



Article

# Uncharacterized Proteins CxORFx: Subinteractome Analysis and Prognostic Significance in Cancers

Pavel Ershov <sup>\*</sup>, Evgeniy Yablokov, Yuri Mezentsev and Alexis Ivanov

Institute of Biomedical Chemistry, Moscow 119121, Russia; evgeniy.yablokov@ibmc.msk.ru (E.Y.); yu.mezentsev@gmail.com (Y.M.); alexei.ivanov@ibmc.msk.ru (A.I.)

\* Correspondence: pavel79@inbox.ru

**Abstract:** Functions of about 10% of all the proteins and their associations with diseases are poorly annotated or not annotated at all. Among these proteins, there is a group of uncharacterized chromosome-specific open-reading frame genes (CxORFx) from the ‘Tdark’ category. The aim of the work was to reveal associations of CxORFx gene expression and ORF proteins’ subinteractomes with cancer-driven cellular processes and molecular pathways. We performed systems biology and bioinformatic analysis of 219 differentially expressed CxORFx genes in cancers, an estimation of prognostic significance of novel transcriptomic signatures and analysis of subinteractome composition using several web servers (GEPIA2, KMplotter, ROC-plotter, TIMER, cBioPortal, DepMap, EnrichR, PepPSy, cProSite, WebGestalt, CancerGeneNet, PathwAX II and FunCoup). The subinteractome of each ORF protein was revealed using ten different data sources on physical protein–protein interactions (PPIs) to obtain representative datasets for the exploration of possible cellular functions of ORF proteins through a spectrum of neighboring annotated protein partners. A total of 42 out of 219 presumably cancer-associated ORF proteins and 30 cancer-dependent binary PPIs were found. Additionally, a bibliometric analysis of 204 publications allowed us to retrieve biomedical terms related to ORF genes. In spite of recent progress in functional studies of ORF genes, the current investigations aim at finding out the prognostic value of CxORFx expression patterns in cancers. The results obtained expand the understanding of the possible functions of the poorly annotated CxORFx in the cancer context.



**Citation:** Ershov, P.; Yablokov, E.; Mezentsev, Y.; Ivanov, A. Uncharacterized Proteins CxORFx: Subinteractome Analysis and Prognostic Significance in Cancers. *Int. J. Mol. Sci.* **2023**, *24*, 10190. <https://doi.org/10.3390/ijms241210190>

Academic Editor: Stanislav Naryzhny

Received: 2 May 2023

Revised: 25 May 2023

Accepted: 26 May 2023

Published: 15 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** uncharacterized proteins; open-reading frame; subinteractome; gene expression; cancers; tumors

## 1. Introduction

Elucidation of molecular mechanisms underlying the malignant transformation of cells is one of the most popular areas in biomedical investigations with preclinical and clinical implications. Aberrant functioning of signal transduction, regulatory and metabolic cascades significantly contribute to malignant transformation [1], therefore, many ways of pharmacological correction of molecular components of cascades are being under development [2–4]. In recent years, significant progress has been made in the identification of proteins, which are directly involved in oncogenic cascades and modulate them by means of protein–protein interactions (PPIs) and post-transcriptional regulation [5–8]. Thus, the exploration of a spectrum of novel cancer-associated proteins is a promising strategy in anticancer drug discovery.

Currently, up to 90% of predicted human proteins (19,467 proteins) have been detected with high-confidence proteomic methods [9,10] and annotated, while 10% of proteins have not yet been annotated or at least poorly annotated in structural, functional and disease-specific context [10,11]. Among these proteins, there is a group of uncharacterized chromosome-specific open-reading frame proteins (CxORFx or ORF proteins) [12–15]. The nomenclature of CxORFx genes (ORF genes), encoding ORF proteins, means a chromosome

number (Cx) and an open-reading frame number (ORFx). It is worth mentioning that not all mapped ORF genes are protein coding. Of particular interest are studies devoted to the elucidation of the possible involvement of uncharacterized ORF genes in cancers. Some ORF genes such as *C20orf27* [16] and *C19orf53* [17] have already been described in association with cell invasion and proliferation. *C8orf48* gene overexpression has led to a reduction of these biological processes in colorectal cancer cells [18]. *C15orf39* protein has been identified as a substrate for mitogen-activated protein kinase MAPK1 [19], which is frequently deregulated in cancers. Recently, Joshi and colleagues [20] showed that *C16orf87* protein interacts with epigenetic master regulator histone deacetylase 1 targeted by anticancer drugs. Additionally, there are data on gene expression (e.g., the Human Proteome Atlas) and protein–protein interactions patterns with the participation of ORF proteins. The above-mentioned facts point to a research relevance in functional and disease-specific characterization of ORF genes, which is still missing for the vast majority of them.

A group of 342 ORF genes, belonging to the ‘T-dark’ category, was retrieved from the PHAROS database. A ‘T-dark’ category includes genes (proteins) that were mentioned in publications and other evidence, but their functional roles and pharmacological targeting have not been fully studied or remain unknown. A list of 219 ORF genes that are differentially expressed in a wide range of cancers was selected. Then, we used systems biology and bioinformatic analysis to reveal associations between ORF genes and malignant transformation of cells as well as to predict the involvement of uncharacterized proteins in cellular functions and molecular cascades through a spectrum of neighboring annotated proteins that physically interact and form binary or higher-order complexes with them.

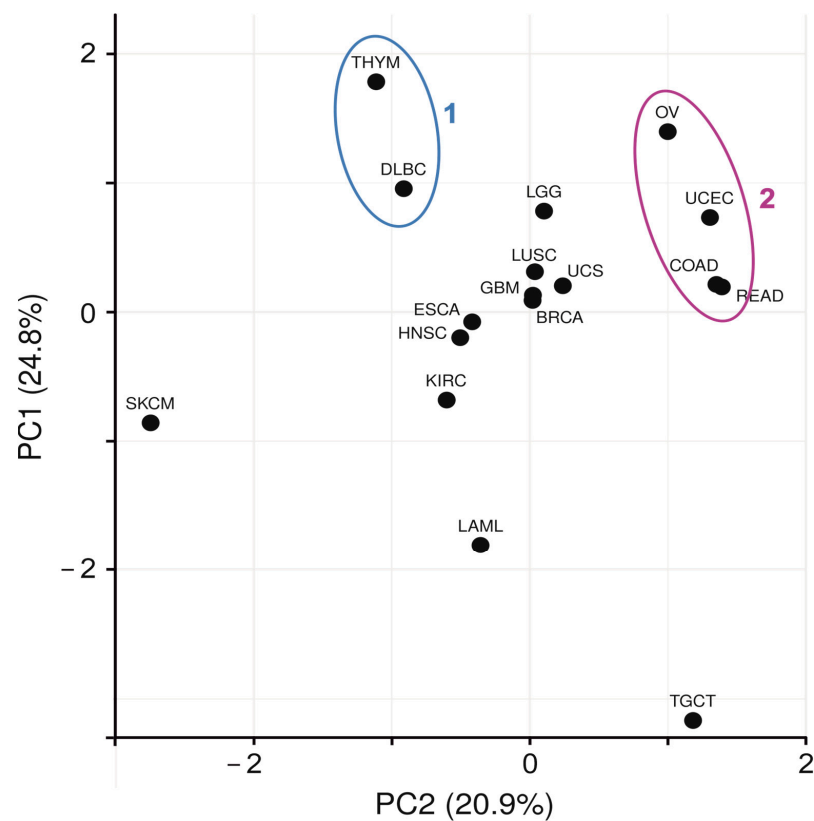
## 2. Results

### 2.1. ORF Genes: Expression Patterns and Their Prognostic Significance

#### 2.1.1. Gene Expression Patterns

For a panoramic view of 219 differentially expressed ORF genes, we used the principal component analysis, which made it possible to separate at least two cancer clusters (Figure 1) with similar expression patterns of ORF genes. One cluster includes 68 genes predominantly up-regulated in diffuse large B cell lymphoma (DLBC) and thymoma (THYM) (cluster 1). Cluster 2 includes four cancer types: 16 up-regulated genes in uterine corpus endometrial carcinoma (UCEC), 11 up-regulated genes in ovarian carcinoma (OV), 9 genes in colorectal adenocarcinoma (COAD) and 12 genes in rectum adenocarcinoma (READ). More specific expression patterns of ORF genes are characteristic of discrete (non-clustered) cancers. Thus, 124 out of 219 genes are down-regulated in testicular germ cell tumor (TGCT), except for one up-regulated (*C6orf132*), and 100 out of them are down-regulated in TGCT only (mentioned here, ‘TGCT-specific 100-gene expression signature’, Figure S1). Twenty-three ORF genes are up-regulated in acute myeloid leukemia (LAML) and ten genes are down-regulated in skin cutaneous melanoma (SKCM) (Figure 1).

There is also a number of ORF genes with different expression in  $\geq$  five cancers (mentioned here, ‘a pan-cancer group’) (Figure 2): preferentially down-regulated genes (*C11orf96*, *C5orf38* and *C8orf88*); preferentially up-regulated genes (*C19orf48*, *C1orf210*, *C4orf48* and *C6orf132*); mixed type of gene expression (*C1orf162*, *C2orf54*, *C2orf74* and *C14orf132*). Figure 2 shows a heat map of gene expression changes at mRNA and total protein levels. It can be seen that there is good consistency between the vectors of expression changes at both levels for *C11orf96*, *C6orf132* and *C8orf88* genes, in particular, in UCEC.



**Figure 1.** A panoramic view of clustered and non-clustered cancers according to the patterns of differentially expressed ORF genes in the coordinates of principle components PC2 and PC1. No scaling is applied to rows; singular value decomposition (SVD) with imputation was used to calculate PCs. X- and Y-axis show PC2 and PC1 that explain 20.9% and 24.8% of the total variance, respectively. Clusters 1 and 2 are highlighted in blue and purple, respectively.

Genes		TGCT	LAML	DLBC	THYM	SKCM	LUSC	LUAD	GBM	UCES	UCS	OV	BRCA	KICH	KIRC	KIRP	COAD	READ	PAAD	ESCA
C11orf96	mRNA																			
	Protein																			
C6orf132	mRNA																			
	Protein																			
C8orf88	mRNA																			
	Protein																			
C14orf132	mRNA																			
C19orf48	mRNA																			
C1orf162	mRNA																			
C1orf210	mRNA																			
C2orf54	mRNA																			
C2orf74	mRNA																			
C4orf48	mRNA																			
C5orf38	mRNA																			

**Figure 2.** Visualization of differentially expressed ORF genes in cancers (pan-cancer group). Up- and down-regulated genes are highlighted with yellow and violet colors, respectively. TGCT—testicular germ cell tumor, LAML—acute myeloid leukemia, DLBC—lymphoid neoplasm diffuse large B-cell lymphoma, THYM—thymoma, SKCM—skin cutaneous melanoma, LUSC—lung squamous cell carcinoma, LUAD—lung adenocarcinoma, GBM—glioblastoma multiforme, UCES—uterine corpus endometrial carcinoma, UCS—uterine carcinosarcoma, OV—ovarian serous cystadenocarcinoma, BRCA—breast invasive carcinoma, KICH—kidney chromophobe carcinoma, KIRC—kidney renal clear cell carcinoma, KIRP—kidney renal papillary cell carcinoma, COAD—colon adenocarcinoma, READ—rectum adenocarcinoma, PAAD—pancreatic adenocarcinoma, and ESCA—esophageal carcinoma.

### 2.1.2. Prognostic and Predictive Significance

The tissue-specific transcriptomic signatures are useful for the prognosis of cancer outcomes (survival rates and disease monitoring) as well as for the prediction of the immune and chemotherapy effectiveness. Prioritization of gene expression signatures with prognostic significance will contribute to the development of multigene transcriptomic panels for cancer molecular subtyping and a personalized approach to cancer treatment. Prognostic significance is often connected with the overall survival of patients (OS, the length of time from the date of diagnosis or the start of treatment for disease, during which patients with a diagnosed disease are still alive) and disease-free survival (DFS, the length of time after primary treatment for disease ends that the patient survives without any signs of disease). Kaplan–Meier analysis is a statistical method for estimating the survival curve with a log-rank test, which provides the comparison of groups with low and high levels of gene expression.

The potential prognostic significance is already known for 85 out of 219 ORF genes encoding chromosome-specific uncharacterized ORF proteins in 16 types of solid tumors (according to the Human Proteome Atlas portal) (Table S1).

