

Article

Gene Co-Expression Analysis of Human *RNASEH2A* Reveals Functional Networks Associated with DNA Replication, DNA Damage Response, and Cell Cycle Regulation

Stefania Marsili, Ailone Tichon, Deepali Kundnani and Francesca Storici *

Georgia Institute of Technology, School of Biological Sciences, Atlanta, GA 30332, USA; marsili@gatech.edu (S.M.); eilonti1@gmail.com (A.T.); dkundnani3@gatech.edu (D.K.)

* Correspondence: storici@gatech.edu

Simple Summary: *RNASEH2A* is the catalytic subunit of the ribonuclease (RNase) H2 ternary complex that plays an important role in maintaining DNA stability in cells. Recent studies have shown that the *RNASEH2A* subunit alone is highly expressed in certain cancer cell types. Via a series of bioinformatics approaches, we found that *RNASEH2A* is highly expressed in human proliferative tissues and many cancers. Our analyses reveal a possible involvement of *RNASEH2A* in cell cycle regulation in addition to its well established role in DNA replication and DNA repair. Our findings underscore that *RNASEH2A* could serve as a biomarker for cancer diagnosis and a therapeutic target.

Abstract: Ribonuclease (RNase) H2 is a key enzyme for the removal of RNA found in DNA-RNA hybrids, playing a fundamental role in biological processes such as DNA replication, telomere maintenance, and DNA damage repair. RNase H2 is a trimer composed of three subunits, *RNASEH2A* being the catalytic subunit. *RNASEH2A* expression levels have been shown to be upregulated in transformed and cancer cells. In this study, we used a bioinformatics approach to identify *RNASEH2A* co-expressed genes in different human tissues to underscore biological processes associated with *RNASEH2A* expression. Our analysis shows functional networks for *RNASEH2A* involvement such as DNA replication and DNA damage response and a novel putative functional network of cell cycle regulation. Further bioinformatics investigation showed increased gene expression in different types of actively cycling cells and tissues, particularly in several cancers, supporting a biological role for *RNASEH2A* but not for the other two subunits of RNase H2 in cell proliferation. Mass spectrometry analysis of *RNASEH2A*-bound proteins identified players functioning in cell cycle regulation. Additional bioinformatic analysis showed that *RNASEH2A* correlates with cancer progression and cell cycle related genes in Cancer Cell Line Encyclopedia (CCLE) and The Cancer Genome Atlas (TCGA) Pan Cancer datasets and supported our mass spectrometry findings.

Keywords: RNASEH; *RNASEH2A*; co-expression; genotype-tissue expression; cell cycle; cancer



Citation: Marsili, S.; Tichon, A.; Kundnani, D.; Storici, F. Gene Co-Expression Analysis of Human *RNASEH2A* Reveals Functional Networks Associated with DNA Replication, DNA Damage Response, and Cell Cycle Regulation. *Biology* **2021**, *10*, 221. <https://doi.org/10.3390/biology10030221>

Academic Editor: Michael Seidman

Received: 22 February 2021

Accepted: 10 March 2021

Published: 13 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

1. Introduction

The development of high-throughput tools to monitor gene expression levels in a specific cell type and tissue has allowed the characterization of gene expression patterns throughout the human body. Such an approach was utilized in several studies to identify molecular signatures of tissues and cells [1–3]. Genes that share a molecular pathway in a given tissue or cell should be co-expressed in a spatial and a temporal manner. Therefore, examining genes that show high co-expression correlation in multiple tissues can be used to identify a shared molecular pathway between those genes. The Genotype-Tissue Expression portal (GTEx) [3] has been used as a suitable platform to identify gene co-expression in the human body based on the availability of the full transcriptome in 53 different human tissues. Gene ontology analysis of the co-expressed genes might shed light on genes with

uncharacterized functions and help cover additional functions of genes for which their role is partially known.

