes *per se* is not enough for understanding mechanisms of carcinogenesis because each gene product only serves as a component of complex biological processes determining cell fate. One of the key systemic genetic approaches in cancer is data analysis at the level of molecular pathways [8–11]. It was found previously that changes at the molecular pathway levels characterize cancers better than individual gene mutations and expression levels due to greater stability and reliability of the additive biomarkers [12–14].

Several databases of molecular pathways were published, frequently specifically devoted to the pathways with specific functions, such as MetaCyc and SynSysNet [15,16]. On the other hand, a more universal database like Kyoto Encyclopedia of Genes and Genomes (KEGG) classifies the pathways by relation to “metabolism, processing of genetic information, processing of environmental information, cellular processes, organizational systems, diseases and drug development” [17]. Many attempts to investigate cancer interactomes on the basis of molecular pathways have been made, showing mosaic picture of tumor-associated lesions [18–22]. In this study, we aimed to perform the most comprehensive analysis of the molecular pathway’s mutation and activation features in cancer. To this end, we used new instruments to algorithmically assess pathway activation levels [23] and pathway mutation instability rates [24] by analyzing high throughput gene expression and whole exome sequencing data. To our knowledge, these metrics were never systemically investigated before for simultaneous large-scale characterization of molecular pathways in cancer. Using descriptions provided by the pathway database administrators and available literature, we classified 419 available molecular pathways into four functional groups of *signaling*, *metabolic*, *cytoskeleton* and *DNA repair* processes (including 278, 72, 47 and 48 pathways, respectively).

The *signaling pathways* transmit different types of signals into the cell, e.g., signals from external cell surface receptors. They control all major biological processes in the cell such as proliferation, cell growth, migration, differentiation and death. Impaired molecular signaling can lead to acquisition of cancer phenotype by the cell and further disease progression [25–28].

*Metabolic pathways* are responsible for the whole repertoire of biochemical reactions in organism. In the living cell they are dynamically controlled by the external signals, concentrations of biomolecules and differentiation programs [29]. In cancer, the regulation of metabolism is strongly biased due to increased consumption of energy required for forced proliferation [30,31]. For example, Warburg effect of replacing oxidative phosphorylation as the major provider of ATP by glycolysis in cancer cells is well known since 1920s [32]. On the other hand, cancer cells frequently de-differentiate and abrogate their specialized molecular functions, thus losing complexity of metabolic patterns [33].

In turn, the *cytoskeleton* and *DNA repair* pathways may be considered as the specific sort of signaling pathways because many of them if not all are directly controlled by external or internal stimuli [34,35]. Imbalanced *cytoskeleton pathways* can lead to defects in mitosis and cellular morphogenesis, altered intercellular contacts and cell motility. This can promote cancer progression, invasion and metastasis [36].

Finally, mutations in *DNA repair* pathways are most likely the key mechanisms in the emergence and development of malignant tumors. Mutations in genes responsible for genome integrity frequently point on individuals with a predisposition to cancer, but also may serve as biomarkers of response on specific anticancer therapies [37]. These mutations also induce multiplicative tumorigenic effects because they initiate accelerated molecular evolution of cancer cells, that can both repress tumor suppressor genes and upregulate oncogenes, while escaping restriction by the immune system [38,39]. Moreover, a link was shown between functional changes in DNA repair genes and switching of cancer cell metabolic programs [40,41].

In this study we for the first time scrutinized mutation frequencies and expression profiles of genes included in 419 molecular pathways belonging to the above four groups with different biological functions. Totally, we analyzed 4910 matched individual cancer gene expression and whole exome sequencing profiles collectively covering 1,252,669 mutations from the TCGA project database [42] in thirteen cancer types. We identified no correlation between mutation enrichment and expression changes both at the individual gene and molecular pathway levels. However, we found that pathway activation profiles were largely congruent among the different cancer types. Overall,

positive median cancer-specific activation levels were seen in the DNA repair, versus similar slightly negative values in the other types of pathways. The DNA repair pathways also demonstrated the highest values of mutation enrichment. However, the signaling and cytoskeleton pathways had the biggest proportions of representatives among the outstandingly frequently mutated genes thus suggesting their initiator roles in carcinogenesis and the auxiliary/supporting roles for the other groups of molecular pathways.

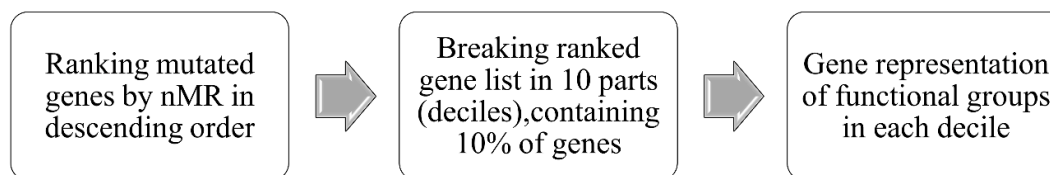
For all the biosamples, we present a database of mutation and expression data at the individual gene and at the molecular pathway levels. At the gene level, normalized mutation rate and case-to-normal ratio data are given. Conversely, at the pathway level we present algorithmically calculated pathway mutation instability rates [24] and pathway activation levels [23], respectively.

## 2. Results

### 2.1. Cancer Mutation Frequencies of Molecular Pathway Genes

We extracted molecular pathway gene contents and architecture from the publicly available databases [17,43–47] and refined by expert manual curation. We classified the pathways extracted into four major functional groups: signaling (278 members), metabolic (72 members), and DNA repair (48 members), cytoskeleton pathways (47 members), listed in Supplementary Table S1. Twenty-six pathways were simultaneously classified as the signaling and DNA repair pathways, and our further analyses we used only the 419 non-overlapping molecular pathways.

We then measured mutation rates of genes belonging to these functional groups of pathways (Supplementary Table S2). To this end, a complete list of non-synonymous somatic mutations mapped in 19,608 genes in 4910 tumor samples was built and processed according to the analytic pipeline outlined in Figure 1.



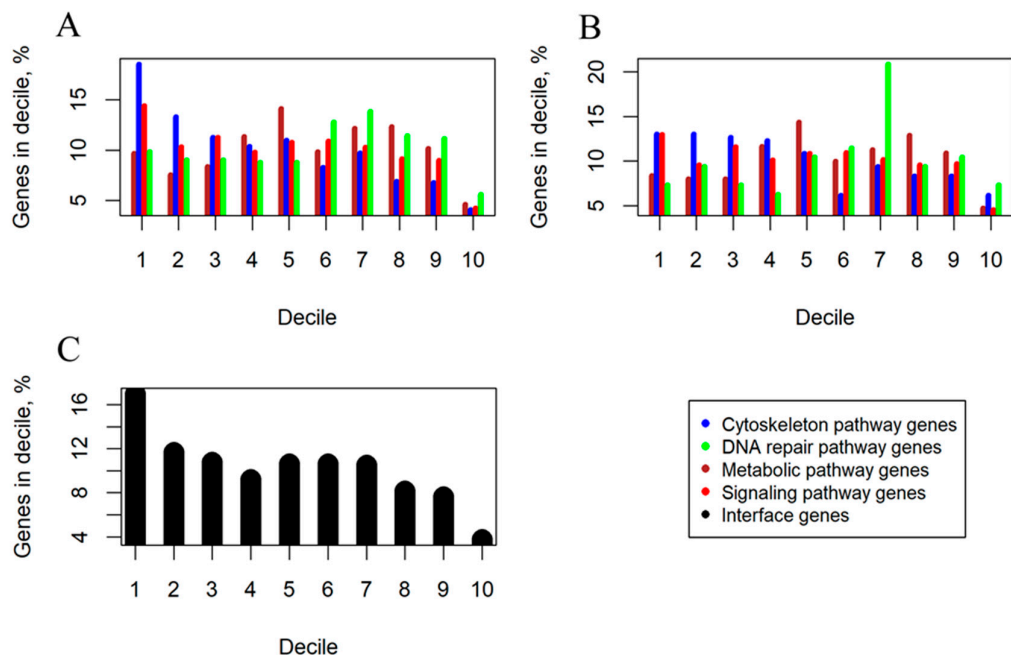
**Figure 1.** Pipeline for bioinformatic analysis of gene mutation frequencies corresponding to functional groups of molecular pathways.

The tumor samples analyzed represented thirteen tumor types and were originally profiled within the The Cancer Genome Atlas (TCGA) project [42] while the respective whole exome mutation profiles were taken from the Catalogue Of Somatic Mutations In Cancer (COSMIC) database [48]. Normalized mutation rates (nMR) of all genes investigated are listed in Supplementary Table S3.

The nMR values were ranked in the descending order and then divided into 10 parts (deciles), each representing 10% of genes. In each decile, we measured proportions of genes belonging to the four functional groups of pathways under investigation (Figure 2). We observed a trend that the *signaling* and *cytoskeleton* pathways had highest representations in the first decile which was decreasing further to the minimal values in the last decile, thus evidencing high proportion of especially frequently mutated genes in these groups (Figure 2). In contrast, the *metabolic* and *DNA repair* pathways showed a different trend, where maximum representations were reached in the central deciles, with relatively modest representations in the first and in the last deciles (Figure 2).

We then re-analyzed these representations and confirmed the above two trends separately for the thirteen cancer types (Figure 3; Supplementary Table S4). These trends were more pronounced for the cancer types represented by bigger number of samples, most probably, due to greater statistical power (Figure 3)

To characterize in brief gene compositions of the molecular pathways, we intersected them for the different groups of pathways (Figure 4). Jaccard (JC) and Szymkiewicz-Simpson (SS) [49] similarity coefficients between the pathway groups are shown on Table 1.



**Figure 2.** Decile analysis of cancer-associated mutations (deciles ordered from high gene mutation rate on the left to low on the right). **(A)** Percentage shares of genes belonging to signaling, metabolic, cytoskeleton, DNA repair pathway functional groups. **(B)** Percentage shares of pathway group-specific, non-overlapping genes. **(C)** Percentage shares of genes present in two or more functional groups of molecular pathways (termed interface genes).