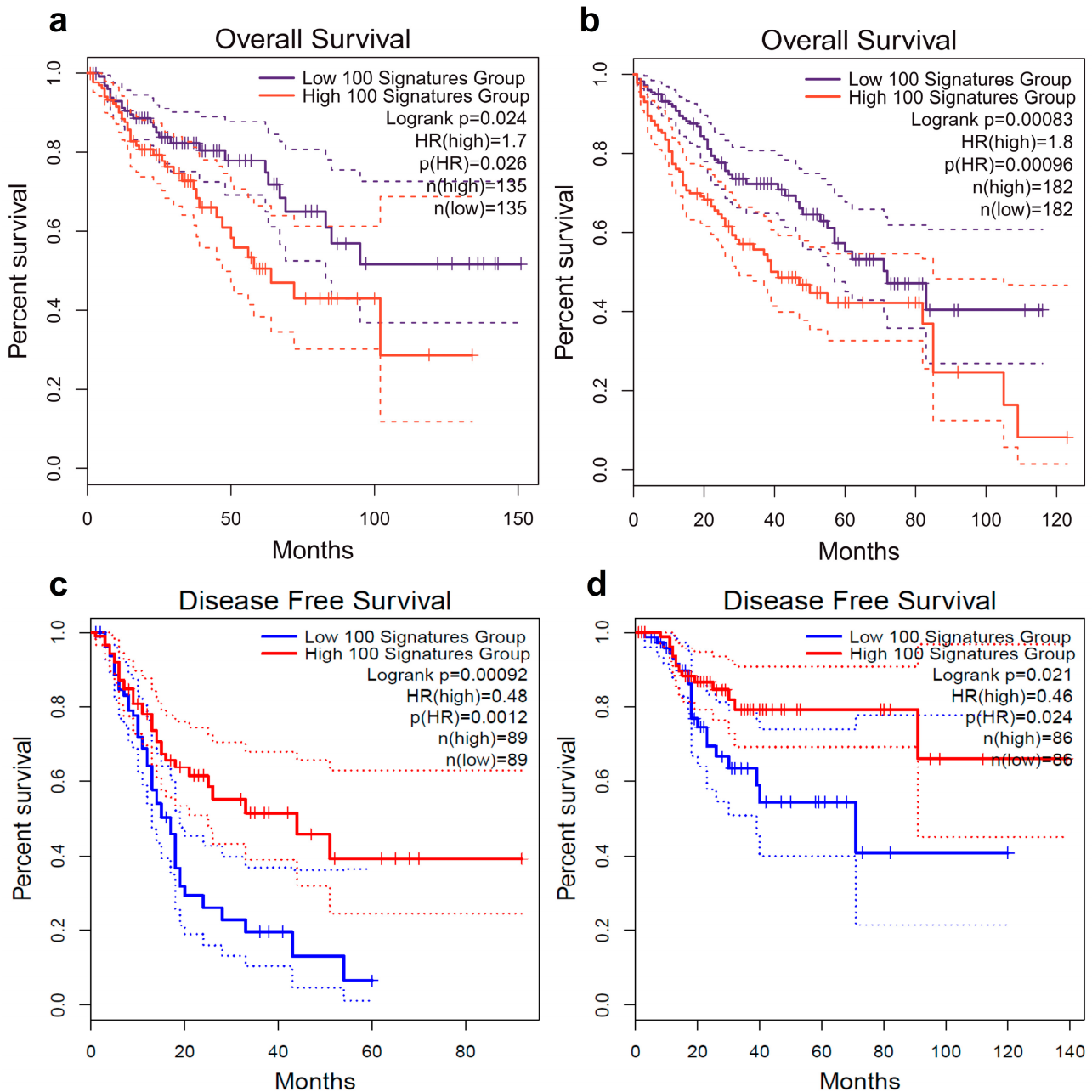
Endometrial, liver, pancreatic and renal cancers have the largest number of ORF genes with prognostic significance—20, 20, 14 and 43 genes, respectively. In addition, many genes have prognostic significance in more than one type of cancer. For example, a high level of *C3orf62* gene expression correlates with higher survival rates in patients with urothelial/gastric/pancreatic/breast, lung/head and neck cancers, while in the case of liver cancer, it correlates with lower survival rates.

We attempted to find novel transcriptomic signatures of differently expressed ORF genes in cancers using a Kaplan–Meier analysis of The Cancer Genome Atlas cohorts. It was shown that genes included in ‘TGCT-specific 100-gene expression signature’ have no comprehensive prognostic significance in the TGCT cohort, except for a ‘panel’ of six genes—*C11orf94*, *C1orf105*, *C20orf144*, *C22orf31*, *C9orf13* and *CXorf49*. The low levels of their expression correlate with better DFS of patients with TGCT (hazard ratio (HR) = 2.9 and *p*-value = 0.0047 for the high expression group, *n* = 136).

Interestingly, in colorectal adenocarcinoma (COAD) or liver hepatocellular carcinoma (LIHC) cohorts, the high levels of gene expression of a total amount of 100 genes correlate with poor OS (HR = 1.7 or 1.8, respectively, *p*-value < 0.05) (Figure 3a,b). Conversely, the low levels of gene expression in pancreatic adenocarcinoma (PAAD) and UCEC correlate with DFS (HR = 0.46–0.48, *p*-value < 0.05) (Figure 3c,d). It should note that with a more stringent cut-off HR-values ( $0.5 > \text{HR} > 2$ ), connections of expression patterns of total amount of 100 genes with DFS are kept in PAAD or UCEC only. Further, we determined that a ‘panel’ of seven genes (*C16orf78*, *C16orf86*, *C16orf96*, *C19orf18*, *C20orf144*, *C22orf42* and *C3orf62*) out of 100 genes make a decisive contribution since their exclusion from analysis leads to a significant decrease in prognostic significance of remaining 93 genes in PAAD or UCEC (*p*-value > 0.05). Moreover, it is found that the low levels of expression of all seven genes, at least in PAAD, correlate with unfavorable DFS (HR = 0.24, *p*-value = 0.0012) and OS (9.8 vs. 16.6 months, HR = 0.37, *p*-value = 0.000018).

Table 1 shows the potential prognostic significance of several ORF genes that allowed us to prioritize two novel transcriptomic gene signatures for PAAD and READ. First, the high levels of *C14orf119* and *C5orf46* gene expression (gene signature I) correlate with poor OS in PAAD (16.6 vs. 30.4 months, HR = 2.1 (1.37–3.18), *p*-value < 0.001). In addition, this gene signature was specific for LIHC cohort only (14.2 vs. 54.1 months, HR = 2.82 (1.69–4.72), *p*-value < 0.001) (a plot not shown). Second, the low levels of *C6orf132*, *C6orf222* and *C4orf19* gene expression (gene signature II) correlate with poor OS in READ (33.1 vs. 52.7 months, HR = 0.36 (0.16–0.8), *p*-value < 0.01) (a plot not shown). A similar HR value is observed in the kidney renal clear cell carcinoma cohort (KIRC) (HR = 0.44 (0.32–0.61), *p*-value < 0.001) (a plot not shown). Thus, transcriptomic gene signatures I and II can be considered sufficiently specific for these types of cancers with a clear difference in survival times almost twice and higher in low and high-expression groups. From Table 1, it also

follows that the expression levels of *C11orf52*, *C9orf116*, *C17orf51* and *C1orf53* genes in the UCEC cohort have prognostic significance, but median five-year survival values were not calculated, probably due to the small number of appropriate clinical cases.



**Figure 3.** Kaplan–Meier analysis of ‘TGCT-specific 100-gene signature’: (a)—colorectal cancer TCGA cohort (OS,  $n = 135$  cases), (b)—liver hepatocellular carcinoma TCGA cohort (OS,  $n = 182$  cases), (c)—pancreatic adenocarcinoma TCGA cohort (DFS,  $n = 89$ ) and (d)—uterine corpus endometrial carcinoma TCGA cohort (DFS,  $n = 86$ ). Original data output returned from GEPIA2 web-server. TCGA—The Cancer Genome Atlas.

**Table 1.** Prognostic significance of ORF genes expression patterns in different cancers \* (five-year overall survival rates).

Gene	Cancer	Hazard Ratio	Low Expression Group, Months **	High Expression Group, Months **	Log-Rank <i>p</i> -Value
<i>C14orf119</i>	PAAD	1.76 (1.16–2.68)	24.4	17.0	0.0069
<i>C5orf46</i>	PAAD	2.1 (1.27–3.46)	16.2	9.77	0.0031
<i>C6orf222</i>	READ	0.41 (0.19–0.89)	24.5	47.8	0.0210
<i>C4orf19</i>	READ	0.44 (0.21–0.96)	33.1	52.2	0.0350
<i>C6orf132</i>	READ	0.37 (0.17–0.8)	33.1	52.2	0.0083
<i>C11orf52</i>	UCEC	0.42 (0.24–0.75)	n/d	n/d	0.0023
<i>C9orf116</i>	UCEC	0.28 (0.14–0.58)	n/d	n/d	0.0003
<i>C17orf51</i>	UCEC	2.51 (1.49–4.34)	n/d	n/d	0.0006
<i>C1orf53</i>	UCEC	2.13 (1.32–3.42)	n/d	n/d	0.0014

\* according to predictions on the KMplotter server, URL: <https://kmplot.com/> (accessed on 1–20 March 2023) using pan-cancer RNAseq dataset; \*\* upper quartile survival; PAAD—pancreatic adenocarcinoma, READ—rectum adenocarcinoma, UCEC—uterine corpus endometrioid carcinoma.

Connections between ORF genes expression levels and immune cell infiltration in the corresponding cancers were also searched for. In general, overexpression of the *C4orf46*, *C9orf40*, *C17orf67* and *C21orf58* gene of thymoma correlates with tumor immune cell infiltration and gene expression levels of *C1orf162*, *C16orf54* and *CXorf21* correlate with immune cell infiltration in many primary and metastatic tumors (Table S2). A high positive correlation is observed between the up-regulation of some ORF genes and increased tumor immune cell infiltration, especially by dendritic cells.

Finally, connections between ORF gene expression in tumor tissues and responses to chemotherapy (predictive significance) were investigated. *C22orf42* gene expression levels in OV correlate with good responses to platinum-based drugs (fold-change and AUC values  $\approx 2$  and 0.7, respectively) (Figure S2).

### 2.1.3. Other Aspects of Differently Expressed ORF Genes

We prioritized a group of differentially expressed ORF genes in cancers, for which the occurrence of significant effects on the viability of a limited number of cell lines (“strongly selective genes” subgroup) or most cell lines (“common essential genes” subgroup) is directly dependent on gene knockouts or knockdowns (Table 2). Up-regulation of some ORF genes in cancers is consistent with activity changes of certain protein kinases that are encoded by driver genes and involved in signaling pathways responsible for cancer progression. The cancer-dependent regulation of *C12orf49* gene expression through nuclear transcription factor Y subunit alpha (NFYA) is not yet known, however, there is evidence of its role in the trans-activation of some cancer-promoting genes [21,22]. Transcription factor GATA3, which can bind to the *C8orf76* gene promoter, is expressed at an early stage of thymus development and participate in thymocyte differentiation [23] as well as in control lymphoid cell differentiation, being a tumor suppressor in B-cell lymphomagenesis. It also follows from Table 2 that at least *C1orf109* and *C8orf33* gene expression may depend on binding in their promotor regions of cAMP-responsive element binding protein 1 (CREB1). Since CREB1 functions in cancer signaling [24], together with data on the involvement of ABL1 and ERBB2 protein kinases in pathways in cancers, it suggests the presence of

upstream cancer-associated tissue-specific transcriptional regulation in DLBC and THYM cancers. Thus, an upstream regulation of the expression of ORF genes, which are critical for cell viability, can occur during malignant transformation of cells.

**Table 2.** Up-stream regulation of expression patterns of ORF genes, which are critical for cell viability.

Genes *	Up-Regulation of Gene Expression in Cancers	Protein Kinase ** Perturbation Correlates with up-Regulation of ORF Gene Expression	Transcription Factors	Kinase Involvement in Cancer-Associated Signaling Pathways #
Common essential genes ***				
<i>C1orf109</i>	THYM	ABL1, ERBB2, FGFR1	CREB1	Pathways in cancer
<i>C1orf131</i>	DLBC, THYM	ERBB2, MAPK1	n/d	ErbB receptor tyrosine kinase; gastric cancer, endometrial cancer, lung cancer
<i>C9orf16</i>	OV	AKT1, FGFR1	n/d	Pathways in cancer, proteoglycans in cancer
<i>C12orf45</i>	DLBC, THYM	AKT1, FGFR2	n/d	Pathways in cancer, PI3K-Akt (phosphoinositide-3-kinase, AKT and mTOR kinases), MAPK (mitogen-activated protein kinase)
<i>C17orf58</i>	DLBC, THYM	SYK	n/d	PI3K-Akt
<i>C19orf53</i>	DLBC, THYM	ATM, JAK2, MET, PDGFRA	n/d	Pathways in cancer, microRNAs in cancer
Strongly selective genes ***				
<i>C1orf112</i>	DLBC, THYM	ABL1, AKT1, JAK1, JAK2	n/d	Pathways in cancer, microRNAs in cancer
<i>C1orf21</i>	LAML, PAAD, THYM	AKT1, ERBB3, MAP2K1, MET, TGFB2	n/d	MAPK, proteoglycans in cancer
<i>C8orf33</i>	DLBC, THYM	ABL1, ERBB2, MET, PDGFRA	CREB1	Pathways in cancer, microRNAs in cancer
<i>C8orf76</i>	DLBC, THYM	CDK4, MAP2K4, MAPK1	GATA2	Kaposi sarcoma-associated herpesvirus infection, Human T-cell leukemia virus 1 infection
<i>C9orf40</i>	DLBC, THYM	ATM, BTK, MET, SYK	n/d	NF-kappa B signaling pathway
<i>C12orf49</i>	DLBC, PAAD	ATM, FGFR3	NFYA, USF1	n/d

\* Protein-coding genes only; \*\* genes containing mutations that are causally implicated in cancer (according to the cancer Gene Census); \*\*\* according to the DepMap portal; # pathway mapping was performed at the Pathway II server, URL: <https://pathway.sbc.su.se/> (accessed on 15–28 February 2023) using the KEGG database v.94.1 as a source. n/d—not determined.

## 2.2. ORF Proteins

Evidence on a protein level is found for 200 out of 219 ORF genes according to immunohistochemistry and/or mass-spectrometry data. Nineteen genes (*C1orf14*, *C1orf53*, *C1orf115*, *C2orf81*, *C4orf45*, *C5orf66*, *C6orf48*, *C6orf99*, *C7orf71*, *C8orf49*, *C10orf25*, *C11orf94*, *C11orf98*, *C16orf82*, *C16orf90*, *C16orf197*, *C19orf67*, *C20orf141* and *C20orf173*) are expressed at a transcript level only, pointing to probable ‘missing’ proteins. For example, the functionally

significant product of the *C8orf49* gene is a long non-coding RNA [25,26]. The distribution of ORF proteins by molecular weights (Mw) is shown in Figure S3. The major part (25%) includes proteins with Mw from 20 to 30 kDa. In total, micro-proteins with Mw from five to 20 kDa are almost 38%. It is known that micro-proteins possess a number of beneficial biological activities, for example, C5orf46 protein possesses antimicrobial activity [27]. It is also interesting that the micro-protein C19orf48 is widely expressed in cancers and can be processed for cytotoxic T lymphocyte (CTL) recognition [28].