RNASEH2A, together with *RNASEH2B* and *RNASEH2C*, composes the holoenzyme ribonuclease (RNase) H2 [4,5]. *RNASEH2B* and *RNASEH2C* are used as the scaffold for *RNASEH2A*, which serves as the catalytic subunit of the RNase H2 complex [6]. The function of RNase H2, together with *RNASEH1*, is to cleave RNA of RNA-DNA hybrids, which can be formed during transcription [7], DNA replication [8], and repair [9]. RNA-DNA hybrids, if not repaired, can harm the genomic DNA by different mechanisms, such as modifying the DNA structure, blocking DNA replication and transcription, causing hyper-recombination and mutation, or chromosome loss [4,10–13]. Phenotypically, it has been postulated that the RNA-DNA hybrids mimic the infection of nucleic acids of viral origin, thus activating the innate immune response [14]. An example of a pathologic condition is the Aicardi-Goutières Syndrome (AGS), a severe autosomal recessive neurological disorder with symptoms similar to in-utero viral infection [15,16]. Genetic studies have linked AGS to mutations in the genes composing RNase H2, and it has been shown that mutations silencing intracellular RNases affect microRNA turnover, resulting in the severe clinical consequences in the brain that characterize the clinical feature of AGS [17]. However, no evidence was found that *RNASEH1* is linked to the AGS pathology [14], highlighting a marked difference between the function of these two enzymes. Another distinctive feature of *RNASEH2A* comes from examining its expression levels in cancer compared to normal tissues. A previous study showed that, in human mesenchymal stem cells transformed by the over-expression of several oncogenes, *RNASEH2A* was among the genes with the highest and the earliest up fold change in their expression [18]. In addition, increased levels of *RNASEH2A* in cancer compared to normal tissues and cells were also reported in cervical cancer [19], prostate cancer [20], colorectal carcinoma [21], and triple negative breast cancer [22]. In support of this finding, a recent study by Luke's group [23] highlights a differential cell-cycle regulation for the activity and the level of the *RNASEH2A* orthologous gene in yeast *Saccharomyces cerevisiae* (*RNH201*), peaking at S and G2 phases of the cell cycle, supporting the hypothesis that *RNASEH2A* plays a role in cancer progression. This is in contrast with the yeast *RNASEH1* ortholog, *RNH1*, activity and level of which are not altered throughout the cell cycle [23]. With the goal to better understand the biological role of *RNASEH2A* in cancer, we analyzed the prevalence of patients with copy number alterations in the *RNASEH2A* gene in The Cancer Genome Atlas (TCGA) Pan Cancer studies in different cancer tissues, hypothesizing that identifying patterns of its expression can shed new insights on its function.

2. Materials and Methods

2.1. Co-Expression Correlation Analysis

We obtained expression data from GTEx portal v7 [3] (<https://gtexportal.org/home/datasets>; accessed date 21 September 2019) which were obtained through RNA-Seq for 53 different tissues. The dataset contained total 56,202 transcripts, out of which 42,548 transcripts had greater than zero transcripts per million (TPM) values in at least one tissue of the 53 different tissues. We calculated Pearson correlation coefficients for every transcript with *RNASEH* transcripts data using R [24]. Then, we analyzed the gene ontology (GO) term of the top 2% or 5% co-expressed genes using Gene Ontology enRichment anaLysis and visualizaTion tool (GORilla) [25].

2.2. Co-Expression Correlation Analysis Verification with STRING Database

To validate the correlation co-expression coefficients of different genes with *RNASEH2A* in GTEx tissues, we used protein co-expression data obtained from STRING protein–protein association network v11 [26]. We looked at top correlated genes including *RNASEH2A* in GTEx dataset, which also had at least 20 exclusively co-expressed proteins in STRING protein association network. We used the default filtering criteria of medium confidence score in STRING database, selecting only co-expressed proteins. With 10 such transcripts

(including *RNASEH2A*) and 20 co-expressed proteins for each from the STRING database, we ended up with a list of 200 proteins. We removed any duplicate proteins listed and counted how many proteins were in the top five percentile of the Pearson's correlation coefficient distribution from the co-expression correlation analysis that was to be validated.