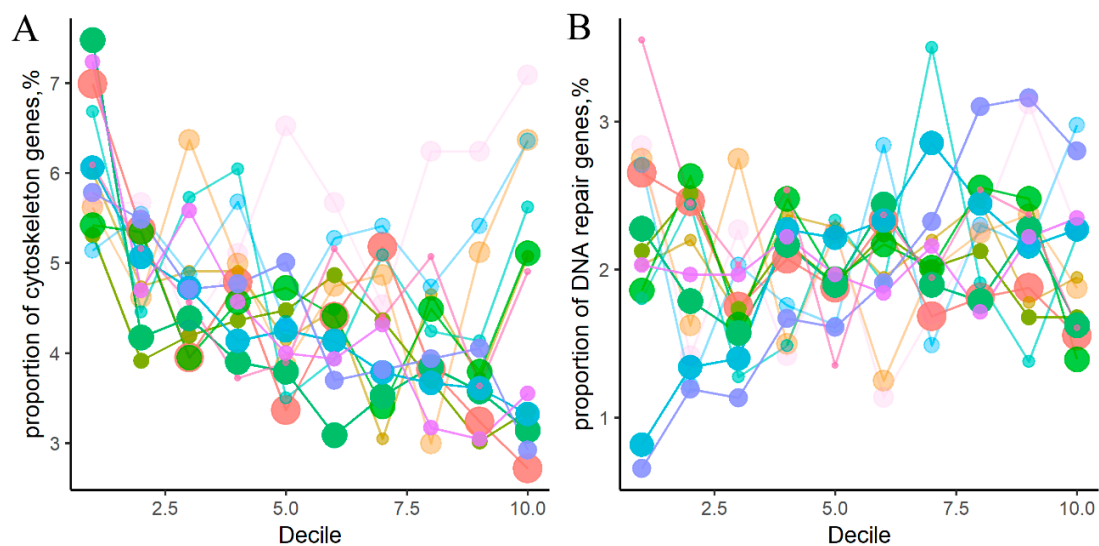
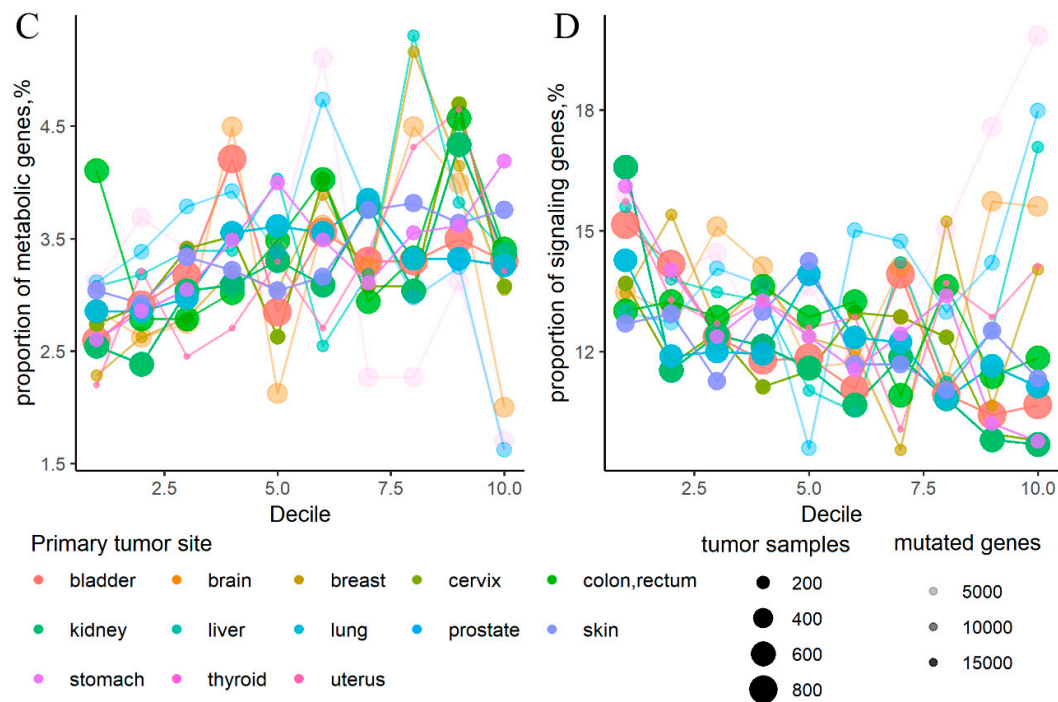
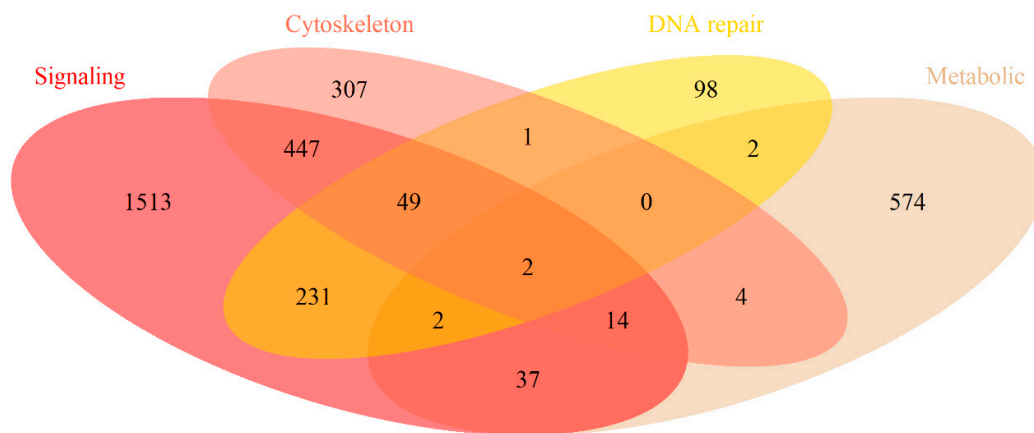


Figure 3. Cont.



**Figure 3.** Percentage share in different cancer types of genes belonging to four groups of molecular pathways: (A) cytoskeleton, (B) DNA repair, (C) metabolic, (D) signaling pathways. Color represents cancer type. Color density is proportionate to number of mutant genes per group per decile. Point width reflects number of samples in cancer type.



**Figure 4.** Intersection of gene composition for different functional groups of molecular pathways.

**Table 1.** Similarity coefficients between gene compositions of different groups of molecular pathways. Jaccard coefficients are shown above and Szymkiewicz-Simpson coefficients—below the main diagonal.

Intersection	Signaling	Metabolic	Cytoskeleton	DNA Repair
Signaling	1	0.019	0.196	0.119
Metabolic	0.087	1	0.014	0.006
Cytoskeleton	0.621	0.031	1	0.045
DNA repair	0.738	0.016	0.135	1

In total, all the 419 pathways contained 3281 unique genes, of them 809 were the interface genes, i.e., simultaneously members of two or more groups of pathways. *Signaling* pathways included 2295 genes, of them 1513 were specific for this group and the rest were the interface genes. *Metabolic* pathways had totally 635 genes and 574 specific genes; *cytoskeleton* pathways—824 and 307 genes, respectively; *DNA repair* pathways—385 and 98 genes, respectively. The highest intersection level was seen for the *signaling* and *cytoskeleton* pathway genes (JC = 0.196, SS = 0.621) and for the *signaling* and *DNA repair* genes (JC = 0.119, SS = 0.738).

To avoid any possible bias linked with the interface genes, we then repeated the decile distribution analysis for the fraction of pathway-specific, non-overlapping genes. The decile-specific gene distribution trends previously seen for the *signaling/cytoskeleton* vs *metabolic/DNA repair* groups of pathways were also clearly detectable for the pathway type-specific genes (Figure 2B). Moreover, the fraction of interface genes that was mostly formed by genes from the intersection of *signaling* and *cytoskeleton* groups of pathways also showed the same trend as for the *signaling/cytoskeleton* groups (Figure 2C). Interestingly, the interface genes had overall higher mutation rates than the pathway type-specific genes (Figures 2B,C), which can relate to their relatively more important functions in tumorigenesis, e.g., because of their simultaneous participation in different types of molecular processes.

Every tumor sample investigated had mutated genes. Among them, 78% of samples had at least *one* mutated gene from cytoskeleton, 53%—from DNA repair, 58%—from metabolic, and 94%—from signaling pathways (Table 2). Moreover, 65% of tumor samples had at least *two* mutated genes from cytoskeleton, 35%—from DNA repair, 41%—from metabolic, and 89%—from signaling pathways, and so on up to 20 mutated genes, further showing prevalence of signaling and cytoskeleton pathways (Table 2).

**Table 2.** Percentage of 4910 tumor samples with mutated genes of four pathway groups.

Minimal Number of Mutated Genes	Cytoskelton	DNA Repair	Metabolic	Signaling	Minimal Number of Mutated Genes	Cytoskelton	DNA Repair	Metabolic	Signaling
1	78%	53%	58%	94%	11	17%	6%	9%	46%
2	65%	35%	41%	89%	12	16%	5%	8%	42%
3	54%	24%	31%	84%	13	15%	5%	8%	39%
4	45%	18%	24%	78%	14	14%	4%	7%	37%
5	38%	14%	19%	72%	15	13%	4%	6%	35%
6	32%	11%	16%	66%	16	12%	4%	6%	33%
7	27%	9%	14%	61%	17	11%	3%	5%	31%
8	24%	8%	12%	56%	18	10%	3%	5%	29%
9	21%	7%	11%	52%	19	10%	3%	5%	28%
10	19%	6%	10%	48%	20	9%	3%	5%	26%

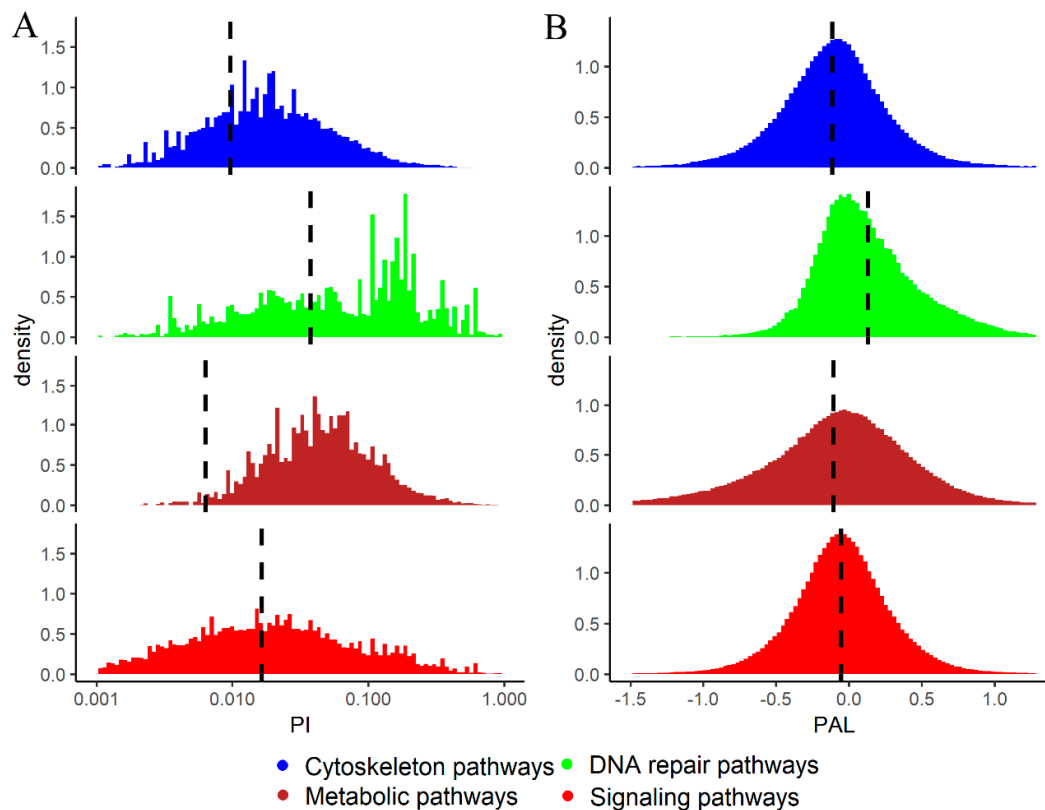
## 2.2. Mutation Enrichment and Pathway Activation Levels

The above 4910 tumor samples were selected for the pathway analysis based on simultaneous availability of both RNA sequencing and whole exome mutation profiles. In addition, from the TCGA data repository we extracted the gene expression profiles for 655 tumor-matched adjacent normal tissue samples. These normal samples represented the same localizations/tissue types as the corresponding primary tumor samples (Supplementary Table S5).

We then used tumor and normal RNA sequencing profiles to analyze differential gene expression and to assess activation levels of molecular pathways. For every comparison, the tumor profiles were normalized on the corresponding normal tissues, e.g., breast cancer profiles were normalized on the normal breast samples, etc. *Pathway activation level (PAL)* values were calculated to characterize molecular pathway activation using differential gene expression data [23,50,51], Supplementary Table S6. *PAL* scores can take positive and negative values and are congruent with the expected differential activation of a pathway [23].

The mutation enrichment of the same molecular pathways in 4910 tumors was assessed using *Pathway instability (PI)* metric that characterizes average mutation load per gene in a pathway [14], Supplementary Table S7. *PI* score can take only positive values and is congruent with the pathway mutation burden.