Data on 237 subcellular localization sites were retrieved from The Human Protein Atlas for 135 out of 200 ORF proteins. Two major groups of proteins with nuclear localization sites (nucleoplasm, nucleolus, nuclear membrane and nuclear bodies) and cytoplasmic sites make up about 41% and 18%, respectively (Figure S4). It can also be noted that seven (C1orf54, C4orf48, C5orf46, C17orf67, C17orf99, C21orf62 and C22orf15) out of 135 proteins are predicted to be secreted. Using the Exocarta server [29,30], seven other secreted ORF proteins are additionally identified: exosome-derived proteins C2orf16, C2orf74, C11orf52, C16orf87 and C19orf18 (from urine samples); C1orf198 and C2orf88 proteins (from blood samples). C2orf74 and C19orf18 proteins appear to be membrane associated since they are predicted to contain typical alpha helices for anchoring in the hydrophobic membrane bilayer. Thus, at least 14 secreted ORF proteins may be involved in intercellular communication and signaling functions. We did not observe data on these secreted ORF proteins as potential biomarkers.

### 2.3. Protein–Protein Interactions of ORF Proteins and Their Subinteractomes

#### 2.3.1. Structural Data

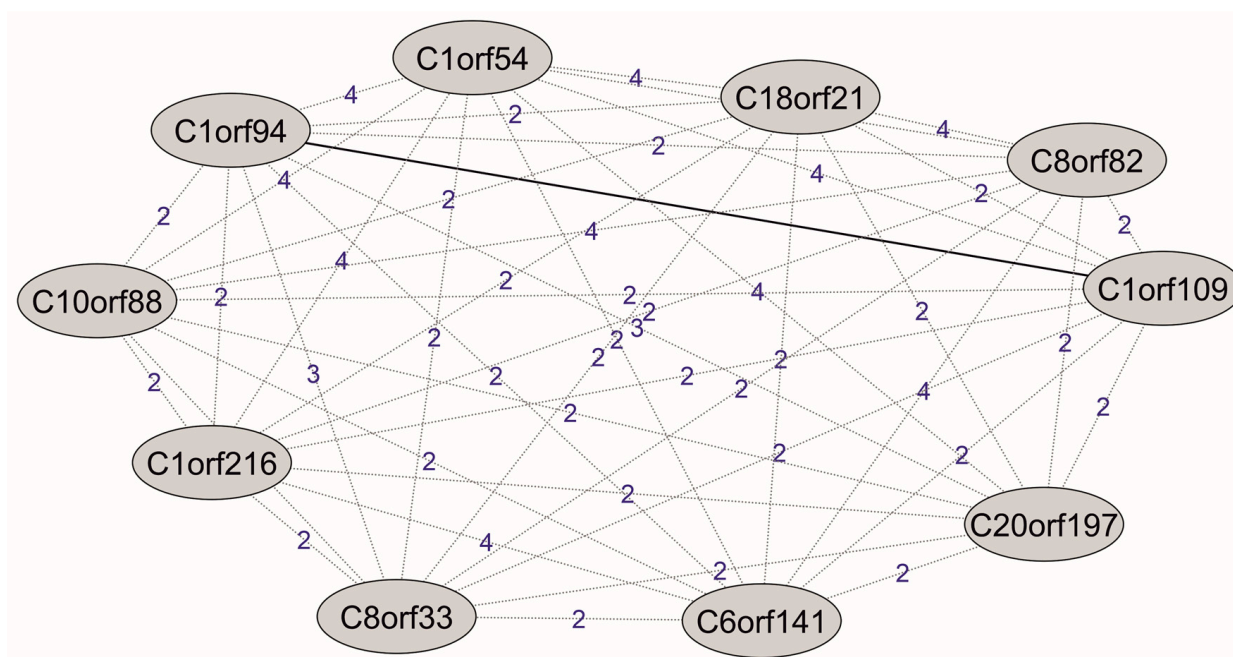
There is still very little information on the protein 3D structures among 219 differentially expressed ORF genes in cancers. The Protein Data Bank, URL: <https://www.rcsb.org/> (accessed on 15 February 2023) contains two crystallographic 3D-models of C1orf123 monomeric protein (PDB ID: 5zlq) [31] and C9orf64 dimeric protein (PDB ID: 7ugk) [32]. Using the interactive tool PDBePISA v.1.52 for analysis of protein–protein interfaces, URL: [https://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) (accessed on 17 February 2023), we analyzed the last model with resolution of 1.78 Å. Despite the relatively small interface area (6% of the total surface area), it is enriched with 13 hydrogen bonds and 10 salt bridges. Interestingly, the dimeric form of C9orf64 protein, according to CavityPlus server [33]) has two predicted ligand-binding pockets on the surface with a high level of drugability (drug scores are 884 and 2237) compared to monomeric form with one ligand-binding pocket (drug score = 281). This implies the higher potential of C9orf64 protein to pharmacological targeting its dimeric form than monomeric one in case of its valid clinical significance. Protein C2orf76 is described as a dimeric protein [34] that is involved in the formation of an aggressive pancreatic cancer phenotype through induction of cell invasion and proliferation. Protein C2orf76 can be considered as a potential drug target, but there is no crystallographic data for in silico drug design.

#### 2.3.2. Subinteractomes of ORF Proteins

The term ‘subinteractome’ means here a set of all the known protein partners that physically interact with a target protein. What is known about the intermolecular interactions with the participation of target ORF proteins? Protein C1orf123 specifically interacts with the heavy-metal-associated domain of a copper chaperone for superoxide dismutase encoded by the CCS gene [31]. Participation of another protein C16orf62 in the formation of a functionally significant heterotrimer with DSCR3 and VPS29 proteins is shown in [35]. Protein C11orf98 can be a putative interaction partner of nucleophosmin (NPM1) and nucleolin (NCL) [36]. Analysis of structural features and the PPIs spectrum of the C11orf96 protein points to the fact that it may be phosphorylated and plays a role in endoplasmic reticulum stress, protein ubiquitination and gene transcription [37]. Protein C1orf112 is probably cancer-associated as it follows from the functional analysis of its 31 protein partners and co-expression data [38]. Thus, uncharacterized ORF proteins can participate in

homo and heterodimeric protein interactions and deciphering their subinteractomes helps to predict functional roles through analysis of the PPIs spectrum.

A number of interactomic browsers (BioPlex, HIPPIE, MINT, hu.MAP, IID, InnateDB and HuRI), as well as interactomic datasets from the publications (see Section 4 in more details), were used to collect information about the subinteractomes of each ORF protein. In general, 4317 unique PPIs for 177 ORF proteins were found, of which eight PPIs are between the ORF proteins themselves. These are eight binary interactions: C1orf94 and C1orf109; C1orf112 and C11orf87; C1orf52 and CXorf56; C2orf74 and C20orf27; C5orf24 and C16orf71; C5orf51 and C10orf82; C1orf131 and C16orf87; C1orf131 and C20orf197. As for the rest of the 42 ORF proteins, there is a lack of any information about their PPIs. There are 517 protein partners forming binary interactions with two different ORF proteins (potential heterotrimeric complexes) and 768 PPIs, in which one protein partner interacts with three or more ORF proteins (potential multimeric complexes). It indicates a wide PPI subnetwork with connections between ORF proteins through common protein partners and the shortest paths. The last ones are among potential multimeric complexes, where ORF proteins can indirectly interact with the first-, second- and higher-order protein partners (Figure 4).



**Figure 4.** Visualization of the shortest paths between ORF proteins in the PPI subnetwork. The solid line indicates the direct interaction between ORF proteins. The dotted lines indicate the paths between ORF proteins through a number of other interacting proteins in the subnetwork.

As a rule, subinteractome of a target protein includes protein partners that form mainly stable PPIs with it, which is due to the peculiarities of experimental approaches used for isolation and identification of protein partners from biosamples. At the same time, it is not completely known what proportion of protein partners form direct interactions with a target protein. Therefore, analysis of subinteractomes was carried out on the assumption that each protein partner has an equal probability to directly interact with a target ORF protein. The distribution of a number of protein partners of ORF proteins within 177 subinteractomes are as follows: <7 protein partners (63 ORF proteins), from 7 to 15 (30 ORF proteins), from 15 to 30 (40 ORF proteins), from 30 to 70 (27 ORF proteins), from 70 to 100 (6 ORF proteins) and over 100 protein partners (11 ORF proteins). So, subinteractomes of 103 ORF proteins, each containing from 7 to 100 protein partners, are chosen for analysis. The choice of the minimal number of protein partners is due to the

sufficiency of functional enrichment analysis. The maximal number is limited in order to reduce the redundancy of output functional terms.

Protein partners included in the subinteractome of any ORF protein are analyzed for direct or indirect associations with ‘cancer hallmarks’. The ‘cancer hallmark’ score (CH-score), which means here the relative frequency of occurrence of CH-associated proteins per 10 protein partners in the subinteractome of an ORF protein, is calculated. For example, CH-score equal to one means that only one CH-associated protein is present among 10 protein partners, including itself. Next, 70 out of 103 subinteractomes are prioritized according to CH-score value  $\geq 2$ , which corresponds to approximately the 25th percentile at a median CH-score value equal to 2.8. Such subinteractomes consist of CH-associated proteins (a ‘core’) and other non-CH-associated protein partners. The results of molecular pathway enrichment analysis of ‘core’ proteins and PPIs cluster modeling as well as tissue-specific co-expression analysis are shown in Table S3. At least 42 subinteractomes of ORF proteins are enriched with molecular pathways in cancers, form highly connected PPIs clusters of CH-associated proteins and have noticeable or high correlation of co-expression between ORF genes and genes, encoding CH-associated proteins (Table S3). These facts indirectly speak in favor of the involvement of ORF proteins in signaling cascades and regulatory networks, which are critical for the malignant transformation of cells.

High correlation coefficients ( $r$ -values  $\geq 0.85$ ) of gene co-expression may additionally hint at the presence of direct interactions between ORF proteins and their protein partners. As it follows from Table S3, among such tissue- and cancer-specific PPIs, there are the following hits: C2orf88 and PRKAR1A (brain tissue), C5orf24 and WDR45B (thymoma), C5orf24 and GOPC (thymoma), C6orf132 and CDH1 (colon transverse), C6orf132 and EZR (colon), CXorf56 and RPRD1B (thymoma) and C16orf87 and ANAPC10 (thymoma). In this regard, it is interesting to find potential binary PPIs, whose gene co-expression patterns are highly cancer dependent. These PPIs can be predicted by the change of correlation coefficients from positive to negative values in cancers relative to normal tissue and vice versa. Thus, a group of 30 potential tissue-specific binary PPIs, involving ORF proteins and CH-associated protein partners, is shown in Table 3. It follows that 23% of PPIs belong to the subgroup of cancer-dependent PPIs in cancers (positive  $r$ -values of gene co-expression). This subgroup is represented with PPIs between ORF proteins and known oncoproteins, for example, C1orf131/JUN (transcription factors), C5orf24/PBX1 (pre-B-cell leukemia transcription factor 1), C2orf88/PRKACA (cAMP-dependent protein kinase catalytic subunit alpha), C6orf132/CDH1 (fizzy-related protein homolog) and C6orf132/EZR (tumor suppressor ezrin). Ezrin, being an adapter protein between the actin cytoskeleton and the plasma membrane, is involved in cancer promotion through the modulation of signaling pathways [39,40]. However, data on the involvement of ezrin in TGCT cancer is absent. Thus, we demonstrated that the interaction of a number of ORF proteins with oncoproteins are cancer-dependent PPIs.

Another subgroup of PPIs presumably perturbed by cancers is characterized by negative or near-zero  $r$ -values in cancers and positive  $r$ -values of gene co-expression in conditionally normal tissues, for example, C1orf131/EIF3L, C19orf53/APEX1, C1orf198/PPP2R1A and C1orf198/ARHGEF1 (Table 3). It should be noted that a significant part of protein partners of ORF proteins in this subgroup are protein-kinases and E3-ubiquitin-protein ligases that function as universal modulators of the posttranslational activity of protein substrates and modulators of the cellular GTPase system. Impaired interaction of C19orf44 protein with the testis-specific multifunctional enzyme GAPDHS (glyceraldehyde-3-phosphate dehydrogenase, 44 PPIs in the BioGRID database,  $\geq 2$  evidence) may be associated with modulation of spermatogenesis and male fertility [41] or metabolic reprogramming in cancers via GAPDH [42]. Thus, using the gene co-expression analysis, we demonstrate that a number of ORF proteins may participate in cancer-dependent PPIs or PPIs perturbed by cancers.