2.3. Gene Expression Analysis in Cancer

Data were obtained from Cancer RNA-Seq Nexus (CRN) database providing phenotype-specific coding-transcript/lncRNA expression profiles, and mRNA-lncRNA co-expression networks in cancer cells [27] were used to examine the expression level of *RNASEH* genes in different type of cancers compared to normal tissues and in cancers at different stage of progression.

2.4. *RNASEH2A* Copy Number Alterations Analysis in TCGA Pan Cancer Dataset

To analyze copy number alterations (CNAs) in *RNASEH2A*, we used The Cancer Genome Atlas (TCGA) Pan Cancer studies [28] involving 32 studies and 10,967 patients through cBioPortal [29,30]. After filtering samples containing both copy number alterations generated by Genomic Identification of Significant Targets in Cancer (GISTIC) and RNA-Seq data, processing, and normalizing by RNA-Seq by expectation maximization (RSEM), data for 9889 patients were used for final analysis.

2.5. Cell and Protein Extracts

Human embryonic kidney, HEK293 were grown in Dulbecco's Modified Eagle Medium (DMEM) medium (Corning Inc., Corning, NY, USA, cat # 45000-304) containing 10% fetal bovine serum (FBS) (Sigma-Aldrich, St. Louis, MO, USA, cat # F0926) and 1% penicillin-streptomycin (Fisher Scientific, Hampton, NH, USA, cat # 15-140-122) in 5% CO₂ incubator at 37 °C and were passaged once a week to maintain cell growth. Whole cell lysate from HEK293 was extracted with NP-40 lysis buffer (50 mM Tris-HCL pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40) together with a cocktail of protease inhibitor (Complete EDTA-free, Roche Applied Science, Indianapolis, IN, USA) followed by centrifugation at 14,000 rpm for 15 min at +4 °C to pellet the debris.

2.6. Plasmids and Transfection

To over-express *RNASEH2A* in HEK293 cell line, we used the *pEGFP-RNASEH2A* plasmid (Addgene, Watertown, MA, USA, cat # 108700). As control, we cloned the *eGFP* plasmid by restricting the *pEGFP-RNASEH2A* using HincII (New England Biolabs, Radnor, PA, USA, cat # R0103S) that cut upstream and downstream of the *RNASEH2A* gene but left the *EGFP* intact. Then, the plasmid was ligated by T4 ligase (New England Biolabs, Ipswich, MA, USA, cat # 101228-176) and transformed into Agilent XL1-Blue Electroporation-Competent Cells (Agilent, Santa Clare, CA, USA, cat # 50-125-045) by electroporation using Gene Pulser Xcell electroporation system (Bio-Rad, Hercules, CA, USA, cat # 1652660). Isolated colonies were grown in Luria-Bertani (LB) medium overnight; plasmids were extracted using the miniprep kit (Fisher Scientific, Hampton, NH, USA, cat # FERK0503) and sent to DNA sequencing to verify the plasmid sequence.

For transient transfection, HEK293 cells were plated at $1-3 \times 10^6$ per 10 cm plate. The following day, 5 µg of the plasmids were incubated with 25 µL of the Lipofectamine transfection reagent (Fisher Scientific, Hampton, NH, USA, cat # 11668019) for 10–15 min at room temperature in Opti-MEM media (Fisher Scientific, Hampton, NH, USA, cat # 31985062), and the plasmid-reagent mix was distributed to the cells. The next day, transfected cells were fed with fresh media and harvested 48 h after transfection.