In the tumor samples investigated, 32.3% of *signaling* pathways, 30.4% of *cytoskeleton* pathways, 29.2% of *DNA repair* pathways, and 9.6% of *metabolic* pathways had mutations in their genes and, therefore, had *PI* values greater than zero. Interestingly, the distribution of *PI* scores demonstrated that overall the *DNA repair* pathways had the highest relative mutation rates per gene, the *signaling* pathways had intermediate scores and the *metabolic* and *cytoskeleton* pathways had the lowest relative mutation levels (Figure 5A).

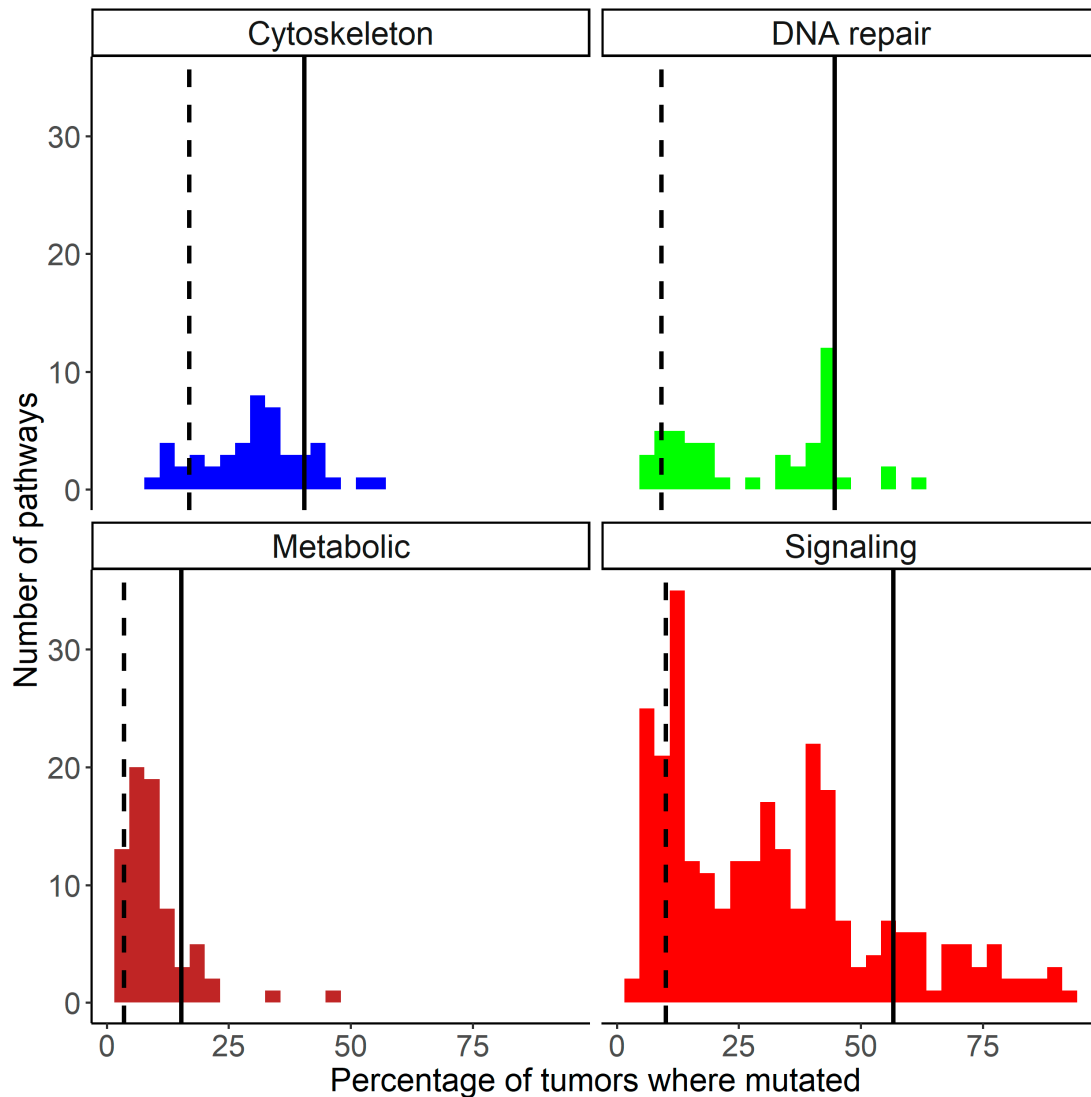


**Figure 5.** (A) Distributions of  $PI > 0$  scores for cytoskeleton, metabolic, DNA repair and signaling molecular pathways in 4910 cancer samples. *PI* axis is given in logarithmic scale. Vertical dashed lines show mean *PI* values calculated for all cases including  $PI = 0$ . (B) Distributions of *PAL* scores for cytoskeleton, metabolic, DNA repair and signaling molecular pathways in the same cancer samples. Vertical dashed lines show mean *PAL* values calculated for the respective group of pathways for all tumor samples.

We then investigated in more details what biological processes were connected with the genes from the most and least mutated pathways. To this end, for every type of pathways we took top and



bottom 15% pathways (Figure 6; Supplementary Table S8). To identify relevant Gene Ontology terms, the gene lists from these top and bottom pathways were analyzed using GOrilla software [52]. The obtained terms were then plotted with REVIGO software [53], Figure 7. The least mutated pathways (Figure 7A–D, left panels) dealt (*cytoskeleton*, Figure 7A) with membrane organization, receptor-mediated endocytosis, actin filament binding and locomotion, calcium-independent cell-cell adhesion via plasma membrane; (*DNA repair*, Figure 7B) with regulation of DNA quality and associated DNA repair, and with rearrangement of immunoglobulin loci; (*metabolic*, Figure 7C) with biosynthesis of nucleotides and DNA replication; (*signaling*, Figure 7D) with negative regulation of glucocorticoid receptor pathway, regulation of hormone levels, circadian rhythms, H3 histone deacetylation and protein catabolism.

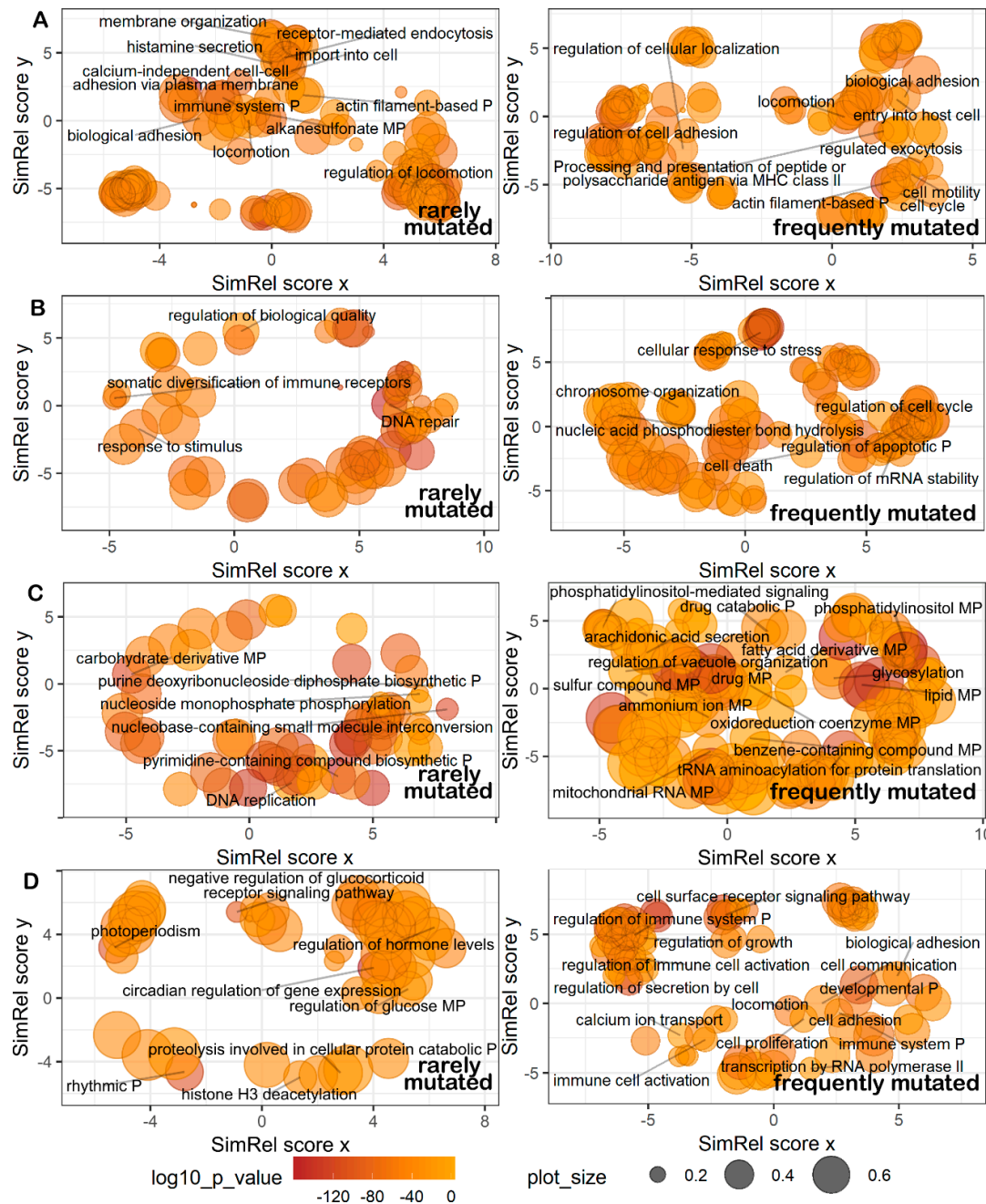


**Figure 6.** Distributions of molecular pathways according to percentage of tumor samples having mutations within them, for cytoskeleton, DNA repair, metabolic, signaling pathways. Vertical solid lines delineate 15% most frequently mutated pathways (**on the right**), vertical dashed lines delineate 15% least mutated pathways (**on the left**).

In turn, the most strongly mutated pathways (Figures 7A–D, right panels) were responsible for (*cytoskeleton*, Figure 7A) cellular adhesion and cell surface receptor binding, exocytosis, cell motility, cell communication and locomotion, cell cycle progression, processing and presentation of antigens