**Table 3.** Potential cancer-associated PPIs involving ORF proteins and CH-associated protein partners.

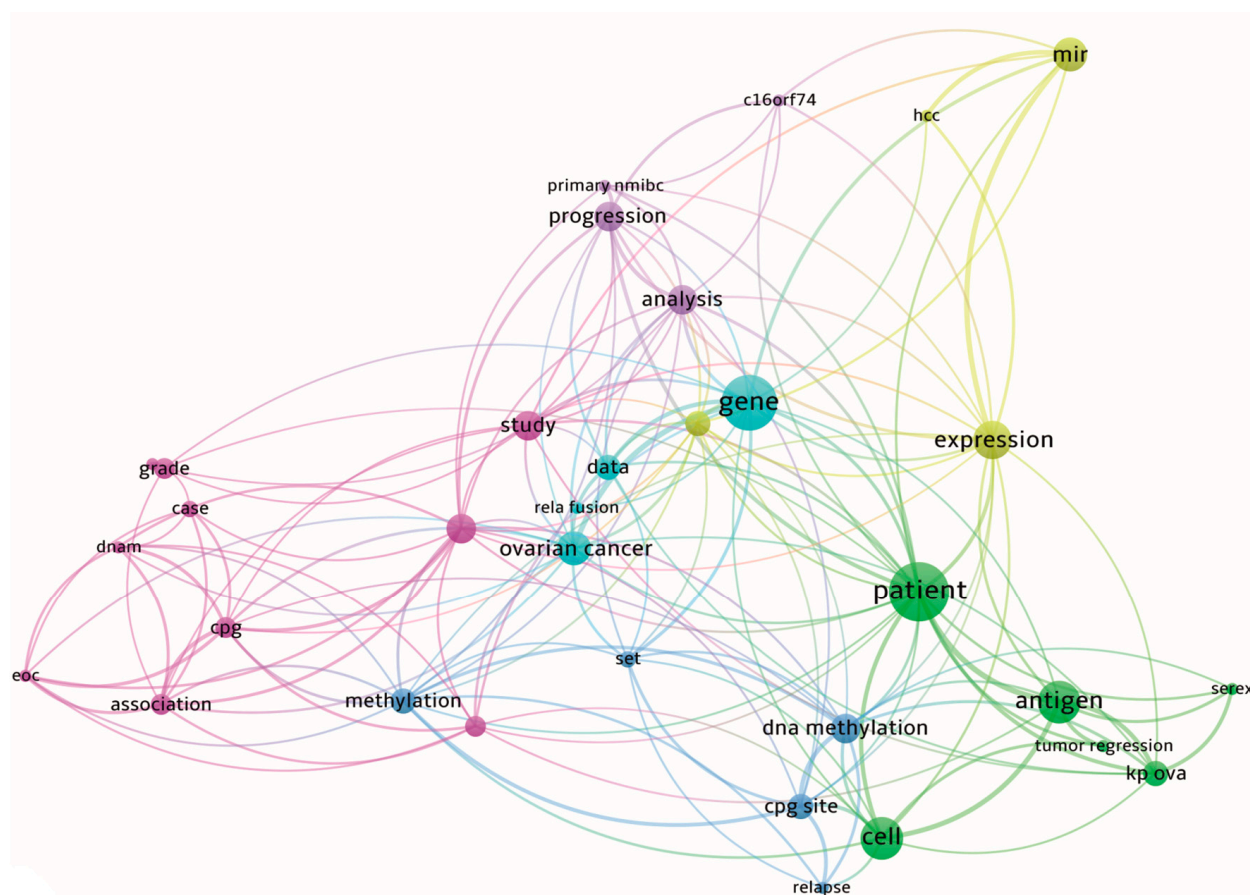
Protein–Protein Interaction		Gene Co-Expression Data *		The Main Function Related to Protein Partner ***
ORF-Protein	Protein Partner **	Normal Tissue	Cancer Tissue	
C1orf131	EIF3L	Whole blood (0.68)	DLBC (−0.51)	Initiation of protein synthesis Transcriptional master regulator Cellular response to oxidative stress
C1orf131	JUN	Whole blood (−0.07)	DLBC (0.61)	
C19orf53	APEX1	Whole blood (0.87)	DLBC (−0.66)	
C1orf226	BCR	Brain (0.53–0.59)	LGG (−0.06)	Regulatory activities toward small GTP-binding proteins Component of E3-ubiquitin-protein ligase complex Dephosphorylation Regulatory subunit of the cAMP-dependent protein kinases
C1orf226	BTRC	Brain (0.60–0.76)	LGG (−0.09)	
C1orf226	PTPRJ	Brain (0.53–0.80)	LGG (0.06)	
C2orf88	PRKAR1A	Brain (0.61–0.85)	LGG (0.03)	
C6orf132	CDH1	Kidney (0.82)	KIRP (0.04)	Cell adhesion
C1orf131	PPP2R1A	Whole blood (0.74)	THYM (−0.6)	Phosphorylation GTPase-activating protein Transcription factor Ubiquitination DNA helicase activity Nuclear protein import
C1orf198	ARHGEF1	Whole blood (0.61)	THYM (−0.61)	
C5orf24	PBX1	Whole blood (0.04)	THYM (0.7)	
C11orf98	WWP2	Whole blood (0.75)	THYM (−0.55)	
C12orf45	SMARCD1	Whole blood (0.73)	THYM (−0.58)	
C12orf45	KPNA1	Whole blood (0.57)	THYM (−0.50)	
C1orf115	MDM2	Testis (0.51)	TGCT (−0.41)	Ubiquitination of p53/TP53 Chromosome segregation and progression into anaphase Component of ubiquitin-protein ligase complex Regulation of Shh signaling Phosphorylation Connections of cytoskeletal structures to the plasma membrane Maintenance of the Golgi complex Potent inhibitor of cell death Calcium signal transduction pathway Transcriptional regulation Ubiquitination Activation of HECT domain-containing E3-ubiquitin-protein ligases ATP-dependent degradation of ubiquitinated proteins Nuclear pore formation Regulation of the switch between different pathways for energy production
C2orf16	SPAG5	Testis (0.68)	TGCT (−0.24)	
C2orf88	FBXW8	Testis (−0.68)	TGCT (0.54)	
C2orf88	GPR161	Testis (−0.64)	TGCT (0.58)	
C2orf88	PRKACA	Testis (−0.10)	TGCT (0.53)	
C6orf132	EZR	Testis (−0.28)	TGCT (0.71)	
C8orf48	OPTN	Testis (0.67)	TGCT (−0.06)	
C10orf67	BCL2L1	Testis (0.63)	TGCT (−0.22)	
C11orf65	CALM3	Testis (0.81)	TGCT (−0.06)	
C16orf71	MEIS2	Testis (−0.41)	TGCT (0.52)	
C16orf90	RNF123	Testis (0.55)	TGCT (−0.23)	
C16orf90	NDFIP2	Testis (0.65)	TGCT (0.03)	
C17orf47	ADRM1	Testis (0.65)	TGCT (−0.12)	
C17orf47	NUP214	Testis (0.80)	TGCT (−0.01)	
C19orf44	GAPDHS	Testis (0.83)	TGCT (−0.08)	
C6orf132	CDH1 (P12830)	Uterus (0.08)	UCS (0.76)	Calcium-dependent cell adhesion

\* Uniprot ID, URL: <https://www.uniprot.org/> (accessed on 1 April 2023), \*\* Spearman correlation coefficient according to the GEPIA2 server, URL: <http://gepia2.cancer-pku.cn/#correlation> (accessed on 1 April 2023), \*\*\* UniprotKB, URL: <https://www.uniprot.org> (accessed on 1 April 2023). Uniprot IDs of protein partners of ORF proteins: ADRM1 (Q16186), APEX1 (P27695), ARHGEF1 (Q92888), BCL2L1 (Q07817), BCR (P11274), BTRC (Q9Y297), CALM3 (P0DP25), CDH1 (P12830), EIF3L (Q9Y262), EZR (P15311), FBXW8 (Q8N3Y1), GAPDHS (O14556), GPR161 (Q8N6U8), JUN (P05412), KPNA1 (P52294), MDM2 (Q00987), MEIS2 (O14770), NDFIP2 (Q9NV92), NUP214 (P35658), OPTN (Q96CV9), PBX1 (P40424), PPP2R1A (P30153), PRKACA (P17612), PRKAR1A (P10644), PTPRJ (Q12913), RNF123 (Q5XPI4), SMARCD1 (Q9H4L7), SPAG5 (Q96R06) and WWP2 (O00308).

### 3. Discussion

Elucidation of cellular roles of uncharacterized chromosome-specific ORF proteins is the focus of many scientific investigations. We found at least 27 publications (see Table 4) devoted to the functional annotation of ORF genes mainly through gene knockdowns or knockouts and other experimental approaches. In Table 3, it follows that there is a large

biodiversity of established or proposed cellular roles of 20 ORF proteins: involvement in signaling pathways and processes such as ciliary motility and ciliogenesis, protein folding and degradation, lipid homeostasis, cell cycle regulation, protein trafficking, mitophagy and apoptosis (Table 4). To explore the publication landscape on 219 target ORF genes or ORF proteins in the cancer context, a bibliometric analysis was performed using the VosViewer software v.1.6.18. A full-text search allows us to select 204 publications, in which 92 out of 219 differently expressed ORF genes are mentioned in association with the malignant transformation of cells. Thus, *C9orf50*, *C9orf64*, *C5orf66*, *C16orf74* and *C10orf55* genes are most mentioned in the cancer context (each gene was mentioned at least in six publications), while 42 out of 92 genes are mentioned in one publication only. A co-occurrence map of 33 terms retrieved from the abstracts and titles of 204 found publications is shown in Figure 5, where these terms are divided into six thematic clusters, containing such terms as ‘patient survival’, ‘staging’, ‘progression’, ‘recurrence’ and ‘predictive significance’. In most publications, ORF genes are described mainly as a part of transcriptomic or methylation signatures correlating with disease prognosis. In a much smaller number of publications, specific cancer-associated features of some ORF genes were studied, for example, the capability to form gene fusions [43], participation in the PPI subnetwork as hub genes [44] and elucidation of upstream regulators (e.g., oncogene micro-RNA miR-556 for *C8orf48* gene [18]). Thus, the bibliometric analysis makes it possible to observe any associations with cancers for half of the target ORF genes mentioned in at least one publication. The conclusion is that the current vector of investigations of uncharacterized chromosome-specific ORF genes is to identify their prognostic significance.



**Figure 5.** A co-occurrence map of 33 terms retrieved from abstracts and titles of publications. Clusters of terms are highlighted with different colors.

**Table 4.** Relevant literature data on the known functional characterization of differentially expressed ORF genes in cancers.

Gene	Thesis	Reference
<i>C1orf54</i>	<i>C1orf54</i> gene-knockout mice exhibit impaired tubular epithelial cell proliferation and delayed recovery after kidney ischemia-reperfusion injury, which lead to deteriorated renal function and increased mortality.	[45]
<i>C1orf109</i>	<i>C1orf109</i> gene controls the late step of human pre-60S maturation in the cytoplasm and the loss of <i>C1orf109</i> global protein synthesis.	[46]
<i>C1orf115</i>	<i>C1orf115</i> protein modulates drug efflux through regulation of the major drug exporter ABCB1/MDR1 by means of physical association with it.	[47]
	CRISPR-Cas9 knockout screens reveal that the loss of <i>RDD1</i> ( <i>C1orf115</i> ) resulted in resistance to five anti-cancer drugs.	[48]
<i>C1orf131</i>	A protein, encoded by <i>C1orf131</i> gene, probably takes part in regulation of rRNA turnover.	[49]
<i>C1orf194</i>	The loss of normal <i>C1orf194</i> protein altered intracellular Ca <sup>2+</sup> homeostasis and up-regulated Ca <sup>2+</sup> handling regulatory proteins.	[50]
<i>C3orf70</i>	<i>C3orf70</i> gene is involved in neural and neurobehavioral development; defects in <i>C3orf70</i> may be associated with neurodevelopmental and neuropsychiatric disorders.	[51]
<i>C5orf51</i>	<i>C5orf51</i> protein interacts with <i>C5orf51</i> interacts with MON1 and CCZ1, members of the RAB7A guanine nucleotide exchange factor (GEF) complex. <i>C5orf51</i> is involved in mitophagy (selective degradation of mitochondria by autophagy).	[52]
<i>C6orf120</i>	<i>C6orf120</i> gene induces apoptosis of CD4 <sup>+</sup> T-cells mediated with endoplasmic reticulum stress	[53]
	Rats with deficiency of <i>C6orf120</i> gene are susceptible to liver injury induced by tetra-chloromethane.	[54]
<i>C8orf76</i>	<i>C8orf76</i> protein binds to the promoter region of <i>SLC7A11</i> gene and up-regulates <i>SLC7A11</i> . <i>C8orf76</i> down-regulation induces G1-S arrest and inhibition of cell proliferation.	[55]
<i>C6orf203</i>	Knockout of <i>C6orf203</i> leads to a decrease in mitochondrial translation and consequent OXPHOS deficiency. <i>C6orf203</i> involves translation via RNA binding without changing the stability of mtRNAs.	[56,57]
<i>C11orf42</i>	<i>C11orf42</i> protein could potentially be related to trafficking, as it is a part of a complex with 30% similarity to the retromer complex (SNX1, SNX2, VPS35, VPS29 and VPS26A), i.e., with 0.3 Jaccard similarity to the known CORUM retromer complex.	[58]
<i>C11orf53</i>	<i>C11orf53</i> protein interacts with POU2F3 and depletion of <i>C11orf53</i> reduced enhancer H3K27ac levels and chromatin accessibility, resulting in a reduction of POU2F3-dependent gene expression.	[59,60]
<i>C11orf88</i>	<i>C11orf88</i> protein is required for motile ciliogenesis and flagellar genesis in vertebrates by mediating the maturation of glycolytic enzyme ENO4	[61]
<i>C11orf94</i>	Deletion of <i>C11orf94</i> gene dramatically decreases male fertility in mice and plays a critical role in sperm binding to the oolemma.	[62]
	Loss of <i>C11orf94</i> gene results in reduces assembly of Izumo1 complexes and male infertility due to impaired gamete fusion.	[63]
<i>C12orf4</i>	<i>C12orf4</i> protein is a cytoplasmic protein implicated in the early signaling events following FcεRI-induced cell activation (FcεRI—the high-affinity IgE receptor).	[64]
	From strong in silico evidence, it follows that <i>C12orf4</i> protein, as a putative ADPR “eraser,” regulates PARP-mediated ADP-ribosylation signaling.	[65]
<i>C12orf45</i>	<i>C12orf45</i> protein acts as a bridge between NOP58 ribonucleoprotein and the particle for arrangement of quaternary structure (PAQosome), a multi-subunit chaperone complex. <i>C12orf45</i> interacts with NOP58 and RUVBL1/2 AAA <sup>+</sup> ATPases.	[66]
<i>C12orf49</i>	<i>C12orf49</i> gene participates in regulation of exogenous lipid uptake through modulation of SREBF2 signaling in response to lipid starvation.	[67]
	<i>C12orf49</i> protein mediates site-1 protease activation, which is essential for the cleavages of protease substrates in lipid homeostasis.	[68]
<i>C12orf73</i>	<i>C12orf73</i> protein interacts with newly synthesized cytochrome b to support initial steps of complex III biogenesis in complex with UQCC1 and UQCC2.	[69]
<i>C16orf62</i>	Heterotrimer (DSCR3, <i>C16orf62</i> and VPS29), interacting with other protein complexes, takes part in prevention of lysosomal degradation and promotion of cell surface recycling of α5β1 integrin.	[35]
<i>C16orf71</i>	Complex <i>C16orf71</i> /Daap1 is identified as a novel axonemal dynein regulator that is critical for ciliary motility.	[19]