2.7. Co-Immunoprecipitation

Whole cell lysate (1 mg) from HEK293 cells transfected with either the *eGFP* or the *pEGFP-RNASEH2A* plasmid was incubated with 30 µL of GFP-Trap Magnetic Agarose beads (ChromoTek, Planegg-Martinsried, Germany, gtm-10 lot # 80912001MA) at 4 °C

with end to end rotation for 1 h. Then, the magnetic beads–GFP protein complex were separated from the rest of the protein extract using a magnetic stand. The beads were washed twice with wash buffer I (50 mM Tris-HCL pH 7.5, 150 mM NaCl, 0.25% NP-40) and an additional 2 washes with wash buffer II (50 mM Tris-HCL pH 7.5, 150 mM NaCl). Then, the beads were resuspended in SDS loading buffer and boiled at 95 °C for 10 min. Finally, the beads were centrifuged at 13,000 rpm for 5 min at room temperature, and the eluted proteins were collected and used for either Western blot or mass spectrometry analysis.

2.8. Protein Analysis (Western Blot)

Total protein concentrations from HEK293 lysate were quantified (Bradford Protein Assay, Pierce Biotechnology, Rockford, IL, USA), and 15 µg of the protein extracts was separated on a 10% SDS-polyacrylamide electrophoresis gel, transferred to 0.45 µm nitrocellulose membrane (Amersham, Buckinghamshire, UK, cat # 10120-006). The blotted membrane was then blocked in 5% non-fat dry milk tris-buffered saline and Tween 20 (TBS-T) (10 mM Tris-Cl (pH 7.5), 100 mM NaCl, 0.1% Tween-20) at room temperature for 1 h and incubated overnight with specific primary antibodies: *anti-RNASEH2A* (Santa Cruz Biotechnology, Dallas, TX, USA, 1:1000) and anti-GFP (B2) (Santa Cruz Biotechnology, Dallas, TX, USA, 1:1000). After washing 3 × 10 min with TBS-T, the membrane was incubated with a mouse secondary antibody conjugated with horseradish peroxidase (Thermo Fisher Scientific, Waltham, MA, 1:10,000) at room temperature for 1 h. Following washing 3 × 10 min with TBS-T, protein signals were visualized using the electrochemiluminescence (ECL) method according to the manufacturer's recommendations (Pierce™ ECL Western Blotting Substrate, Thermo Fisher Scientific, Waltham, MA, USA) and exposed on autoradiograph films (Denville Scientific, South Plainfield, NJ, USA).

2.9. Mass Spectrometry Analysis

The gel lanes were excised from the gel, and the proteins were reduced, alkylated, and digested with trypsin as previously described [31]. The peptides were analyzed by nano-LC-MS/MS and peptide identification as previously described [32]. The raw files were searched using the Mascot algorithm (v2.5.1) against a protein database constructed by combining the bovine reference database (Uniprot.com, 37,941 entries, downloaded 22 May 2019), the human reference database (Uniprot.com, 73,911 entries, downloaded 22 May 2019), the *GFP-RNASEH2A* protein, and a contaminant database (cRAP, downloaded 21 November 2016 from <http://www.thegpm.org>) via Proteome Discoverer 2.1. A 1% false discovery rate (FDR) (“High Confidence”) was used for both the peptides and the proteins. At least 3 peptide sequences for the *RNASEH2A* protein were considered for the analysis.

Spectral count data were analyzed for mass spectrometry runs on two sets of immunoprecipitation experiments described previously in the methods. Counts in each run were normalized based on mean of total spectral counts in all runs in a respective set of the experiment. We then filtered hits specific to *Homo sapiens*, added a count of 1 and log₂ transformed the spectral counts. We then used these transformed data to get differentially present protein candidates in immunoprecipitation product of HEK293 with *GFP-RNASEH2A* vs. HEK293 with GFP using linear fitting and empirical Bayes method from limma, a bioconductor package [33–35]. To get the final list of protein candidates, we used the filtering criteria of log fold change > 1.5 and FDR adjusted *p* value < 0.05.