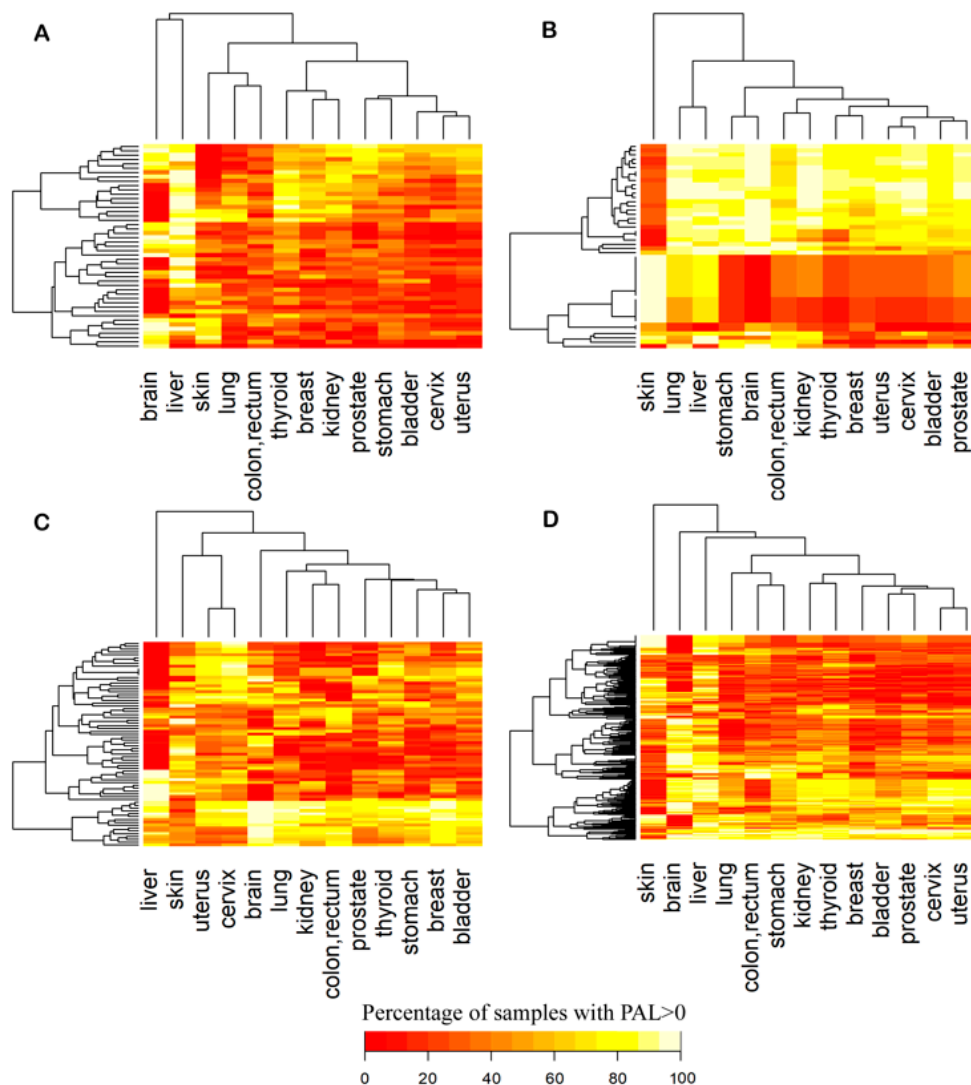
via the major histocompatibility complex (MHC) class II; (*DNA repair*, Figure 7B) cellular response to stress, chromosome organization, hydrolysis of phosphodiester bonds; (*metabolic*, Figure 7C) drug response and catabolism, arachidonic acid secretion, fatty acid derivative pathways, glycosylation, sulfur and benzene-containing compound pathways, production of mitochondrial RNA, and tRNA aminoacylation for protein translation; (*signaling*, Figure 7D) regulation through phosphatidylinositol, cell surface receptor pathways, growth and proliferation, regulation of calcium transport, transcription by RNA polymerase II, and regulation of immune cell activation.



**Figure 7.** Representation of Gene Ontology annotations for genes from 15% least (**left**) and 15% most frequently (**right**) mutated (**A**) cytoskeleton, (**B**) DNA repair, (**C**) metabolic, (**D**) signaling pathways. Bubble size indicates the frequency of the gene ontology term in the underlying GOA database [54] (bubbles corresponding to more general terms are shown larger). “P” stands for process, “MP” —for metabolic process. Too general and non-informative gene ontology terms are not defined.

At the level of pathway activation (PAL scores), most of the pathway types were mostly downregulated in cancer, except for the *DNA repair* pathways that were mostly upregulated (Figure 5B). The general downregulation observed was maximal for the *cytoskeleton* pathways (Figure 5B). Overall, averaged PAL scores showed cancer-specific activation in approximately 60.3% of *DNA repair*, 44.7% of *metabolic*, 42.8% of *signaling* and 37.5% of *cytoskeleton* pathways.

We then investigated in more detail up- and downregulated pathways in each group. Because the same pathway can be differently regulated in different samples, we tried to focus on cancer type-specific changes. To this end, for every pathway in every cancer type, we calculated percentage of samples in which the pathway was upregulated, as reflected by  $PAL > 0$  (Figure 8). Most of DNA repair pathways were activated in the majority of samples (Figure 8B), whereas cytoskeleton, metabolic and signaling pathways showed mosaic picture (Figures 8A,C,D). We also found that most of the pathways showed uniform, cancer type-independent activation pattern in tumor samples (Figure 8).

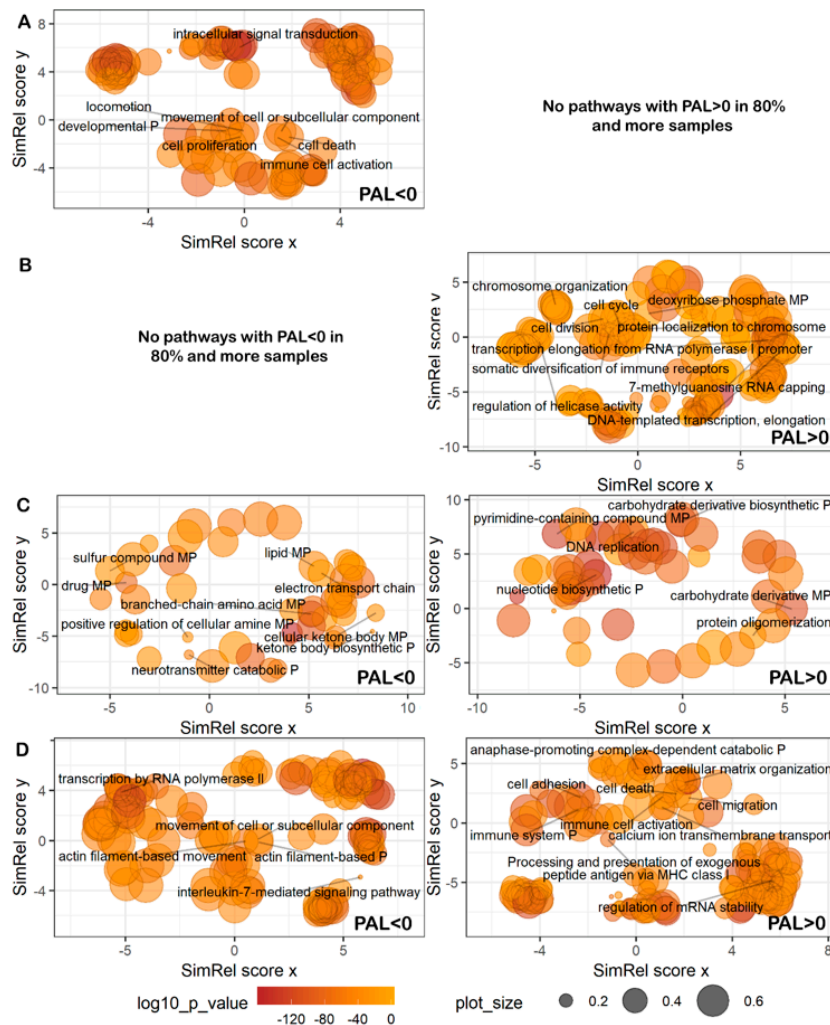


**Figure 8.** Percentage shares of cancer samples with  $PAL > 0$  in thirteen cancer types for: (A) cytoskeleton, (B) DNA repair, (C) metabolic, (D) signaling pathways under investigation. For every pathway, color scale corresponds to percentage of samples where it is upregulated ( $PAL > 0$ ).

We then performed Gene Ontology analyses of up/downregulated pathways. To this end, we took fractions of top pathways with  $PAL > 0$  in more than 80% of cancer samples under investigation, and of bottom pathways with  $PAL < 0$  in more than 80% of the samples. In this way we selected

totally 8 up/14 downregulated signaling pathways, 4/2 metabolic pathways, 15/0 DNA repair pathways, and 0/2 cytoskeleton pathways, accordingly. As before, the genes from these pathways were subjected for gene ontology analysis using GOrilla software [52] the results of which were visualized using REVIGO [53] viewer (Figure 9; Supplementary Table S9). The most strongly cancer-downregulated pathways related to (*cytoskeleton*, Figure 9A) movement of cell or subcellular components and locomotion, regulation of development and cell proliferation; (*metabolic*, Figure 9C) cellular ketone body metabolism, neurotransmitter catabolic processes, sulfur pathways, drug response pathways; (*signaling*, Figure 9D) regulation of transcription by RNA polymerase II, interleukin-7-mediated pathway. We found no strongly cancer-downregulated DNA repair pathways.

Among the most strongly cancer-upregulated pathways (Figure 9A–D, right panels), there were no *cytoskeleton* pathways, whereas the others dealt with (*DNA repair*, Figure 9B) response to DNA templated transcription and elongation, 7-methylguanosine RNA capping, protein localization on chromosomes and chromosomal organization, rearrangement of immune receptor loci; (*metabolic*, Figure 9C) oligosaccharide biosynthesis and protein oligomerization, nucleotide biosynthesis and DNA replication; (*signaling*, Figure 9D) cell migration and adhesion, regulation of cell death, calcium ion transmembrane transport, extracellular matrix organization, regulation of mRNA stability, processing and presentation of exogenous peptide antigen via MHC class I, and activation of immune cells.

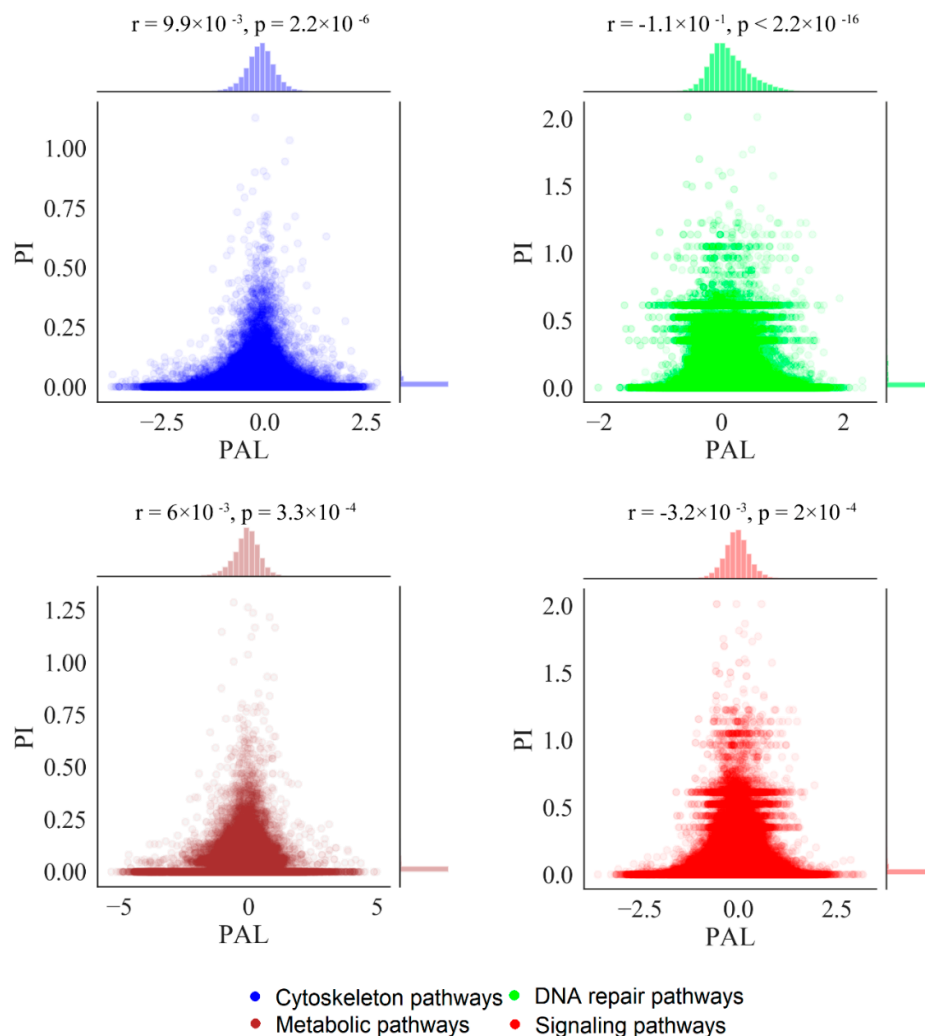


**Figure 9.** Representation of Gene Ontology annotations for genes from most strongly downregulated (PAL < 0; left) and upregulated (PAL > 0; right) (A) cytoskeleton, (B) DNA repair, (C) metabolic, (D) signaling pathways. Bubble size indicates the frequency of the Gene Ontology term in the underlying

GOA database [54] (bubbles corresponding to more general terms are shown larger). “P” stands for process, “MP” —for metabolic process. Too general and non-informative gene ontology terms are not defined. No cytoskeleton pathways with PAL < 0 and DNA repair pathways with PAL > 0 could be identified in 80% and more samples.