At the same time, cellular roles remain unknown for many ORF proteins under both physiological and pathological conditions, so one of the ways to predict them is

to analyze the already-known functions of protein partners that physically interact with uncharacterized ORF proteins. Functional terms, describing subinteractomes of ORF proteins, are shown in Table S4. For example, the protein partners of secreted C2orf88 protein are related to post-translational modifications mediated through protein kinase A as well as important cellular functions such as ciliary motility. C4orf17 protein's function may be associated with chromatin modifications through histone methylation, which follows from the high representation of ATRX, CHD3, KAT2B, KDM1A, PRMT1 and SUV39H1 proteins in the C4orf17's subinteractome. Protein CXorf56 along with its interacting partner C1orf52 may be involved in the formation of spliceosome complex. The vast majority of interacting partners of C16orf90 protein are responsible for post-translational modifications (neddylation-dependent protein degradation (BIRC2, NDFIP1, NDFIP2 and XIAP) and ubiquitin-dependent protein degradation (BIRC2, RNF123, TRAF2 and XIAP). In this regard, the C16orf90 protein may act as a part of protein-modifying enzymes complex. Protein C11orf98 is involved in the regulation of DNA-binding transcription factor activity that stems from the spectrum of its protein partners (ESR1, ESR2, FOXA1, JUN and WWP2). On the other hand, functional coupling of subinteractomes of different ORF proteins may indicate identical GO terms (Table S4): 'GO:0000151~ubiquitin ligase complex' (C16orf87 and C6orf222), 'GO:0016197~endosomal transport' (C11orf49 and C1orf210), 'GO:0016311~dephosphorylation' (C1orf21, C2orf74 and C20orf27), 'GO:0016607~nuclear speck' (C1orf226 and C19orf53), 'WP254~apoptosis' (C10orf67 and C16orf90), 'PA443358~aneuploidy' (C1orf226, C6orf222 and C16orf87). The most common term 'GO:1901987~regulation of cell cycle phase transition' is typical for subinteractomes of C16orf87, C3orf62, C16orf87, CXorf56, C11orf49 and C16orf87 proteins.

It is important to find out whether ORF proteins are associated with signaling pathways in normal and pathological conditions. The separate over-representation analysis is performed for each subinteractome of ORF protein. It follows from Table S4 that there are a variety of functional terms covering  $\geq 25\%$  of the total amount of protein partners in the subinteractome. The subinteractome of C1orf21 protein is enriched with the 'WP3932~focal adhesion-PI3K-Akt-mTOR-signaling pathway' term. This pathway connects the membrane receptor and actin cytoskeleton through a cascade of protein kinases and phosphatases and affects the cell shape and motility. In neoplastic cells, aberrant activation of this pathway leads to the formation of a drug-resistant cancer phenotype [70,71]. An analysis of 79 candidate protein partners of C11orf52 points to its involvement in the Wnt-signaling pathway through desmoglein 1 (DSG1) [72]. In this regard, the subinteractome of C11orf52 protein contains HRAS ('Harvey RA Sarcoma virus oncogene'), which is one of the most mutated genes in solid tumors and SRSF protein kinase 1 (SRPK1), which was identified as part of HRAS signaling [73]. Thus, both C1orf21 and C11orf52 proteins can theoretically act as adapter proteins or modulators of these signaling pathways.

There is an opportunity for ORF proteins to interact with each other. Assumptions on binary interactions between C2orf74 and C20orf27, C5orf24 and C16orf71 proteins are strengthened by some interesting findings. First, the subinteractomes of these ORF proteins contain proteins associated with 'cancer hallmark' and oncoproteins (e.g., the case of C5orf24 and its protein partners XPO1, PBX1, MYC, CIC, GOPC, CREB1 and STK11, Table S3). Second, a comparative analysis of subinteractomes indicates the existence of 29 and 14 common protein partners for potential interactions C2orf74/C20orf27 and C5orf24/C16orf71, respectively. Over-representation analysis shows the related GO terms for both ORF proteins in each pair: 'GO:0016311~dephosphorylation' and 'GO:1903293~phosphatase complex' (C2orf74 and C20orf27) as well as 'PA444750~leukemia' and 'PA444761~leukemia, myeloid' (C5orf24 and C16orf71) (Table S4). As for another potential interaction C1orf131/C16orf87, only histone acetyltransferase KAT5 [74] is a common protein partner that can post-translationally modify both ORF proteins. The subinteractome of C16orf87 protein contains HDAC1 and HDAC2 histone deacetylases as well as 16 different histones, thus indicating functional associations of C16orf87 protein with epigenetic modulation of chromatin. C1orf131 gene belongs to a subgroup of

common essential genes (Table 2) and its subinteractome is over-represented by the term ‘GO:0022613~ribonucleoprotein complex biogenesis’ covering 11 out of 43 protein partners (25%) (Table S4).

Another point is that genetic alterations (GAs) more frequently occur in cancer-associated genes. Analysis of the cBioPortal pan-cancer cohort shows that the median GA frequency for a sample of 159 out of 219 differentially expressed ORF genes is about 1.3%. Gene amplifications are the prevalent type of GAs. It is interesting to note that GA frequency is up to 6% in *C3orf70*, *C8orf33*, *C8orf76* and *C8orf82* genes according to the most affected cancer types. By the way, an average proportion of GAs for known oncogenes (e.g., *KIT*, *MYC*, *CTNNB1*, *MDM2* and *APC*) in the same pan-cancer cohort ranges from four to 9%. At least the function of genetically unstable *C8orf76* is associated with cell proliferation [55], so amplification of the *C8orf76* gene in cancers may be an independent risk factor of enhanced cell proliferation. According to the Cancer Gene Census portal, only the *C15orf65* gene contains driver somatic mutations in primary mediastinal B-cell and Hodgkin’s lymphoma. Currently, the ClinVar database, URL: <https://www.ncbi.nlm.nih.gov/clinvar/> (accessed on 23 March 2023), aggregates records of missense mutations in these ORF genes with ‘uncertain’ or ‘likely benign’ clinical significance. A search for novel genetic variants of *C3orf70*, *C8orf33*, *C8orf76* and *C8orf82* associated with systemic diseases, including cancers, is of practical interest for medical genomics.

Identified gene expression signatures involving chromosome-specific ORF genes with potential prognostic and/or predictive significance can be further translated into multi-gene transcriptomic panels to predict disease dynamics or tumor responses to therapeutics. There is plenty of patents disclosing such panels, e.g., WO2021164492A1 (*C10orf99*), WO2018097166A1 (*C17orf67*) and WO2020226333A1 (*C14orf45*, *C16orf61*, *C7orf63*, *C10orf76* and *C12orf72*) in the Espacenet database (<https://worldwide.espacenet.com>). In this study, two highly specific gene expression signatures (*C14orf119* and *C5orf46*) and (*C16orf78*, *C16orf86*, *C16orf96*, *C19orf18*, *C20orf144*, *C22orf42* and *C3orf62*) were found for PAAD that are associated with the overall survival of patients and disease-free survival, respectively. Therefore, verification of gene expression signatures in a larger cohort of patients (balanced by subgroups with different cancer stages, grades, mutation burden, etc.) will provide the design of a clinically relevant transcriptomic panel for molecular subtyping of one of the highly aggressive solid tumors.

We focused on fourteen secreted ORF proteins. Peptide compositions of, at least, seven proteins (*C4orf48*, *C5orf46*, *C22orf15*, *C2orf16*, *C11orf52*, *C1orf198* and *C2orf88*) were identified by mass spectrometry in human blood plasma or urine samples according to the Peptide Atlas portal, URL: <https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/> (accessed on 21 May 2023). So far, the diagnostic significance of these proteins in the ‘liquid biopsy’ paradigm for cancer detection has not been studied, which also follows from the materials of the collaborative program Early Detection Research Network (EDRN) [75], and the above mentioned secreted ORF proteins may represent novel molecular entities in this scientific field.

Among other biological features of some uncharacterized chromosome-specific ORF proteins is the high representation of oncoproteins (encoded by highly mutated genes in cancers) in their subinteractomes. This fact may indirectly indicate the involvement of ORF proteins in cancer-associated signaling pathways in yet unknown roles. These include eight ORF proteins: *CXorf56* (*BRD4*, *EZH2*, *JUN*, *LMNA*, *MYC*, *NSD2*, *RECQL4*, *RET*); *C8orf48* (*CCNE1*, *FGFR3*, *HRAS* and *TSC1*); *C6orf132* (*CDH1*, *ESR1* and *EZR*), *C5orf24* (*CIC*, *CREB1*, *GOPC*, *MYC*, *PBX1*, *STK11* and *XPO1*); *C1orf226* (*BARD1*, *BCR*, *BRD3*, *FOXO3*, *HRAS* and *PRCC*); *C16orf71* (*CDKN2C*, *CREB1*, *EZH2*, *PBX1*, *PTEN*); *C11orf98* (*FOXA1*, *JUN* and *MYC*); *C1orf123* (*CDKN1A*, *CLTC*, *HEY1* and *RAF1*); *C1orf131* (*CDC73*, *JUN*, *KRAS*, *PPP2R1A* and *SMARCB1*). Thus, functional annotation of these ORF proteins via the expansion of data on the tissue-specific and cancer-specific spectrum of protein–protein interactions will clarify their roles in signaling pathways and, possibly, allow us to select new molecular targets for pharmacological interventions.

## 4. Materials and Methods

### 4.1. Gene Expression Data Analysis

PHAROS database v.3.15, a web-based interface for exploring and visualizing the Target Central Resource Database (TCRD) data generated by the Illuminating the Druggable Genome project (IDG) [76], was used for retrieval of ORF gene list.

Web-based server Gene Expression Profiling Interactive Analysis (GEPIA2) [77] adapted for the analysis of The Cancer Genome Atlas (TCGA) cohorts was used for selection of differentially expressed genes (DE-genes) in cancer and normal tissues according to the following settings:  $|\log_2FC| = 2$  (four-fold change),  $p$ -value cut-off  $< 0.05$ , match TCGA normal and GTEx (Gene Tissue Expression) datasets.

Survival analysis (Kaplan-Meier plots) was carried out using the web-server KMplotter [78].

ROCplotter server [79] was used to explore the gene expression patterns and response to therapy of patients with glioblastoma, breast, ovarian and colorectal cancers. AUC (Area Under the Curve) and gene expression fold change values equal to 0.7 and 2.0 were set as cut-off values, respectively.

TIMER (Tumor IMMune Estimation Resource) server is a comprehensive resource for systematical analysis of immune infiltrates across diverse cancer types [80,81]. “Gene” or “Survival” modules were used to explore the correlation between gene expression patterns and abundance of immune infiltrates as well as between immune infiltrates and clinical outcomes.

Patterns of genetic alterations (missense and truncating mutations, and copy-number variations (deletion or amplification)) were explored in pan-cancer MSK-IMPACT Clinical Sequencing Cohort (10,953 patients/10,967 samples) cBioPortal server [82].

A Cancer Dependency Map or DepMap portal, URL: <https://depmap.org/portal/> (accessed on 1–5 February 2023), allows for defining genes that are essential for cell viability using data on systematic loss-of-function screens (RNA interference and CRISPR-Cas9) in the cell lines representing the heterogeneity of tumors [83]. We used the DepMap portal for definition of ORF genes as “common essential genes” or “strongly selective genes”, implying the significant effects of gene depletion studies in most or several of the screened cell lines, respectively.

Gene expression changes following kinase perturbations assays (inhibition, activation, knockdown, knockout, over-expression and mutation) from the GEO datasets (Gene Expression Omnibus) were explored using kinase perturbations data annotated by the Ma’ayan lab [84], which are available at EnrichR database, URL: <https://maayanlab.cloud/Enrichr/> (accessed on 23–25 February 2023).

Data on the interactions of transcription factors (TFs) with the promotor regions of ORF genes was retrieved from EnrichR database (section ‘Transcription’, subsections ‘ENCODE and ChEA Consensus TFs from ChIP-X’, ‘TRANSFAC and JASPAR PWMs’ and ‘ENCODE TF ChIP-seq 2015’). The resulting list contained matched TFs in all three subsections.

### 4.2. Protein Expression Data analysis

The Human Proteome Atlas (HPA) [85] was used for exploration of protein subcellular localization. In case of absence of available data, predictions were made in the Compartment web-server [86], URL: <https://compartments.jensenlab.org/> (accessed on 29–31 January 2023), which also integrates the results of automat text mining.

Data on protein expression in normal tissues were obtained from the PepPSy server [87], URL: <http://peppsy.genouest.org/query> (accessed on 6 February 2023). This server enables the determining in which human tissues investigators should look for unseen (‘missing’) proteins.

Data on the relative protein abundance in cancer and normal tissues were retrieved from the National Cancer Institute’s Clinical Proteomic Tumor Analysis Consortium (CPTAC) and the National Cancer Institute’s International Cancer Proteogenome Consortium

(ICPC) datasets using cProSite web-based tool [88], URL: <https://cprosite.ccr.cancer.gov/> (accessed on 20 January 2023).

The TOPCONS web-server was used to predict transmembrane topology and membrane domains (TM-helices) of proteins [89], URL: <https://topcons.cbr.su.se/> (accessed on 3 February 2023).