2.10. RNASEH2A Correlation with Cancer Progression and Cell Cycle Related Genes in CCLE and TCGA Pan Cancer Dataset

To validate the results of our study, we further made use of 2 large cancer related datasets to investigate the correlations between *RNASEH2A* and 39 other genes, including common cancer proliferative markers with few genes involved in each cell cycle phase as well as genes upregulated and downregulated in cancer as controls. We also included the genes for the proteins we identified as putative binding partners from mass spectrometry analysis. We used Broad Institute Cancer Cell Line Encyclopedia (CCLE) dataset containing

RNA-Seq expression levels in Reads Per Kilobase of transcript, per Million mapped reads (RPKM) in 1019 cancer cell lines from 26 different tissues of origin [36]. The second dataset we used was the TCGA Pan Cancer Study containing batch normalized RNA-Seq expression levels quantified by RNA-Seq by Expectation Maximization (RSEM) in 32 studies from 10,071 cancer patients. [29]. We added a count of 1 to the values and log transformed the expression counts in both the datasets. These transformed data were used to draw correlations between gene expressions with rcorr function from Hmisc package in R [37]. We also used hierarchical clustering the group the genes in the correlation plot.

3. Results

3.1. RNASEH2A Is Co-Expressed with Genes that Function in Cell Cycle Regulation

To study *RNASEH2A* function based on co-expressed genes, we used an approach (Figure 1) that utilizes data obtained from GTEx database [3] consisting of RNA-seq output of 42,548 nuclear and mitochondrial transcripts including coding, non-coding, and isoforms (raw data in Table S1).

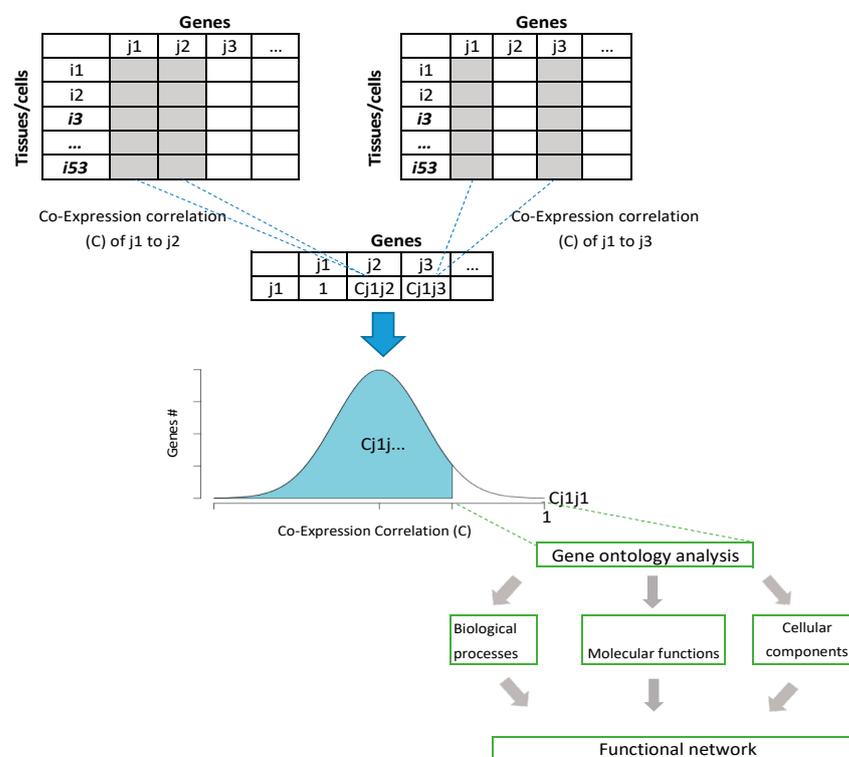


Figure 1. Co-expression correlation analysis approach. Diagram showing the process of analyzing gene functional networks using co-expression analysis. The analysis is based on any database that shows expression levels on multiple tissues and cells from the same organism.