Remarkably, we found no connection between mutation burden (PI scores) and the respective activation of the specific molecular processes (PAL scores). For example, the processes of transcription by RNA polymerase II, chromosomal organization, processing and presentation of antigens via MHC class II and activation of immune cells were featured for the genes from *strongly mutated* and at the same time *upregulated* molecular pathways. The drug response, sulfur pathway categories included genes from *strongly mutated, downregulated* pathways. The rearrangement of immunoglobulin loci, biosynthesis of nucleotides, DNA replication, and circadian rhythmic processes involved genes of *poorly mutated, upregulated* pathways.

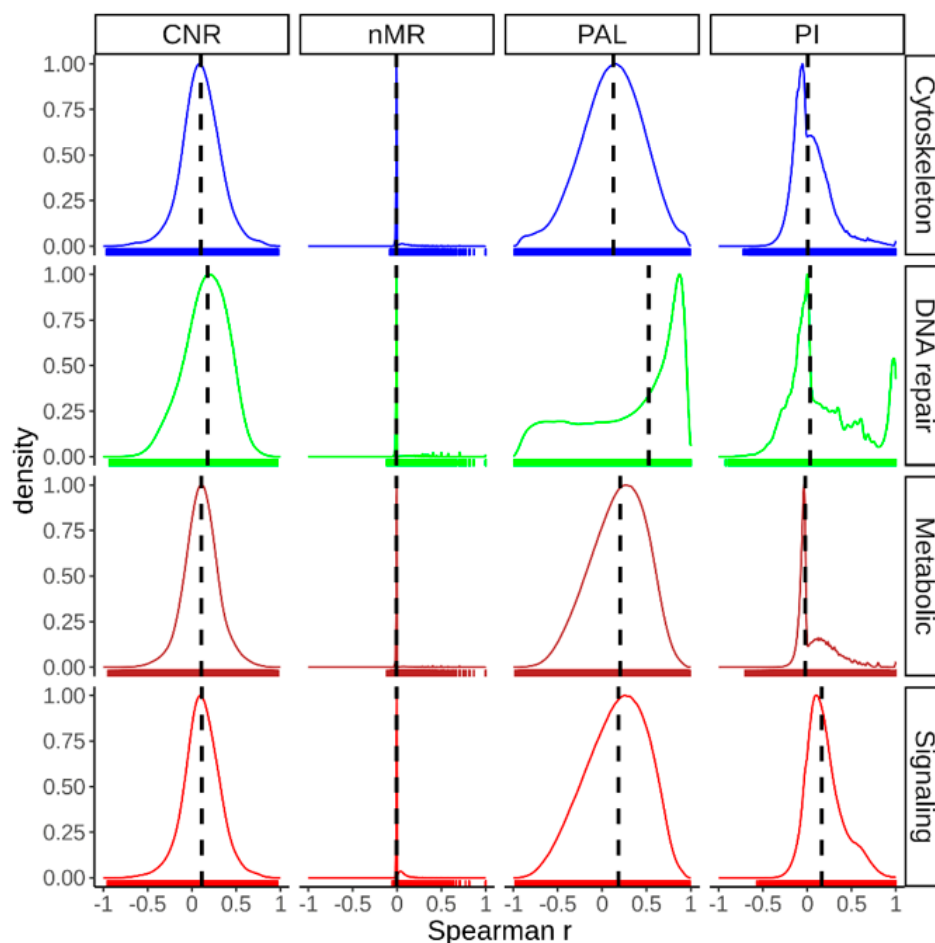
For all types of the pathways, we found no correlation between mutation burden and expression changes at the level of individual genes, i.e., case-to-normal ratio (CNR) and normalized mutation rate (nMR) values (Supplementary Figure S1). We also found no correlation at the pathway level—between the PAL and PI values, for all four pathway types (Figure 10).



**Figure 10.** Correlation of PI and PAL scores for different types of molecular pathways in 4910 tumor samples.

We then assessed correlations between all tumor samples for mutation burden at gene (nMR) and pathway (PI) levels, and also for expression changes at gene (CNR) and pathway (PAL) levels (Figure 11). To this end for every pair of tumor samples we calculated pairwise Spearman correlation coefficient, separately in all four pathway groups (Figure 11). For mutation signatures, the samples were highly heterogeneous with near-zero medians of Spearman correlation at the gene (nMR) and the pathway (PI) levels, except for slightly higher PI of signaling pathways (Spearman correlation 0.16). In contrast, for gene expression, the medians of Spearman correlation coefficients varied from 0.1 till 0.18 at the gene level (CNR), and much higher at the pathway level (PAL): from 0.12 till 0.52. These results evidence significantly greater inter-tumoral similarity at the level of pathway activation compared to the individual gene expression level and, especially, in comparison with both types of mutational data (Figure 11). These trends were also confirmed in thirteen analyses performed for all separate cancer types (Supplementary File S1). Interestingly, in most of the analyses, intertumoral activation profiles of the DNA repair pathways showed significantly higher congruences than it was observed for the other pathway types (Figure 11, Supplementary File S1).

Taken together, these data suggest that for both mutation and gene expression data, the intertumoral similarities are much higher at the pathway level (Figure 11, Supplementary File S1). On the other hand, DNA repair pathways demonstrate the most congruent activation patterns among the tumors.



**Figure 11.** Distribution of pairwise Spearman correlation coefficients between all possible combinations of tumor samples under investigation for mutation and expression data at both gene wise (nMR, CNR) and pathway (PAL, PI) level in cytoskeleton, DNA repair, metabolic and signaling pathway groups. Vertical dashed lines indicate the medians of the distributions.



### 3. Discussion

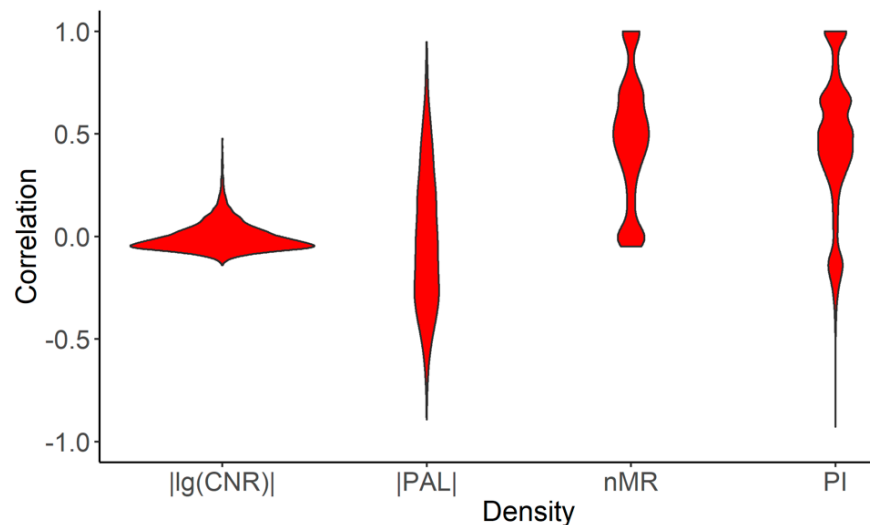
Juxtaposition of clinical and molecular tumor phenotypes provides a basis for further improvement of cancer treatments. High-throughput cancer gene expression and mutation data can help identifying new oncogenes and driver mutations [55]. In turn, further analysis of cancer data at the level of molecular pathways helps understanding pathological molecular changes in a quantitative way [50,56–58].

Nowadays, many molecular pathways were reported and collected in specific databases. Despite the existence of universal methods to pathway annotation like PathwayCommons, these databases were mostly generated separately and utilize different approaches to pathways nomenclature, and the same or very similar pathways can be included in different databases under different names [59]. Functional classification of the pathways also depends on a database, e.g., HumanCyc repository contains only metabolic pathways [45]. The structure and composition of the different pathway databases are continuously being revised in search for uniformity and comprehensiveness [59,60].

However, most studies of cancer biology focus on only single pathways or small groups of pathways [61,62]. For example, in 2018, a detailed study of the mechanisms and patterns of somatic alterations in 10 signaling pathways in 33 cancer types revealed patterns of co-occurrence and mutual exclusivity, driver changes, single and multiple potentially targeted mutations [22]. However, high-throughput simultaneous comparison of mutation and activation features of different types of molecular pathways was missing [63,64].

In this study, we performed more thorough analysis in terms of number of molecular pathways investigated. We explored 419 molecular pathways using new instruments to algorithmically assess pathway activation levels [23] and pathway mutation instability scores [24] by analyzing high throughput gene expression and whole exome sequencing data. To our knowledge, these metrics were never systemically investigated before for simultaneous large-scale characterization of molecular pathways in cancers. This enabled us to perform the comprehensive comparative characterization of four major functional groups of pathways in thirteen human cancer types. For the first time, for thirteen cancer primary localizations we investigated general mutation and activation features specifically for each type of the *signaling*, *cytoskeleton*, *metabolic* and *DNA repair* pathways.

We then compared our results with the previous study “Oncogenic Signaling Pathways in The Cancer Genome Atlas” on genetic alterations scoring in ten signaling pathways using TCGA data [22]. The pathways were ranked on the basis of specific alteration score, where genes in each individual tumor were defined as either altered or not altered. Altered genes had at least one of the following: copy-number alterations, mutations, fusions or specific features of epigenetic silencing. Consequently, altered pathway was defined as pathway having at least one altered gene. Alterations (binary alteration score, BAS) were marked as binary values. Therefore, binary alteration scores were calculated both at the gene and the pathway levels. Totally 9125 tumor samples of 38 tumor subtypes were investigated, among them 4382 were also investigated in our study. To compare the results obtained, for those overlapping 4382 tumor samples, we calculated our functional metrics (nMR, CNR, PAL, PI) for the same ten pathways that were investigated in the previous study [22]. Then we performed correlation analysis on gene and pathway levels between BAS from the previous study [22] and our metrics (Figure 12). We found positive correlations between BAS and our mutation metrics both on gene and pathway levels, but no correlation was found with the expression data (Figure 12).



**Figure 12.** Distribution of correlation coefficients calculated for 4382 tumor samples for ten signaling pathways from [22] and enclosed individual genes between: (i) binary alteration score (BAS) at the gene level with absolute values of  $\lg(\text{CNR})$  and normalized mutation rate (nMR); (ii) BAS at the pathway level and absolute values of pathway activation level (PAL) and pathway instability (PI).

We also compared our results with the trends previously revealed for BAS in different tumor subtypes. To this end, we calculated on the pathway level average BAS, PAL and PI scores per tumor subtype (Figure 13). At the gene level, we compared average BAS, CNR and nMR per tumor subtype (Supplementary File S2). Overall, we observed similarities for the mutational scores, but no common trends for the comparison with the expression data. Thus, previously reported complex binary alteration score partially resembles to mutation metrics like nMR and PI because BAS includes mutation data as compound, but it doesn't reflect changes at the expression level (Figures 12 and 13).

Totally, in our study we analyzed 4910 individual cancer gene expression and matching whole exome sequencing profiles collectively covering ~1.3 million mutations in thirteen cancer types. We found that for both mutation and gene expression data, the intertumoral similarities were much higher at the pathway level than at the level of individual genes. We identified common trends for the representatives of *signaling/cytoskeleton* and *metabolism/DNA repair* groups of pathways at the level of mutation data. Most importantly, the *signaling/cytoskeleton* group members were outstandingly enriched by the most highly mutated genes and deficient by the genes with the low mutation levels. This suggests their initiator roles in carcinogenesis and the auxiliary/supporting roles for the other groups of molecular pathways.