#### 4.3. Protein Subinteractome Analysis

Data on human PPIs (evidence type—physical interactions) were retrieved from several interactomic browsers: BioPlex v.3.0 [90], Human Integrated Protein–Protein Interaction Reference or HIPPIE v.2.3 (cut-off score value > 0.7) [91], The Molecular INTERaction database (MINT) [92], Human Protein Complex Map (hu.MAP 2.0) [93], ‘NURSA Human Endogenous Complexome’ and ‘PPI Hub Proteins’ in the ‘Pathways’ of Enrichr database [94,95], Integrated Interactions Database (IID v.2021-05) [96], InnateDB [97], reference interactome map of human binary protein interactions (HuRI) [98] and interactomics profiling data from publications [99,100].

Over-representation analysis (ORA) of a list of gene names was performed using the WebGestalt server (WEB-based Gene SeT AnaLysis Toolkit) [101] with gene ontology terms (biological process noRedundant, molecular function noRedundant and cellular component noRedundant), disease terms (DisGeNet and GLAD4U databases), phenotype terms (Human Phenotype Ontology database), network terms (CORUM database) and pathway terms (WikiPathway cancer database) with the following settings: reference gene list—‘genome protein-coding’; minimum number of genes for a category—‘4’; multiple test adjustment—‘Benjamini-Hochberg’; significance level—‘FDR < 0.1’; redundancy reduction—‘affinity propagation’. A term, covering > 10% of the total number of proteins in the subinteractome of an ORF protein, was considered a significant term.

CancerGeneNet [102], URL: <https://signor.uniroma2.it/CancerGeneNet/> (accessed on 1–20 March 2023), is a web-server aiming at linking frequently mutated genes in cancers with cancer phenotypes and allows searching for a graph path between any gene-of-interest and ‘hallmarks of cancer’ (i.e., sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis) [103,104]. CancerGeneNet was used to select protein partners of each ORF protein by associations with ‘hallmarks of cancer’.

PathwAX II, a web-server for pathway annotation based on crosstalk derived through FunCoup’s genome wide functional association networks [105], URL: <https://pathwax.sbc.su.se/> (accessed on 15–28 February 2023), was used to explore the pathways that involve proteins, selected by CancerGeneNet web-server, with the following settings: database source—‘KEGG v.94.1’; filter—‘enrichment’, ‘depletion’; multiple hypothesis test correction—‘Benjamini-Hochberg’ and cut-off = 0.01.

The Cancer Gene Census (CGC) database aims to catalogue genes containing mutations that have been causally implicated in cancers (i.e., the driver genes) [106]. CGC was used to search for driver genes, encoding protein partners of ORF proteins, selected by the CancerGeneNet web-server.

Selection of gene co-expression hypotheses in normal and cancer tissues was performed using the GEPIA2 web-server [77] with the following settings: Spearman correlation coefficient > |0.5| at *p*-value < 0.05.

The FunCoup—Functional Coupling framework integrates 10 different evidence types derived from high-throughput genomics and proteomics data in a naive Bayesian integration procedure [107]. The FunCoup v.5.0, URL: <https://funcoup5.scilifelab.se/> (accessed on 21–28 March 2023), was used for PPIs subnetwork construction with the following settings: species—‘Homo sapiens’, confidence score—‘0.7’, expansion depth—‘0’, nodes per expansion step—‘30’, evidence—‘Homo sapiens’, link types—‘all types’.

#### 4.4. Others

The search for full-text publications indexed in the PubMed Central system was performed using the Litsence text-mining tool [108].

Bibliometric analysis and construction of co-occurrence maps of terms in research type publications were performed using the VOSviewer software v.1.6.18 [109].

Principal component analysis was performed using ClustVis [110].

#### 5. Conclusions

Currently, there is no detailed understanding of the molecular mechanisms underlying many systemic diseases, as well as malignant transformation of cells. Chromosome-specific open reading frames genes, which functions are not yet fully characterized, may be involved in carcinogenesis. In this work, we performed a systems biology analysis of 219 target ORF genes that are differentially expressed in various cancers. ORF gene expression patterns correlate with disease prognosis, tumor infiltration by immune cells and responses to chemotherapy. Analysis of individual subinteractomes of ORF proteins revealed interactions with a number of proteins, which are associated with ‘hallmarks of cancer’ as well as proteins encoded by highly mutated genes and mapped in cancer-associated signaling cascades. Some of the ORF proteins are involved in cancer-dependent PPIs and PPIs perturbed by cancers and may have functions related to their protein partners. The findings represent an analytical ‘cross-section’ of the current background on uncharacterized ORF genes in the cancer context that is important for prioritizing further fundamental research in this field.

**Supplementary Materials:** The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241210190/s1>.

**Author Contributions:** Conceptualization, P.E.; methodology, P.E.; investigation, P.E., E.Y. and Y.M.; writing—original draft preparation, P.E.; writing—review and editing, Y.M. and E.Y.; supervision, A.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No.122030100168-2).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data on subinteractomes of uncharacterized chromosome-specific open reading frame genes proteins is deposited in doi:10.6084/m9.figshare.22677442.

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

#### References

1. Messina, B.; Lo Sardo, F.; Scalera, S.; Memeo, L.; Colarossi, C.; Mare, M.; Blandino, G.; Ciliberto, G.; Maugeri-Saccà, M.; Bon, G. Hippo Pathway Dysregulation in Gastric Cancer: From Helicobacter Pylori Infection to Tumor Promotion and Progression. *Cell Death Dis.* **2023**, *14*, 21. [CrossRef]
2. He, Y.; Sun, M.M.; Zhang, G.G.; Yang, J.; Chen, K.S.; Xu, W.W.; Li, B. Targeting PI3K/Akt Signal Transduction for Cancer Therapy. *Signal. Transduct. Target. Ther.* **2021**, *6*, 425. [CrossRef] [PubMed]
3. Ortega, M.A.; Fraile-Martínez, O.; Asúnsolo, Á.; Buján, J.; García-Hondurilla, N.; Coca, S. Signal Transduction Pathways in Breast Cancer: The Important Role of PI3K/Akt/MTOR. *J. Oncol.* **2020**, *2020*, 9258396. [CrossRef]
4. Yang, L.; Shi, P.; Zhao, G.; Xu, J.; Peng, W.; Zhang, J.; Zhang, G.; Wang, X.; Dong, Z.; Chen, F.; et al. Targeting Cancer Stem Cell Pathways for Cancer Therapy. *Signal. Transduct. Target. Ther.* **2020**, *5*, 8. [CrossRef]
5. Chai, H.; Pan, C.; Zhang, M.; Huo, H.; Shan, H.; Wu, J. Histone Methyltransferase SETD1A Interacts with Notch and Promotes Notch Transactivation to Augment Ovarian Cancer Development. *BMC Cancer* **2023**, *23*, 96. [CrossRef] [PubMed]
6. Wang, W.; Lei, Y.; Zhang, G.; Li, X.; Yuan, J.; Li, T.; Zhong, W.; Zhang, Y.; Tan, X.; Song, G. USP39 Stabilizes  $\beta$ -Catenin by Deubiquitination and Suppressing E3 Ligase TRIM26 Pre-mRNA Maturation to Promote HCC Progression. *Cell Death Dis.* **2023**, *14*, 63. [CrossRef]
7. Han, L.; Li, F. Origin Recognition Complex Subunit 1 (ORC1) Augments Malignant Behaviors of Lung Adenocarcinoma Cells via Targeting Wnt Signaling. *Bioengineered* **2022**, *13*, 13520–13533. [CrossRef]

8. Shields, S.; Conroy, E.; O'Grady, T.; McGoldrick, A.; Connor, K.; Ward, M.P.; Useckaite, Z.; Dempsey, E.; Reilly, R.; Fan, Y.; et al. BAG3 Promotes Tumour Cell Proliferation by Regulating EGFR Signal Transduction Pathways in Triple Negative Breast Cancer. *Oncotarget* **2018**, *9*, 15673–15690. [[CrossRef](#)] [[PubMed](#)]
9. Paik, Y.-K.; Overall, C.M.; Deutsch, E.W.; Hancock, W.S.; Omenn, G.S. Progress in the Chromosome-Centric Human Proteome Project as Highlighted in the Annual Special Issue IV. *J. Proteome Res.* **2016**, *15*, 3945–3950. [[CrossRef](#)]
10. Paik, Y.-K.; Lane, L.; Kawamura, T.; Chen, Y.-J.; Cho, J.-Y.; LaBaer, J.; Yoo, J.S.; Domont, G.; Corrales, F.; Omenn, G.S.; et al. Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. *J. Proteome Res.* **2018**, *17*, 4042–4050. [[CrossRef](#)]
11. Duek, P.; Mary, C.; Zahn-Zabal, M.; Bairoch, A.; Lane, L. Functionathon: A Manual Data Mining Workflow to Generate Functional Hypotheses for Uncharacterized Human Proteins and Its Application by Undergraduate Students. *Database* **2021**, *2021*, baab046. [[CrossRef](#)]
12. Hu, F.; Lu, J.; Matheson, L.S.; Díaz-Muñoz, M.D.; Saveliev, A.; Xu, J.; Turner, M. ORFLine: A Bioinformatic Pipeline to Prioritize Small Open Reading Frames Identifies Candidate Secreted Small Proteins from Lymphocytes. *Bioinformatics* **2021**, *37*, 3152–3159. [[CrossRef](#)]
13. Martinez, T.F.; Chu, Q.; Donaldson, C.; Tan, D.; Shokhirev, M.N.; Saghatelian, A. Accurate Annotation of Human Protein-Coding Small Open Reading Frames. *Nat. Chem. Biol.* **2020**, *16*, 458–468. [[CrossRef](#)]
14. Sing, C.N.; Garcia, E.J.; Lipkin, T.G.; Huckaba, T.M.; Tsang, C.A.; Coughlin, A.C.; Yang, E.J.; Boldogh, I.R.; Higuchi-Sanabria, R.; Pon, L.A. Identification of a Modulator of the Actin Cytoskeleton, Mitochondria, Nutrient Metabolism and Lifespan in Yeast. *Nat. Commun.* **2022**, *13*, 2706. [[CrossRef](#)] [[PubMed](#)]
15. Prensner, J.R.; Enache, O.; Ji, Z.; Krug, K.; Clauser, K.R.; Yang, X.; Piccioni, F.; Root, D.E.; Golub, T.R. Abstract 4344: Integrative Functional Proteogenomics for Unannotated or Uncharacterized Proteins in Cancer. *Cancer Res.* **2019**, *79*, 4344. [[CrossRef](#)]
16. Gao, J.; Wang, Y.; Zhang, W.; Zhang, J.; Lu, S.; Meng, K.; Yin, X.; Sun, Z.; He, Q.-Y. C20orf27 Promotes Cell Growth and Proliferation of Colorectal Cancer via the TGFβR-TAK1-NFκB Pathway. *Cancers* **2020**, *12*, 336. [[CrossRef](#)]
17. Tang, Y.; Liao, S.; Liu, G.; Xiong, X.; Liu, H.; Li, F.; Tan, Z.; Kong, X.; Yin, Y.; Tan, B. Advanced Single-Cell Pooled CRISPR Screening Identifies C19orf53 Required for Cell Proliferation Based on MTORC1 Regulators. *Cell Biol. Toxicol.* **2022**, *38*, 43–68. [[CrossRef](#)]
18. Lei, L.; An, G.; Zhu, Z.; Liu, S.; Fu, Y.; Zeng, X.; Cao, Q.; Yan, B. C8orf48 Inhibits the Tumorigenesis of Colorectal Cancer by Regulating the MAPK Signaling Pathway. *Life Sci.* **2021**, *266*, 118872. [[CrossRef](#)]
19. Lee, C.; Cox, R.M.; Papoulas, O.; Horani, A.; Drew, K.; Devitt, C.C.; Brody, S.L.; Marcotte, E.M.; Wallingford, J.B. Functional Partitioning of a Liquid-like Organelle during Assembly of Axonemal Dyneins. *eLife* **2020**, *9*, e58662. [[CrossRef](#)] [[PubMed](#)]
20. Joshi, P.; Greco, T.M.; Guise, A.J.; Luo, Y.; Yu, F.; Nesvizhskii, A.I.; Cristea, I.M. The Functional Interactome Landscape of the Human Histone Deacetylase Family. *Mol. Syst. Biol.* **2013**, *9*, 672. [[CrossRef](#)] [[PubMed](#)]
21. Poluri, R.T.K.; Paquette, V.; Allain, É.P.; Lafront, C.; Joly-Beauparlant, C.; Weidmann, C.; Droit, A.; Guillemette, C.; Pelletier, M.; Audet-Walsh, É. KLF5 and NFYA Factors as Novel Regulators of Prostate Cancer Cell Metabolism. *Endocr. Relat. Cancer* **2021**, *28*, 257–271. [[CrossRef](#)]
22. Li, Y.; Xiao, X.; Chen, H.; Chen, Z.; Hu, K.; Yin, D. Transcription Factor NFYA Promotes G1/S Cell Cycle Transition and Cell Proliferation by Transactivating Cyclin D1 and CDK4 in Clear Cell Renal Cell Carcinoma. *Am. J. Cancer Res.* **2020**, *10*, 2446–2463.
23. Pai, S.-Y.; Truitt, M.L.; Ting, C.-N.; Leiden, J.M.; Glimcher, L.H.; Ho, I.-C. Critical Roles for Transcription Factor GATA-3 in Thymocyte Development. *Immunity* **2003**, *19*, 863–875. [[CrossRef](#)] [[PubMed](#)]
24. Zheng, T.; Huang, J.; Xiang, X.; Li, S.; Yu, J.; Qu, K.; Xu, Z.; Han, P.; Dong, Z.; Liu, Y.; et al. Systematical Analysis Reveals a Strong Cancer Relevance of CREB1-Regulated Genes. *Cancer Cell Int.* **2021**, *21*, 530. [[CrossRef](#)]
25. Xia, L.; Wang, Y.; Meng, Q.; Su, X.; Shen, J.; Wang, J.; He, H.; Wen, B.; Zhang, C.; Xu, M. Integrated Bioinformatic Analysis of a Competing Endogenous RNA Network Reveals a Prognostic Signature in Endometrial Cancer. *Front. Oncol.* **2019**, *9*, 448. [[CrossRef](#)] [[PubMed](#)]
26. Zhou, H.; Zhang, C.; Li, H.; Chen, L.; Cheng, X. A Novel Risk Score System of Immune Genes Associated with Prognosis in Endometrial Cancer. *Cancer Cell Int.* **2020**, *20*, 240. [[CrossRef](#)]
27. Zhong, K.; Wang, Y.; Wang, Z.; Zhang, Z.; Zhao, S.; Li, H.; Huang, J.; Guo, W.; Zheng, X.; Guo, G.; et al. AP-64, Encoded by C5orf46, Exhibits Antimicrobial Activity against Gram-Negative Bacteria. *Biomolecules* **2021**, *11*, 485. [[CrossRef](#)] [[PubMed](#)]
28. Tykodi, S.S.; Fujii, N.; Vigneron, N.; Lu, S.M.; Mito, J.K.; Miranda, M.X.; Chou, J.; Voong, L.N.; Thompson, J.A.; Sandmaier, B.M.; et al. C19orf48 Encodes a Minor Histocompatibility Antigen Recognized by CD8+ Cytotoxic T Cells from Renal Cell Carcinoma Patients. *Clin. Cancer Res.* **2008**, *14*, 5260–5269. [[CrossRef](#)] [[PubMed](#)]
29. Keerthikumar, S.; Chisanga, D.; Ariyaratne, D.; Al Saffar, H.; Anand, S.; Zhao, K.; Samuel, M.; Pathan, M.; Jois, M.; Chilamkurti, N.; et al. ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J. Mol. Biol.* **2016**, *428*, 688–692. [[CrossRef](#)]
30. Simpson, R.J.; Kalra, H.; Mathivanan, S. ExoCarta as a Resource for Exosomal Research. *J. Extracell. Vesicles* **2012**, *1*, 18374. [[CrossRef](#)] [[PubMed](#)]
31. Furukawa, Y.; Lim, C.; Tosha, T.; Yoshida, K.; Hagai, T.; Akiyama, S.; Watanabe, S.; Nakagome, K.; Shiro, Y. Identification of a Novel Zinc-Binding Protein, C1orf123, as an Interactor with a Heavy Metal-Associated Domain. *PLoS ONE* **2018**, *13*, e0204355. [[CrossRef](#)] [[PubMed](#)]