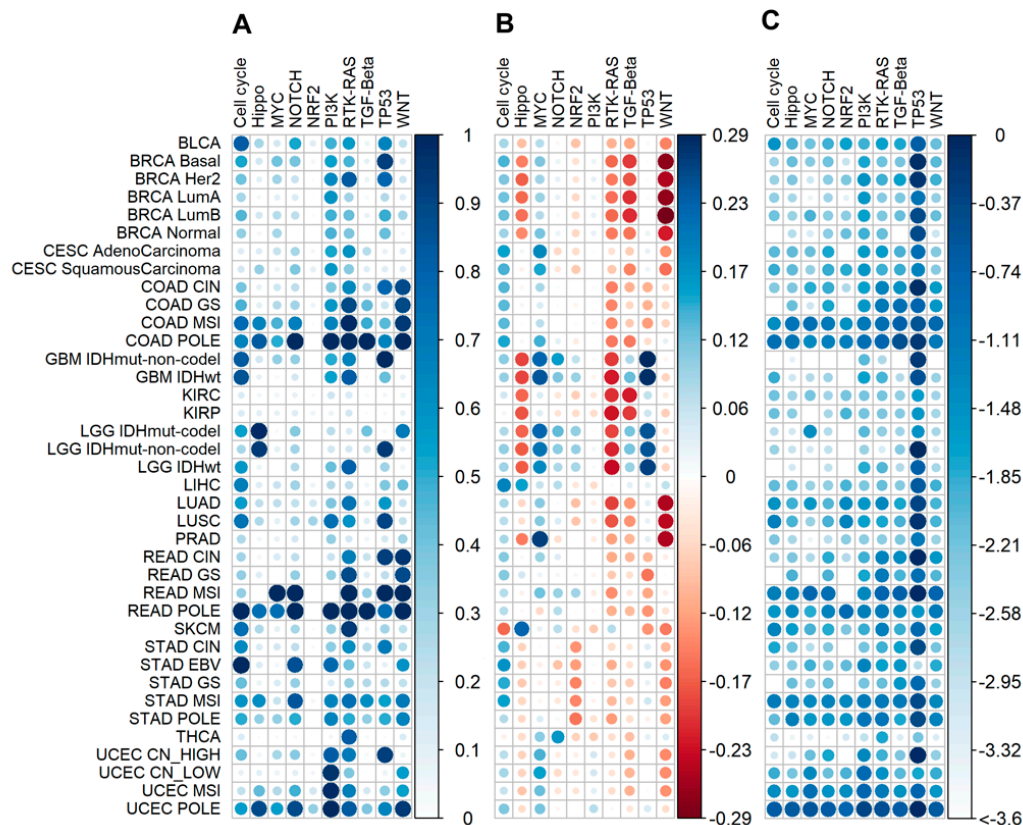
However, at the level of gene expression, the *DNA repair* group showed markedly upregulated activation levels in cancers, whereas for the other groups of pathways, the downregulated members prevailed. On the other hand, DNA repair pathways also demonstrated the most congruent activation patterns among all the tumor samples.

Our results also confirmed largely downregulated activities of the *metabolic* molecular pathways in cancer. In many reports, a focus was made on the metabolic alterations in tumors, so that cancer is even called a metabolic disorder [65–68]. We identified here multiple differentially regulated/mutated metabolic pathways in cancer, and several related biochemical processes were suggested by the gene ontology analyses. We found that metabolic pathways had the lowest mutational burden, and revealed that only the pathways related to nucleotide metabolism were significantly systemically up-regulated that argues some previous reports about hyperactivation of common metabolic background [65–68].

More specifically, the least mutated pathways dealt (*cytoskeleton*) with membrane organization, receptor-mediated endocytosis; (*DNA repair*) with regulation of DNA quality and associated DNA



repair, and with rearrangement of immunoglobulin loci; (*metabolic*) with biosynthesis of nucleotides and DNA replication; (*signaling*) with negative regulation of glucocorticoid receptor pathway, regulation of hormone levels, circadian rhythms, regulation of H3 histone deacetylation and protein catabolism.



**Figure 13.** Comparison of pathway functional metrics for 4382 tumor samples common for this and previous study [22] calculated for ten signaling pathways from [22]. (A) Average BAS (on pathway level) per tumor subtype. (B) Average PAL per tumor subtype. (C) Average lg(PI) per tumor subtype. Molecular tumor subtypes were referred according to [22]: BLCA—Urothelial bladder cancer; BRCA—Breast cancer; CESC—Cervical cancer; KIRC—Chromophobe renal cell carcinoma; KIRC—Clear cell kidney carcinoma; COAD—Colon adenocarcinoma; READ—Rectal adenocarcinoma; SKCM—Cutaneous melanoma; GBM—Glioblastoma multiforme; LIHC—Liver hepatocellular carcinoma; LGG—Lower Grade Glioma; LUAD—Lung adenocarcinoma; LUSC—Lung squamous cell carcinoma; KIRC—Papillary kidney carcinoma; PRAD—Prostate adenocarcinoma; STAD—Stomach adenocarcinoma; THCA—Papillary thyroid carcinoma; UCEC—Uterine corpus endometrial carcinoma; CIN—Chromosomal Instability; CN\_HIGH—copy-number high; CN\_LOW—copy-number low; EBV—Epstein-Barr Virus; GS—Genomically Stable; Her2—Her2-enriched; IDHwt—IDH1-wild-type; IDH mutant-codel—IDH mutant with codeletion of chromosome arm 1p and 19q; IDH mutant-non-codel—IDH mutant with euploid 1p/19q; LumA—Luminal A; LumB—Luminal B; MSI—Microsatellite Instability; POLE—polymerase  $\epsilon$  mutant subtype.

In turn, the most strongly mutated pathways were responsible for (*cytoskeleton*) exocytosis, cell cycle progression, processing and presentation of antigens via MHC class II; (*DNA repair*) cellular response to stress, chromosome organization, hydrolysis of phosphodiester bonds; (*metabolic*) drug response and catabolism, arachidonic acid secretion, fatty acid derivative pathways, glycosylation, sulfur and benzene-containing compound pathways, production of mitochondrial RNA, and tRNA aminoacylation for protein translation; (*signaling*) regulation through phosphatidylinositol, cell surface receptor pathways, growth and proliferation, transport of calcium, transcription by RNA

polymerase II, and regulation of immune cell activation. On the other hand, the most strongly cancer-downregulated pathways related to (*cytoskeleton*) movement of cell or subcellular components and locomotion, regulation of development and cell proliferation; (*metabolic*) cellular ketone body metabolism, neurotransmitter catabolic processes, sulfur pathways, drug response pathways; (*signaling*) transcription by RNA polymerase II, interleukin-7-mediated pathway.

Among the most strongly cancer-upregulated pathways, there were no *cytoskeleton* pathways, whereas the others dealt with (*DNA repair*) rearrangement of immune receptor loci, protein localization on chromosomes and chromosomal organization; (*metabolic*) oligosaccharide biosynthesis, nucleotide biosynthesis and DNA replication; (*signaling*) cell migration and adhesion, regulation of cell death, calcium ion transmembrane transport, extracellular matrix organization, regulation of mRNA stability, processing and presentation of exogenous peptide antigen via MHC class I, and activation of immune cells.

Finally, our study poses a challenge of analyzing the whole cancer interactome model where separate pathways would be connected into a single graph with known connectivity and functional relationships between its nodes. In the future, this type of analysis could be beneficial for personalized finding of causative cancer mutations and clinically actionable therapeutic molecular targets.

#### 4. Materials and Methods

##### 4.1. Mutation Data

DNA mutation data were extracted from the Catalogue Of Somatic Mutations In Cancer (COSMIC) project database [69]. We used genome-wide screen dataset, including whole exome and genome sequencing data, database version 76. The whole dataset included information for 19 434 tumor samples of different localizations obtained from different sources but here we considered only the samples obtained from The Cancer Genome Atlas (TCGA) project because of their uniform experimental and analytic pipeline [42]. We took only tumor samples that had the matching RNA sequencing data for the same tumor samples and also had corresponding RNA sequencing data for the normal samples of the same tissue type on Genomic Data Commons (GDC) data portal [70]. We extracted non-synonymous somatic mutations by selecting the following mutation types: “substitution-missense”, “deletion-frameshift”, “substitution-nonsense”, “insertion-in frame”, “deletion-in frame”, “insertion-frameshift”, “complex-deletion in frame”, “nonstop extension”, “complex”, “complex-compound substitution”, “complex-frameshift”, “complex-insertion in frame” (Supplementary Tables S10). The final dataset contained 1,252,669 mutation records for 19,608 genes from 4910 individual tumor samples. The tumor samples selected had the following primary localizations: bladder, breast, brain, cervix, kidney, colon and rectum, liver, lung, prostate, skin, stomach, thyroid and uterus (Supplementary Table S5). Every patient case corresponded to only one tumor sample. Individual mutation profiles are listed in Supplementary Table S11.

##### 4.2. Gene Expression Data

Gene expression profiles were obtained from GDC data portal [70]. We downloaded RNA sequencing data (HTSeq counts modification) for the tumor samples that also had somatic mutation profiles from COSMIC database [48]; the normal tissues were taken for the same tissue types as the tumor samples investigated. Totally, gene expression data were collected for 655 normal and 4910 tumor tissue samples (13 cancer types, Supplementary Table S5). Individual gene expression profile IDs are listed in Supplementary Table S12.

##### 4.3. Pathway Databases

The gene structures and molecular architectures of 419 intracellular pathways were extracted from the public databases Reactome [43], NCI Pathway Interaction Database [44], Kyoto Encyclopedia of Genes and Genomes [17], HumanCyc [45], Biocarta [46], Bio-rad [71], Qiagen [47] (Supplementary Table S1).

To increase statistical accuracy, we considered only 419 molecular pathways including ten or more genes. These pathways were classified into four functional groups by expert supervision: 278 signaling, 72 metabolic, 48 DNA repair and 47 cytoskeleton pathways (Supplementary Table S1; gene lists are available in Supplementary Table S2). 26 pathways were marked as simultaneously signaling and DNA repair pathways.

#### 4.4. Calculation of Mutation Frequency Metrics

*Pathway instability (PI)* reflects mutation enrichment of a molecular pathway [14]. *PI* calculation utilizes mutation frequencies of genes forming a pathway. To assess mutation frequencies of individual genes, we introduced *normalized mutation rate (nMR)* value expressed by the formula:

$$\begin{aligned} nMR_n &= \frac{N \text{ mut}(n, g) \times 1000}{N \text{ samples}(g) \times \text{Length CDS}(n)} nMR_n \\ &= \frac{N \text{ mut}(n, g) \times 1000}{N \text{ samples}(g) * \text{Length CDS}(n)} \end{aligned} \quad (1)$$

where  $nMR_n$  is *normalized mutation rate* of gene  $n$ ;  $N \text{ mut}(n, g)$  is total number of mutations for gene  $n$  in group of samples  $g$ ;  $N \text{ samples}(g)$  is number of samples in group  $g$ ;  $\text{Length CDS}(n)$  is length of coding DNA sequence (CDS) of gene  $n$  in nucleotides. *PI* levels were calculated as follows:

$$PI_p = \frac{\sum_n nMR_n PG_{p,n}}{N_p} \quad (2)$$

where  $PI_p$  is *PI* for pathway  $p$ ;  $nMR_n$  is as described above;  $PG_{p,n}$  is pathway-gene indicator that equals to one if gene  $n$  belongs to pathway  $p$ , or equals to zero if not;  $N_p$ —total number of gene products belonging to pathway  $p$ . The calculations were made using Amazon and Microsoft Azure cloud services.

#### 4.5. Calculation of Pathway Activation Level

*Pathway activation level (PAL)* characterizes cumulative changes of expression levels of genes belonging to a certain molecular pathway [12,23,50,72,73]. *PAL* is calculated as follows [23]:

$$PAL_p = \sum_n ARR_{np} * \lg(CNR_n) / \sum_n |ARR_{np}| \quad (3)$$

where  $PAL_p$  is *PAL* for pathway  $p$ ,  $CNR_n$  is *case-to-normal ratio*, the ratio of gene  $n$  expression level in a tumor sample under study to an average level for the control group; *ARR* (*activator/repressor role*) is Boolean flag that depends on function of gene  $n$  product in pathway  $p$ . *ARR* is  $-1$  if gene product  $n$  inhibits pathway  $p$ ;  $1$  if  $n$  activates pathway;  $0$  if  $n$  has ambiguous or unclear role in a pathway;  $0.5$  or  $-0.5$ , if  $n$  is rather activator of a pathway or its inhibitor, respectively. The calculations were made using Amazon and Microsoft Azure cloud services.

#### 4.6. Gene Ontology Analysis

We performed analysis of biological processes involving gene sets under investigation using GOrilla tool for Gene Ontology terms analysis [52]. We used modification of two unranked gene lists: *target dataset* was the gene set of interest and *background* was list of protein coding human genes from Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) database, version 20170717 [74],  $p$ -value threshold was set as  $10^{-3}$ .

Gorilla outputs were visualized using REVIGO viewer with the default settings (*allowed similarity* =  $0.7$ , semantic similarity measure was SimRel), except for dispensability cut-off level ( $0.25$ ).

We excluded from visualization the non-informative general terms like “biological processes” and irrelevant terms like “kidney morphogenesis” by expert curation.

#### 4.7. Comparison with the Study “Oncogenic Signaling Pathways in The Cancer Genome Atlas”.

We used Genomic Alteration matrixes of gene and pathway levels from the study “Oncogenic Signaling Pathways in The Cancer Genome Atlas” [22] to compare the results of this study. Genomic Alteration matrixes contained information about the following alterations: copy-number alterations, mutations, fusions or epigenetic silencing in the form of binary alteration score (BAS). If one or more alterations was in the gene in the tumor sample, BAS (gene level) for this gene and this tumor sample was “1”, otherwise—“0”. If one or more genes had BAS equal to one in the pathway in the tumor sample, BAS (pathway level) for this pathway and this tumor sample was “1”, otherwise—“0”. We used BAS for ten signaling pathways: (1) cell cycle, (2) Hippo signaling, (3) Myc signaling, (4) Notch signaling, (5) oxidative stress response/ Nuclear factor erythroid 2-related factor 2 (NRF2) pathway, (6) PI-3-Kinase signaling, (7) receptor-tyrosine kinase (RTK)/ Rat sarcoma (RAS) kinase/ Mitogen-Activated Protein (MAP) kinase signaling, (8) Transforming growth factor beta (TGF $\beta$ ) signaling, (9) P53 and (10)  $\beta$ -catenin/ Wntless-related integration site (WNT) signaling, and 186 enclosed genes [22].

The study [22] analyzed 9125 tumor samples of 64 tumor subtypes, among them 4382 (representing 38 tumor subtypes) were also included in our study. This dataset of 4382 samples of 38 tumor subtypes was used for the comparative analysis. Tumor sample barcodes and molecular subtypes are given in Supplementary Table S13. For ten above pathways, we calculated expression and mutation functional metrics: PAL, PI and CNR, nMR for 186 genes (Supplementary Table S14). Average BAS on gene level, BAS on pathway level, CNR, nMR, PAL and PI per tumor subtype were calculated as arithmetic mean for tumor samples of a tumor subtype under consideration.

## 5. Conclusions

Using high throughput mutation and gene expression data, we established pathway activation and mutation enrichment metrics of 278 signaling, 72 metabolic, 48 DNA repair and 47 cytoskeleton molecular pathways in 4910 individual cancer samples RNA sequencing and whole exome sequencing profiles representing thirteen cancer types. We found no correlation between mutation enrichment and expression changes both at the individual gene and molecular pathway levels. We observed generally congruent pathway activation profiles for different cancer types. The highest cancer-specific activation levels were seen in the DNA repair pathways, which had also the highest mutation enrichment levels. Moreover, the DNA repair pathways also demonstrated the most congruent activation patterns among all the tumor samples. However, the signaling and cytoskeleton pathways had the highest proportions of outstandingly frequently mutated genes thus suggesting their initiator roles in carcinogenesis and the supporting roles for the other groups of molecular pathways.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Table S1: Referenced list of molecular pathways investigated in this study, Table S2: Gene lists of pathway groups: signaling, metabolic, DNA repair, cytoskeleton pathways, Table S3: Averaged normalized mutation rates (nMRs) for 19 608 genes, for 4910 cancer samples, Table S4: Averaged normalized mutation rates (nMRs) for 19 608 genes, given separately for 13 cancer types, Table S5: Structure of dataset of 4910 cancer samples and 655 normal samples, IDs of expression and mutation data samples, Table S6: PAL scores for 4910 tumor samples, Table S7: PI scores for 4910 tumor samples, Table S8: Top and bottom 15% frequently mutated pathways and their gene compositions, Table S9: Top and bottom cancer-upregulated pathways and their gene compositions, Table S10: Mutation types in cancers, Table S11: Individual mutation profiles of 4910 tumor samples, Table S12: Links to individual gene expression profiles of 4910 cancer samples and 661 matched normal samples, Table S13: Barcodes and molecular subtypes of 4832 tumor samples, which were included in our study and in [22], Table S14: Expression and mutation functional metrics: PAL, PI for ten signaling pathways and CNR, nMR for 186 genes, for 4382 tumor samples. Supplementary File S1: Cancer type-specific distributions of pairwise Spearman correlation coefficients between all possible combinations of tumor samples under investigation for mutation

and expression data at both gene wise (nMR, CNR) and pathway (PAL, PI) level in cytoskeleton, DNA repair, metabolic and signaling pathway groups. Vertical dashed lines indicate the medians of the distributions, Supplementary File S2: Comparison of gene functional metrics for 4382 tumor samples common for this and previous study [22] calculated for 186 genes from [22]: Average BAS (on gene level), average CNR, average lg(nMR) per tumor subtype. Molecular tumor subtypes were referred according to [22]: BLCA—Urothelial bladder cancer; BRCA—Breast cancer; CESC—Cervical cancer; KICH—Chromophobe renal cell carcinoma; KIRC—Clear cell kidney carcinoma; COAD—Colon adenocarcinoma; READ—Rectal adenocarcinoma; SKCM—Cutaneous melanoma; GBM—Glioblastoma multiforme; LIHC—Liver hepatocellular carcinoma; LGG—Lower Grade Glioma; LUAD—Lung adenocarcinoma; LUSC—Lung squamous cell carcinoma; KIRP—Papillary kidney carcinoma; PRAD—Prostate adenocarcinoma; STAD—Stomach adenocarcinoma; THCA—Papillary thyroid carcinoma; UCEC—Uterine corpus endometrial carcinoma; CIN—Chromosomal Instability; CN\_HIGH—copy-number high; CN\_LOW—copy-number low; EBV—Epstein-Barr Virus; GS—Genomically Stable; Her2—Her2-enriched; IDHwt—IDH1-wild-type; IDH mutant-codel—IDH mutant with codeletion of chromosome arm 1p and 19q; IDH mutant-non-codel—IDH mutant with euploid 1p/19q; LumA—Luminal A; LumB—Luminal B; MSI—Microsatellite Instability; POLE—polymerase  $\epsilon$  mutant subtype. Supplementary Figure S1. Correlation of nMR and CNR values for 4910 cancer samples under investigation.

**Author Contributions:** Conceptualization, M.A.Z., V.S.T., A.P.S., S.A.R. and A.A.B.; Data curation, M.A.Z., V.S.T. and M.I.S.; Formal analysis, M.A.Z., V.S.T., D.V.K., D.E.K., M.I.S. and A.A.B.; Funding acquisition, A.A.B.; Investigation, M.A.Z., V.S.T., D.V.K., D.E.K. and M.I.S.; Methodology, M.A.Z., V.S.T., M.I.S. and A.A.B.; Software, M.A.Z. and V.S.T.; Supervision, S.A.R. and A.A.B.; Writing—original draft, M.A.Z., A.P.S. and M.I.S.; Writing—review & editing, S.A.R. and A.A.B. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** This work was supported by Amazon and Microsoft Azure grants for cloud-based computational facilities. We thank Oncobox/OmicsWay research program in machine learning and digital oncology for software and pathway databases for this study. Financial support was provided by the Russian Foundation for Basic Research Grant 19-29-01108.

**Conflicts of Interest:** The authors declare no conflict of interests.

## References

1. Sonnenschein, C.; Soto, A.M. Theories of carcinogenesis: An emerging perspective. *Semin. Cancer Biol.* **2008**, *18*, 372–377.
2. Vineis, P.; Schatzkin, A.; Potter, J.D. Models of carcinogenesis: An overview. *Carcinogenesis* **2010**, *31*, 1703–1709.
3. Bedessem, B.; Ruphy, S. SMT or TOFT? How the Two Main Theories of Carcinogenesis are Made (Artificially) Incompatible. *Acta Biotheor.* **2015**, *63*, 257–267.
4. Antontseva, E.V.; Matveeva, M.Y.; Bondar, N.P.; Kashina, E.V.; Leberfarb, E.Y.; Bryzgalov, L.O.; Gervas, P.A.; Ponomareva, A.A.; Cherdyntseva, N.V.; Orlov, Y.L.; et al. Regulatory single nucleotide polymorphisms at the beginning of intron 2 of the human KRAS gene. *J. Biosci.* **2015**, *40*, 873–883.
5. Fouad, Y.A.; Aanei, C. Revisiting the hallmarks of cancer. *Am. J. Cancer Res.* **2017**, *7*, 1016–1036.
6. Kulikov, A.V.; Luchkina, E.A.; Gogvadze, V.; Zhivotovsky, B. Mitophagy: Link to cancer development and therapy. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 432–439.
7. Sharma, S.; Kelly, T.K.; Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **2010**, *31*, 27–36.
8. 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.
9. Shtam, T.; Naryzhny, S.; Samsonov, R.; Karasik, D.; Mizgirev, I.; Kopylov, A.; Petrenko, E.; Zabrodskaya, Y.; Kamyshinsky, R.; Nikitin, D.; et al. Plasma exosomes stimulate breast cancer metastasis through surface interactions and activation of FAK signaling. *Breast Cancer Res. Treat.* **2019**, *174*, 129–141.
10. Petrov, I.; Suntsova, M.; Mutorova, O.; Sorokin, M.; Garazha, A.; Il'nitskaya, E.; Spirin, P.; Larin, S.; Kovalchuk, O.; Prassolov, V.; et al. Molecular pathway activation features of pediatric acute myeloid leukemia (AML) and acute lymphoblast leukemia (ALL) cells. *Aging* **2016**, *8*, 2936–2947.
11. Kurz, S.; Thieme, R.; Amberg, R.; Groth, M.; Jahnke, H.-G.; Pieroh, P.; Horn, L.-C.; Kolb, M.; Huse, K.; Platzer, M.; et al. The anti-tumorigenic activity of A2M-A lesson from the naked mole-rat. *PLoS ONE* **2017**, *12*, e0189514.