32. Hung, S.-H.; Elliott, G.I.; Ramkumar, T.R.; Burtnyak, L.; McGrenaghan, C.J.; Alkuzweny, S.; Quaiyum, S.; Iwata-Reuyl, D.; Pan, X.; Green, B.D.; et al. Structural Basis of Qng1-Mediated Salvage of the Micronutrient Queuosine from Queuosine-5'-Monophosphate as the Biological Substrate. *Nucleic Acids Res.* **2023**, *51*, 935–951. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Xu, Y.; Wang, S.; Hu, Q.; Gao, S.; Ma, X.; Zhang, W.; Shen, Y.; Chen, F.; Lai, L.; Pei, J. CavityPlus: A Web Server for Protein Cavity Detection with Pharmacophore Modelling, Allosteric Site Identification and Covalent Ligand Binding Ability Prediction. *Nucleic Acids Res.* **2018**, *46*, W374–W379. [\[CrossRef\]](#)
34. Kushibiki, T.; Nakamura, T.; Tsuda, M.; Tsuchikawa, T.; Hontani, K.; Inoko, K.; Takahashi, M.; Asano, T.; Okamura, K.; Murakami, S.; et al. Role of Dimerized C16orf74 in Aggressive Pancreatic Cancer: A Novel Therapeutic Target. *Mol. Cancer* **2020**, *19*, 187–198. [\[CrossRef\]](#)
35. McNally, K.E.; Faulkner, R.; Steinberg, F.; Gallon, M.; Ghai, R.; Pim, D.; Langton, P.; Pearson, N.; Danson, C.M.; Nägele, H.; et al. Retriever Is a Multiprotein Complex for Retromer-Independent Endosomal Cargo Recycling. *Nat. Cell Biol.* **2017**, *19*, 1214–1225. [\[CrossRef\]](#)
36. Chu, Q.; Rathore, A.; Diedrich, J.K.; Donaldson, C.J.; Yates, J.R.; Saghatelian, A. Identification of Microprotein-Protein Interactions via APEX Tagging. *Biochemistry* **2017**, *56*, 3299–3306. [\[CrossRef\]](#)
37. Yang, H.; Zhu, J.; Guo, H.; Tang, A.; Chen, S.; Zhang, D.; Yuan, L.; Liu, G. Molecular Cloning, Characterization, and Functional Analysis of the Uncharacterized C11orf96 Gene. *BMC Vet. Res.* **2022**, *18*, 170. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Edogbanya, J.; Tejada-Martinez, D.; Jones, N.J.; Jaiswal, A.; Bell, S.; Cordeiro, R.; van Dam, S.; Rigden, D.J.; de Magalhães, J.P. Evolution, Structure and Emerging Roles of C1ORF112 in DNA Replication, DNA Damage Responses, and Cancer. *Cell. Mol. Life Sci.* **2021**, *78*, 4365–4376. [\[CrossRef\]](#)
39. Li, M.J.; Xiong, D.; Huang, H.; Wen, Z.Y. Ezrin Promotes the Proliferation, Migration, and Invasion of Ovarian Cancer Cells. *Biomed. Environ. Sci.* **2021**, *34*, 139–151. [\[CrossRef\]](#)
40. Tanaka, C.; Kobori, T.; Tameishi, M.; Urashima, Y.; Ito, T.; Obata, T. Ezrin Modulates the Cell Surface Expression of Programmed Cell Death Ligand-1 in Human Cervical Adenocarcinoma Cells. *Molecules* **2021**, *26*, 5648. [\[CrossRef\]](#)
41. Margaryan, H.; Dorosh, A.; Capkova, J.; Manaskova-Postlerova, P.; Philimonenko, A.; Hozak, P.; Peknicova, J. Characterization and Possible Function of Glyceraldehyde-3-Phosphate Dehydrogenase-Spermatogenic Protein GAPDHS in Mammalian Sperm. *Reprod. Biol. Endocrinol.* **2015**, *13*, 15. [\[CrossRef\]](#)
42. Fukushima, A.; Kim, H.-D.; Chang, Y.-C.; Kim, C.-H. Revisited Metabolic Control and Reprogramming Cancers by Means of the Warburg Effect in Tumor Cells. *Int. J. Mol. Sci.* **2022**, *23*, 10037. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Olsen, T.K.; Panagopoulos, I.; Gorunova, L.; Micci, F.; Andersen, K.; Kilen Andersen, H.; Meling, T.R.; Due-Tønnessen, B.; Scheie, D.; Heim, S.; et al. Novel Fusion Genes and Chimeric Transcripts in Ependymal Tumors. *Genes. Chromosom. Cancer* **2016**, *55*, 944–953. [\[CrossRef\]](#)
44. Lian, W.; Jin, H.; Cao, J.; Zhang, X.; Zhu, T.; Zhao, S.; Wu, S.; Zou, K.; Zhang, X.; Zhang, M.; et al. Identification of Novel Biomarkers Affecting the Metastasis of Colorectal Cancer through Bioinformatics Analysis and Validation through QRT-PCR. *Cancer Cell Int.* **2020**, *20*, 105. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Xie, H.; Wang, Y.; Zhang, H.; Fan, Q.; Dai, D.; Zhuang, L.; Tao, R.; Chen, Q.; Shen, W.; Lu, L.; et al. Tubular Epithelial C1orf54 Mediates Protection and Recovery from Acute Kidney Injury. *J. Cell. Mol. Med.* **2018**, *22*, 4985–4996. [\[CrossRef\]](#)
46. Ni, C.; Schmitz, D.A.; Lee, J.; Pawłowski, K.; Wu, J.; Buszczak, M. Labeling of Heterochronic Ribosomes Reveals C1ORF109 and SPATA5 Control a Late Step in Human Ribosome Assembly. *Cell Rep.* **2022**, *38*, 110597. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Masud, S.N.; Chandrashekhar, M.; Aregger, M.; Tan, G.; Zhang, X.; Mero, P.; Pirman, D.A.; Zaslaver, O.; Smolen, G.A.; Lin, Z.-Y.; et al. Chemical Genomics with Pyrvinium Identifies C1orf115 as a Regulator of Drug Efflux. *Nat. Chem. Biol.* **2022**, *18*, 1370–1379. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Lau, M.-T.; Ghazanfar, S.; Parkin, A.; Chou, A.; Rouaen, J.R.; Littleboy, J.B.; Nessem, D.; Khuong, T.M.; Nevoltris, D.; Schofield, P.; et al. Systematic Functional Identification of Cancer Multi-Drug Resistance Genes. *Genome Biol.* **2020**, *21*, 27. [\[CrossRef\]](#)
49. Singh, S.; Vanden Broeck, A.; Miller, L.; Chaker-Margot, M.; Klinge, S. Nucleolar Maturation of the Human Small Subunit Processome. *Science* **2021**, *373*, eabj5338. [\[CrossRef\]](#)
50. Sun, S.-C.; Ma, D.; Li, M.-Y.; Zhang, R.-X.; Huang, C.; Huang, H.-J.; Xie, Y.-Z.; Wang, Z.-J.; Liu, J.; Cai, D.-C.; et al. Mutations in C1orf194, Encoding a Calcium Regulator, Cause Dominant Charcot-Marie-Tooth Disease. *Brain* **2019**, *142*, 2215–2229. [\[CrossRef\]](#)
51. Ashikawa, Y.; Shiromizu, T.; Miura, K.; Adachi, Y.; Matsui, T.; Bessho, Y.; Tanaka, T.; Nishimura, Y. C3orf70 Is Involved in Neural and Neurobehavioral Development. *Pharmaceuticals* **2019**, *12*, 156. [\[CrossRef\]](#)
52. Yan, B.-R.; Li, T.; Coyaud, E.; Laurent, E.M.N.; St-Germain, J.; Zhou, Y.; Kim, P.K.; Raught, B.; Brumell, J.H. C5orf51 Is a Component of the MON1-CCZ1 Complex and Controls RAB7A Localization and Stability during Mitophagy. *Autophagy* **2022**, *18*, 829–840. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Li, X.; Qiao, Y.; Chang, L.-S.; Xiao, F.; Lu, L.-H.; Hao, X.-H.; Zhang, R.-W.; Wu, H.; Wei, H.-S. Role of C6ORF120, an N-Glycosylated Protein, Is Implicated in Apoptosis of CD4<sup>+</sup> T Lymphocytes. *Chin. Med. J.* **2011**, *124*, 3560–3567. [\[PubMed\]](#)
54. Zhang, J.; Zhang, M.-K.; Ma, H.-M.; Song, X.-C.; Wu, Y.-N.; Zhang, R.; He, L.-L.; Ye, X.-H.; Gao, M.-X.; Li, X. C6orf120 Gene Deficiency May Be Vulnerable to Carbon Tetrachloride Induced Acute Hepatic Injury in Rats. *Arch. Med. Sci.* **2022**, *18*, 1626–1637. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Li, D.; Pan, J.; Zhang, Y.; Li, Y.; Jin, S.; Zhong, C.; Chen, P.; Ma, J.; Hu, W.; Fan, X.; et al. C8orf76 Modulates Ferroptosis in Liver Cancer via Transcriptionally Up-Regulating SLC7A11. *Cancers* **2022**, *14*, 3410. [\[CrossRef\]](#)

56. Gopalakrishna, S.; Pearce, S.F.; Dinan, A.M.; Schober, F.A.; Cipullo, M.; Spähr, H.; Khawaja, A.; Maffezzini, C.; Freyer, C.; Wredenberg, A.; et al. C6orf203 Is an RNA-Binding Protein Involved in Mitochondrial Protein Synthesis. *Nucleic Acids Res.* **2019**, *47*, 9386–9399. [\[CrossRef\]](#)
57. Kotrys, A.V.; Cysewski, D.; Czarnomska, S.D.; Pietras, Z.; Borowski, L.S.; Dziembowski, A.; Szczesny, R.J. Quantitative Proteomics Revealed C6orf203/MTRES1 as a Factor Preventing Stress-Induced Transcription Deficiency in Human Mitochondria. *Nucleic Acids Res.* **2019**, *47*, 7502–7517. [\[CrossRef\]](#)
58. Palukuri, M.V.; Marcotte, E.M. SuperComplex: A Supervised Machine Learning Pipeline for Molecular Complex Detection in Protein-Interaction Networks. *PLoS ONE* **2021**, *16*, e0262056. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Szczepanski, A.P.; Tsuboyama, N.; Watanabe, J.; Hashizume, R.; Zhao, Z.; Wang, L. POU2AF2/C11orf53 Functions as a Coactivator of POU2F3 by Maintaining Chromatin Accessibility and Enhancer Activity. *Sci. Adv.* **2022**, *8*, eabq2403. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Zhou, C.; Huang, H.; Wang, Y.; Sendinc, E.; Shi, Y. Selective Regulation of Tuft Cell-like Small Cell Lung Cancer by Novel Transcriptional Co-Activators C11orf53 and COLCA2. *Cell Discov.* **2022**, *8*, 112. [\[CrossRef\]](#)
61. Narita, K.; Nagatomo, H.; Kozuka-Hata, H.; Oyama, M.; Takeda, S. Discovery of a Vertebrate-Specific Factor That Processes Flagellar Glycolytic Enolase during Motile Ciliogenesis. *iScience* **2020**, *23*, 100992. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Hao, H.; Shi, B.; Zhang, J.; Dai, A.; Li, W.; Chen, H.; Ji, W.; Gong, C.; Zhang, C.; Li, J.; et al. The Vertebrate- and Testis- Specific Transmembrane Protein C11ORF94 Plays a Critical Role in Sperm-Oocyte Membrane Binding. *Mol. Biomed.* **2022**, *3*, 27. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Contreras, W.; Wiesehöfer, C.; Schreier, D.; Leinung, N.; Peche, P.; Wennemuth, G.; Gentzel, M.; Schröder, B.; Mentrup, T. C11orf94/Frey Is a Key Regulator for Male Fertility by Controlling Izumo1 Complex Assembly. *Sci. Adv.* **2022**, *8*, eabo6049. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Mazuc, E.; Guglielmi, L.; Bec, N.; Perez, V.; Hahn, C.S.; Mollevi, C.; Parrinello, H.; Desvignes, J.-P.; Larroque, C.; Jupp, R.; et al. In-Cell Intrabody Selection from a Diverse Human Library Identifies C12orf4 Protein as a New Player in Rodent Mast Cell Degranulation. *PLoS ONE* **2014**, *9*, e104998. [\[CrossRef\]](#)
65. Dudkiewicz, M.; Pawłowski, K. A Novel Conserved Family of Macro-like Domains-Putative New Players in ADP-Ribosylation Signaling. *PeerJ* **2019**, *7*, e6863. [\[CrossRef\]](#)
66. Abel, Y.; Paiva, A.C.F.; Bizarro, J.; Chagot, M.-E.; Santo, P.E.; Robert, M.-C.; Quinternet, M.; Vandermoere, F.; Sousa, P.M.F.; Fort, P.; et al. NOPCHAP1 Is a PAQosome Cofactor That Helps Loading NOP58 on RUVBL1/2 during Box C/D SnoRNP Biogenesis. *Nucleic Acids Res.* **2021**, *49*, 1094–1113. [\[CrossRef\]](#)
67. Aregger, M.; Lawson, K.A.; Billmann, M.; Costanzo, M.; Tong, A.H.Y.; Chan, K.; Rahman, M.; Brown, K.R.; Ross, C.; Usaj, M.; et al. Systematic Mapping of Genetic Interactions for de Novo Fatty Acid Synthesis Identifies C12orf49 as a Regulator of Lipid Metabolism. *Nat. Metab.* **2020**, *2*, 499–513. [\[CrossRef\]](#)
68. Xiao, J.; Xiong, Y.; Yang, L.-T.; Wang, J.-Q.; Zhou, Z.-M.; Dong, L.-W.; Shi, X.-J.; Zhao, X.; Luo, J.; Song, B.-L. POST1/C12ORF49 Regulates the SREBP Pathway by Promoting Site-1 Protease Maturation. *Protein Cell* **2021**, *12*, 279–296. [\[CrossRef\]](#)
69. Dennerlein, S.; Poerschke, S.; Oeljeklaus, S.; Wang, C.; Richter-Dennerlein, R.; Sattmann, J.; Bauermeister, D.; Hanitsch, E.; Stoldt, S.; Langer, T.; et al. Defining the Interactome of the Human Mitochondrial Ribosome Identifies SMIM4 and TMEM223 as Respiratory Chain Assembly Factors. *Elife* **2021**, *10*, e68213. [\[CrossRef\]](#)
70. Paplomata, E.; O'Regan, R. The PI3K/AKT/MTOR Pathway in Breast Cancer: Targets, Trials and Biomarkers. *Adv. Med. Oncol.* **2014**, *6*, 154–166. [\[CrossRef\]](#)
71. Dong, C.; Wu, J.; Chen, Y.; Nie, J.; Chen, C. Activation of PI3K/AKT/MTOR Pathway Causes Drug Resistance in Breast Cancer. *Front. Pharm.* **2021**, *12*, 628690. [\[CrossRef\]](#)
72. Yang, Y.; Hwang, H.; Im, J.E.; Lee, K.; Bhoo, S.H.; Yoo, J.S.; Kim, Y.-H.; Kim, J.Y. Flashlight into the Function of Unannotated C11orf52 Using Affinity Purification Mass Spectrometry. *J. Proteome Res.* **2021**, *20*, 5340–5346. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Santra, T.; Herrero, A.; Rodriguez, J.; von Kriegsheim, A.; Iglesias-Martinez, L.F.; Schwarzl, T.; Higgins, D.; Aye, T.-T.; Heck, A.J.R.; Calvo, F.; et al. An Integrated Global Analysis of Compartmentalized HRAS Signaling. *Cell Rep.* **2019**, *26*, 3100–3115.e7. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Wei, X.; Cai, S.; Boohaker, R.J.; Fried, J.; Li, Y.; Hu, L.; Pan, Y.; Cheng, R.; Zhang, S.; Tian, Y.; et al. KAT5 Promotes Invasion and Metastasis through C-MYC Stabilization in ATC. *Endocr. Relat. Cancer* **2019**, *26*, 141–151. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Srivastava, S.; Wagner, P.D. The Early Detection Research Network: A National Infrastructure to Support the Discovery, Development, and Validation of Cancer Biomarkers. *Cancer Epidemiol. Biomark. Prev.* **2020**, *29*, 2401–2410. [\[CrossRef\]](#)
76. Kelleher, K.J.; Sheils, T.K.; Mathias, S.L.; Yang, J.J.; Metzger, V.T.; Siramshetty, V.B.; Nguyen, D.-T.; Jensen, L.J.; Vidović, D.; Schürer, S.C.; et al. Pharos 2023: An Integrated Resource for the Understudied Human Proteome. *Nucleic Acids Res.* **2023**, *51*, D1405–D1416. [\[CrossRef\]](#)
77. Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* **2019**, *47*, W556–W560. [\[CrossRef\]](#)
78. Nagy, Á.; Munkácsy, G.; Györfy, B. Pancancer Survival Analysis of Cancer Hallmark Genes. *Sci. Rep.* **2021**, *11*, 6047. [\[CrossRef\]](#)
79. Fekete, J.T.; Györfy, B. ROCplot.Org: Validating Predictive Biomarkers of Chemotherapy/Hormonal Therapy/Anti-HER2 Therapy Using Transcriptomic Data of 3,104 Breast Cancer Patients. *Int. J. Cancer* **2019**, *145*, 3140–3151. [\[CrossRef\]](#)
80. Li, T.; Fan, J.; Wang, B.; Traugh, N.; Chen, Q.; Liu, J.S.; Li, B.; Liu, X.S. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* **2017**, *77*, e108–e110. [\[CrossRef\]](#)

81. Li, B.; Severson, E.; Pignon, J.-C.; Zhao, H.; Li, T.; Novak, J.; Jiang, P.; Shen, H.; Aster, J.C.; Rodig, S.; et al. Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol.* **2016**, *17*, 174. [\[CrossRef\]](#)
82. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [\[CrossRef\]](#)
83. Tsherniak, A.; Vazquez, F.; Montgomery, P.G.; Weir, B.A.; Kryukov, G.; Cowley, G.S.; Gill, S.; Harrington, W.F.; Pantel, S.; Krill-Burger, J.M.; et al. Defining a Cancer Dependency Map. *Cell* **2017**, *170*, 564–576.e16. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Rouillard, A.D.; Gundersen, G.W.; Fernandez, N.F.; Wang, Z.; Monteiro, C.D.; McDermott, M.G.; Ma'ayan, A. The Harmonizome: A Collection of Processed Datasets Gathered to Serve and Mine Knowledge about Genes and Proteins. *Database* **2016**, *2016*, baw100. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-Based Map of the Human Proteome. *Science* **2015**, *347*, 1260419. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Binder, J.X.; Pletscher-Frankild, S.; Tsafou, K.; Stolte, C.; O'Donoghue, S.I.; Schneider, R.; Jensen, L.J. COMPARTMENTS: Unification and Visualization of Protein Subcellular Localization Evidence. *Database* **2014**, *2014*, bau012. [\[CrossRef\]](#)
87. Sallou, O.; Duek, P.D.; Darde, T.A.; Collin, O.; Lane, L.; Chalmel, F. PepPSy: A Web Server to Prioritize Gene Products in Experimental and Biocuration Workflows. *Database* **2016**, *2016*, baw070. [\[CrossRef\]](#)
88. Wang, D.; Qian, X.; Du, Y.-C.N.; Sanchez-Solana, B.; Chen, K.; Park, B.; Chen, B.; Jenkins, L.; Luo, J.; Tripathi, B.K.; et al. Abstract 3912: CProSite: A Web Based Interactive Platform for on-Line Proteomics and Phosphoproteomics Data Analysis. *Cancer Res.* **2022**, *82*, 3912. [\[CrossRef\]](#)
89. Tsirigos, K.D.; Peters, C.; Shu, N.; Käll, L.; Elofsson, A. The TOPCONS Web Server for Consensus Prediction of Membrane Protein Topology and Signal Peptides. *Nucleic Acids Res.* **2015**, *43*, W401–W407. [\[CrossRef\]](#)
90. Schweppe, D.K.; Huttlin, E.L.; Harper, J.W.; Gygi, S.P. BioPlex Display: An Interactive Suite for Large-Scale AP-MS Protein-Protein Interaction Data. *J. Proteome Res.* **2018**, *17*, 722–726. [\[CrossRef\]](#)
91. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing Meaningfulness and Reliability of Protein-Protein Interaction Networks. *Nucleic Acids Res.* **2017**, *45*, D408–D414. [\[CrossRef\]](#)
92. Chatr-aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular INTeraction Database. *Nucleic Acids Res.* **2007**, *35*, D572–D574. [\[CrossRef\]](#)
93. Drew, K.; Wallingford, J.B.; Marcotte, E.M. Hu.MAP 2.0: Integration of over 15,000 Proteomic Experiments Builds a Global Compendium of Human Multiprotein Assemblies. *Mol. Syst. Biol.* **2021**, *17*, e10016. [\[CrossRef\]](#)
94. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool. *BMC Bioinform.* **2013**, *14*, 128. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Pastrello, C.; Kotlyar, M.; Jurisica, I. Informed Use of Protein-Protein Interaction Data: A Focus on the Integrated Interactions Database (IID). *Methods Mol. Biol.* **2020**, *2074*, 125–134. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Breuer, K.; Foroushani, A.K.; Laird, M.R.; Chen, C.; Sribnaia, A.; Lo, R.; Winsor, G.L.; Hancock, R.E.W.; Brinkman, F.S.L.; Lynn, D.J. InnateDB: Systems Biology of Innate Immunity and beyond—Recent Updates and Continuing Curation. *Nucleic Acids Res.* **2013**, *41*, D1228–D1233. [\[CrossRef\]](#)
98. Luck, K.; Kim, D.-K.; Lambourne, L.; Spirohn, K.; Begg, B.E.; Bian, W.; Brignall, R.; Cafarelli, T.; Campos-Laborie, F.J.; Charlotiaux, B.; et al. A Reference Map of the Human Binary Protein Interactome. *Nature* **2020**, *580*, 402–408. [\[CrossRef\]](#)
99. Drew, K.; Lee, C.; Huizar, R.L.; Tu, F.; Borgeson, B.; McWhite, C.D.; Ma, Y.; Wallingford, J.B.; Marcotte, E.M. Integration of over 9,000 Mass Spectrometry Experiments Builds a Global Map of Human Protein Complexes. *Mol. Syst. Biol.* **2017**, *13*, 932. [\[CrossRef\]](#)
100. Havugimana, P.C.; Goel, R.K.; Phanse, S.; Youssef, A.; Padhorny, D.; Kotelnikov, S.; Kozakov, D.; Emili, A. Scalable Multiplex Co-Fractionation/Mass Spectrometry Platform for Accelerated Protein Interactome Discovery. *Nat. Commun.* **2022**, *13*, 4043. [\[CrossRef\]](#)
101. Wang, J.; Duncan, D.; Shi, Z.; Zhang, B. WEB-Based GENE SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **2013**, *41*, W77–W83. [\[CrossRef\]](#)
102. Iannuccelli, M.; Micarelli, E.; Surdo, P.L.; Palma, A.; Perfetto, L.; Rozzo, I.; Castagnoli, L.; Licata, L.; Cesareni, G. CancerGeneNet: Linking Driver Genes to Cancer Hallmarks. *Nucleic Acids Res.* **2020**, *48*, D416–D421. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The next Generation. *Cell* **2011**, *144*, 646–674. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Senga, S.S.; Grose, R.P. Hallmarks of Cancer—the New Testament. *Open. Biol.* **2021**, *11*, 200358. [\[CrossRef\]](#)
105. Ogris, C.; Castresana-Aguirre, M.; Sonnhammer, E.L.L. PathwAX II: Network-Based Pathway Analysis with Interactive Visualization of Network Crosstalk. *Bioinformatics* **2022**, *38*, 2659–2660. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Sondka, Z.; Bamford, S.; Cole, C.G.; Ward, S.A.; Dunham, I.; Forbes, S.A. The COSMIC Cancer Gene Census: Describing Genetic Dysfunction across All Human Cancers. *Nat. Rev. Cancer* **2018**, *18*, 696–705. [\[CrossRef\]](#) [\[PubMed\]](#)
107. Persson, E.; Castresana-Aguirre, M.; Buzzao, D.; Guala, D.; Sonnhammer, E.L.L. FunCoup 5: Functional Association Networks in All Domains of Life, Supporting Directed Links and Tissue-Specificity. *J. Mol. Biol.* **2021**, *433*, 166835. [\[CrossRef\]](#)

108. Allot, A.; Chen, Q.; Kim, S.; Vera Alvarez, R.; Comeau, D.C.; Wilbur, W.J.; Lu, Z. LitSense: Making Sense of Biomedical Literature at Sentence Level. *Nucleic Acids Res.* **2019**, *47*, W594–W599. [[CrossRef](#)]
109. Arruda, H.; Silva, E.R.; Lessa, M.; Proença, D.; Bartholo, R. VOSviewer and Bibliometrix. *J. Med. Libr. Assoc.* **2022**, *110*, 392–395. [[CrossRef](#)]
110. Metsalu, T.; Vilo, J. ClustVis: A Web Tool for Visualizing Clustering of Multivariate Data Using Principal Component Analysis and Heatmap. *Nucleic Acids Res.* **2015**, *43*, W566–W570. [[CrossRef](#)]

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