12. 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.
13. 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*, 1810–1823.
14. 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.* **2019**, *8*, 658.
15. von Eichborn, J.; Dunkel, M.; Gohlke, B.O.; Preissner, S.C.; Hoffmann, M.F.; Bauer, J.M.J.; Armstrong, J.D.; Schaefer, M.H.; Andrade-Navarro, M.A.; Le Novère, N.; et al. SynSysNet: Integration of experimental data on synaptic protein–protein interactions with drug–target relations. *Nucleic Acids Res.* **2012**, *41*, D834–D840.
16. Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C.A.; Holland, T.A.; Keseler, I.M.; Kothari, A.; Kubo, A.; et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2014**, *42*, D459–D471.
17. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
18. Bulashevskaya, S.; Szakacs, O.; Brors, B.; Eils, R.; Kovacs, G. Pathways of urothelial cancer progression suggested by Bayesian network analysis of allelotyping data. *Int. J. Cancer* **2004**, *110*, 850–856.
19. Haddad, D.; Socci, N.; Chen, C.-H.; Chen, N.G.; Zhang, Q.; Carpenter, S.G.; Mittra, A.; Szalay, A.A.; Fong, Y. Molecular network, pathway, and functional analysis of time-dependent gene changes associated with pancreatic cancer susceptibility to oncolytic vaccinia virotherapy. *Mol. Ther.-Oncolytics* **2016**, *3*, 16008.
20. Creixell, P.; Reimand, J.; Haider, S.; Wu, G.; Shibata, T.; Vazquez, M.; Mustonen, V.; Gonzalez-Perez, A.; Pearson, J.; Sander, C.; et al. Pathway and network analysis of cancer genomes. *Nat. Methods* **2015**, *12*, 615–621.
21. Zhang, Q.; Burdette, J.E.; Wang, J.-P. Integrative network analysis of TCGA data for ovarian cancer. *BMC Syst. Biol.* **2014**, *8*, 1338.
22. Sanchez-Vega, F.; Mina, M.; Armenia, J.; Chatila, W.K.; Luna, A.; La, K.C.; Dimitriadou, S.; Liu, D.L.; Kantheti, H.S.; Saghafein, S.; et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**, *173*, 321–337.e10.
23. Borisov, N.; Sorokin, M.; Garazha, A.; Buzdin, A. Quantitation of Molecular Pathway Activation Using RNA Sequencing Data. *Methods Mol. Biol.* **2020**, *2063*, 189–206.
24. Zolotovskaia, M.; Sorokin, M.; Garazha, A.; Borisov, N.; Buzdin, A. Molecular Pathway Analysis of Mutation Data for Biomarkers Discovery and Scoring of Target Cancer Drugs. *Methods Mol. Biol.* **2020**, *2063*, 207–234.
25. Dreesen, O.; Brivanlou, A.H. Signaling pathways in cancer and embryonic stem cells. *Stem. Cell Rev.* **2007**, *3*, 7–17.
26. Whittaker, S.; Marais, R.; Zhu, A.X. The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene* **2010**, *29*, 4989–5005.
27. Petrov, I.; Suntsova, M.; Ilnitskaya, E.; Roumiantsev, S.; Sorokin, M.; Garazha, A.; Spirin, P.; Lebedev, T.; Gaifullin, N.; Larin, S.; et al. Gene expression and molecular pathway activation signatures of MYCN-amplified neuroblastomas. *Oncotarget* **2017**, *8*, 83768–83780.
28. 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.
29. DeBerardinis, R.J.; Thompson, C.B. Cellular metabolism and disease: What do metabolic outliers teach us? *Cell* **2012**, *148*, 1132–1144.
30. Daye, D.; Wellen, K.E. Metabolic reprogramming in cancer: Unraveling the role of glutamine in tumorigenesis. *Semin. Cell Dev. Biol.* **2012**, *23*, 362–369.
31. Sciacovelli, M.; Gaude, E.; Hilvo, M.; Frezza, C. The Metabolic Alterations of Cancer Cells. In *Methods in Enzymology*; Galluzzi, L., Kroemer, G., Eds.; Academic Press, 2014; Volume 542, pp. 1–23.
32. Xu, X.D.; Shao, S.X.; Jiang, H.P.; Cao, Y.W.; Wang, Y.H.; Yang, X.C.; Wang, Y.L.; Wang, X.S.; Niu, H.T. Warburg effect or reverse Warburg effect? A review of cancer metabolism. *Oncol. Res. Treat.* **2015**, *38*, 117–122.
33. Dang, C. V Links between metabolism and cancer. *Genes Dev.* **2012**, *26*, 877–890.

34. Welf, E.S.; Haugh, J.M. Signaling pathways that control cell migration: Models and analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2011**, *3*, 231–240.
35. Karagiannis, T.C.; El-Osta, A. DNA damage repair and transcription. *Cell. Mol. Life Sci.* **2004**, *61*, 2137–2147.
36. Hall, A. The cytoskeleton and cancer. *Cancer Metastasis Rev.* **2009**, *28*, 5–14.
37. Jeggo, P.A.; Pearl, L.H.; Carr, A.M. DNA repair, genome stability and cancer: A historical perspective. *Nat. Rev. Cancer* **2016**, *16*, 35–42.
38. Galanos, P.; Pappas, G.; Polyzos, A.; Kotsinas, A.; Svolaki, I.; Giakoumakis, N.N.; Glytsou, C.; Pateras, I.S.; Swain, U.; Souliotis, V.L.; et al. Mutational signatures reveal the role of RAD52 in p53-independent p21-driven genomic instability. *Genome Biol.* **2018**, *19*, 37.
39. Dong, L.-Q.; Shi, Y.; Ma, L.-J.; Yang, L.-X.; Wang, X.-Y.; Zhang, S.; Wang, Z.-C.; Duan, M.; Zhang, Z.; Liu, L.-Z.; et al. Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma. *J. Hepatol.* **2018**, *69*, 89–98.
40. Turgeon, M.-O.; Perry, N.J.S.; Poulogiannis, G. DNA Damage, Repair, and Cancer Metabolism. *Front. Oncol.* **2018**, *8*, 15.
41. Cui, J.; Qu, Z.; Harata-Lee, Y.; Aung, T.N.; Shen, H.; Adelson, D. Cell Cycle, Energy Metabolism and DNA Repair Pathways in Cancer Cells are Suppressed by Compound Kushen Injection. *BMC Cancer* **2019**, *19*, 103.
42. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. (Poznan Poland)* **2015**, *19*, A68–A77.
43. Croft, D.; Mundo, A.F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M.R.; et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **2014**, *42*, D472–D477.
44. Schaefer, C.F.; Anthony, K.; Krupa, S.; Buchoff, J.; Day, M.; Hannay, T.; Buetow, K.H. PID: The Pathway Interaction Database. *Nucleic Acids Res.* **2009**, *37*, D674–D679.
45. Romero, P.; Wagg, J.; Green, M.L.; Kaiser, D.; Krummenacker, M.; Karp, P.D. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **2004**, *6*, R2.
46. Nishimura, D. BioCarta. *Biotech Softw. Internet Rep.* **2001**, *2*, 117–120.
47. QIAGEN-Pathway-Central. Available online: <https://www.qiagen.com/us/shop/genes-and-pathways/pathway-central/> (accessed on 19 September 2018).
48. Forbes, S.A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C.G.; Ward, S.; Dawson, E.; Ponting, L.; et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2017**, *45*, D777–D783.
49. Pita-Juárez, Y.; Altschuler, G.; Kariotis, S.; Wei, W.; Koler, K.; Green, C.; Tanzi, R.E.; Hide, W. The Pathway Coexpression Network: Revealing pathway relationships. *PLoS Comput. Biol.* **2018**, *14*, e1006042.
50. 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*, 110–124.
51. 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. *Mol. Case Stud.* **2019**, *5*, a003434.
52. 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.
53. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **2011**, *6*, e21800.
54. Huntley, R.P.; Sawford, T.; Mutowo-Meullenet, P.; Shypitsyna, A.; Bonilla, C.; Martin, M.J.; O'Donovan, C. The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* **2015**, *43*, D1057–D1063.
55. Bailey, M.H.; Tokheim, C.; Porta-Pardo, E.; Sengupta, S.; Bertrand, D.; Weerasinghe, A.; Colaprico, A.; Wendl, M.C.; Kim, J.; Reardon, B.; et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **2018**, *173*, 371–385.e18.
56. Ramamoorthi, G.; Sivalingam, N. Molecular mechanism of TGF- $\beta$  signaling pathway in colon carcinogenesis and status of curcumin as chemopreventive strategy. *Tumor Biol.* **2014**, *35*, 7295–7305.
57. Dong, T.; Zhang, Z.; Zhou, W.; Zhou, X.; Geng, C.; Chang, L.K.; Tian, X.; Liu, S. WNT10A/ $\beta$ -catenin pathway in tumorigenesis of papillary thyroid carcinoma. *Oncol. Rep.* **2017**, *38*, 1287–1294.



58. Buzdin, A.A.; Zhavoronkov, A.A.; Korzinkin, M.B.; Venkova, L.S.; Zenin, A.A.; Smirnov, P.Y.; Borisov, N.M. Oncofinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front. Genet.* **2014**, *5*, 55.
59. Chowdhury, S.; Sarkar, R.R. Comparison of human cell signaling pathway databases—evolution, drawbacks and challenges. *Database* **2015**, *2015*, doi:10.1093/database/bau126.
60. Wittig, U. Analysis and comparison of metabolic pathway databases. *Brief. Bioinform.* **2001**, *2*, 126–142.
61. McQuerry, J.A.; Jenkins, D.F.; Yost, S.E.; Zhang, Y.; Schmolze, D.; Johnson, W.E.; Yuan, Y.; Bild, A.H. Pathway activity profiling of growth factor receptor network and stemness pathways differentiates metaplastic breast cancer histological subtypes. *BMC Cancer* **2019**, *19*, 881.
62. Sever, R.; Brugge, J.S. Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* **2015**, *5*, a006098.
63. Cai, H.; Jing, C.; Chang, X.; Ding, D.; Han, T.; Yang, J.; Lu, Z.; Hu, X.; Liu, Z.; Wang, J.; et al. Mutational landscape of gastric cancer and clinical application of genomic profiling based on target next-generation sequencing. *J. Transl. Med.* **2019**, *17*, 189.
64. Vikova, V.; Jourdan, M.; Robert, N.; Requirand, G.; Boireau, S.; Bruyer, A.; Vincent, L.; Cartron, G.; Klein, B.; Elemento, O.; et al. Comprehensive characterization of the mutational landscape in multiple myeloma cell lines reveals potential drivers and pathways associated with tumor progression and drug resistance. *Theranostics* **2019**, *9*, 540–553.
65. Collier, H.A. Is cancer a metabolic disease? *Am. J. Pathol.* **2014**, *184*, 4–17.
66. Seyfried, T.N.; Flores, R.E.; Poff, A.M.; D'Agostino, D.P. Cancer as a metabolic disease: Implications for novel therapeutics. *Carcinogenesis* **2014**, *35*, 515–527.
67. Seyfried, T.N.; Shelton, L.M. Cancer as a metabolic disease. *Nutr. Metab. (Lond)* **2010**, *7*, 7.
68. Wishart, D.S. Is Cancer a Genetic Disease or a Metabolic Disease? *EBioMedicine* **2015**, *2*, 478–479.
69. Forbes, S.A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **2015**, *43*, D805–D811.
70. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A. V.; et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400–416.e11.
71. PrimePCR Pathways-260 pathway maps|Life Science Research|Bio-Rad. Available online: <https://www.bio-rad.com/en-us/prime-pcr-assays/pathway/primepcr-pathways> (accessed on 23 November 2019).
72. Zhu, Q.; Izumchenko, E.; Aliper, A.M.; Makarev, E.; Paz, K.; Buzdin, A.A.; Zhavoronkov, A.A.; Sidransky, D. Pathway activation strength is a novel independent prognostic biomarker for cetuximab sensitivity in colorectal cancer patients. *Hum. Genome Var.* **2015**, *2*, 15009.
73. 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.
74. Yates, B.; Braschi, B.; Gray, K.A.; Seal, R.L.; Tweedie, S.; Bruford, E.A. Genenames.org: The HGNC and VGNC resources in 2017. *Nucleic Acids Res.* **2017**, *45*, D619–D625.