After analysis of the top 2% genes (851) with highest correlation coefficient, using the GOrilla analysis tool [25], we observed a significant enrichment of genes involved in the biological processes of cellular response to DNA damage stimulus ($p = 2.77 \times 10^{-5}$) and DNA replication ($p = 2.66 \times 10^{-4}$), compatible with the known functions of *RNASEH2A*. In addition, we found enrichment of genes involved in mitotic cell-cycle process ($p = 2.80 \times 10^{-14}$), regulation of microtubule cytoskeleton organization ($p = 1.39 \times 10^{-7}$), and chromosome segregation ($p = 9.00 \times 10^{-4}$) (Table 1 and full data in Table S2). For molecular function, we found enrichment of catalytic activity on DNA ($p = 9.59 \times 10^{-6}$) but also microtubule binding ($p = 4.25 \times 10^{-5}$) and kinase binding ($p = 3.28 \times 10^{-4}$) (Table 1 and full data in Table S2). For cellular component, we found gene enrichment in chromosomal part (4.06×10^{-8}), spindle pole ($p = 1.13 \times 10^{-6}$), as well as nucleoplasm ($p = 3.52 \times 10^{-6}$), midbody ($p = 5.01 \times 10^{-6}$), microtubule organizing center ($p = 5.13 \times 10^{-6}$), and condensed

chromosomes outer kinetochore ($p = 8.72 \times 10^{-6}$) (Table 1 and full data in Table S2). Based on these results, we suggest that *RNASEH2A* is involved, among others, in three main functional networks related to DNA replication, DNA damage repair, and regulation of chromosome segregation in mitosis.

Table 1. Gene ontology (GO) term analysis of the top 2% co-expressed genes of *RNASEH2A* reveals the possible role of *RNASEH2A* in cell cycle regulation. The analysis shows *RNASEH2A* functional networks based on process (blue panel), function (orange panel), and component (green panel). The analysis was made by Gene Ontology enRichment anaLysis and visualiZation tool (GORilla). Statistic is shown by both *p*-value and false discovery rate (FDR).

Process			
GO Term	Description	<i>p</i> -Value	FDR <i>q</i> -Value
GO 1903047	mitotic cell cycle process	2.80×10^{-14}	5.93×10^{-11}
GO 0070507	regulation of microtubule cytoskeleton organ	1.39×10^{-7}	8.40×10^{-5}
GO 0006974	cellular response to DNA damage stimulus	2.77×10^{-5}	5.86×10^{-3}
GO 0007059	chromosome segregation	9.00×10^{-5}	1.36×10^{-2}
GO 0006260	DNA replication	2.66×10^{-4}	3.12×10^{-2}
Function			
GO Term	Description	<i>p</i> -Value	FDR <i>q</i> -Value
GO 0140097	catalytic activity, acting on DNA	9.59×10^{-6}	9.06×10^{-3}
GO 0008017	microtubule binding	4.52×10^{-5}	1.42×10^{-2}
GO 0019900	kinase binding	3.28×10^{-4}	7.74×10^{-2}
Component			
GO Term	Description	<i>p</i> -Value	FDR <i>q</i> -Value
GO 0044427	chromosomal part	4.06×10^{-8}	1.36×10^{-5}
GO 0000922	spindle pole	1.13×10^{-6}	2.52×10^{-4}
GO 0005654	nucleoplasm	3.52×10^{-6}	5.88×10^{-4}
GO 0030496	midbody	5.01×10^{-6}	6.70×10^{-4}
GO 0005815	microtubule organizing center	5.13×10^{-6}	5.72×10^{-4}
GO 0000940	condensed chromosome outer kinetochore	8.72×10^{-6}	7.28×10^{-4}
GO 0000796	condensin complex	8.71×10^{-5}	4.16×10^{-3}

To validate the correlation of transcripts in GTEx tissues with *RNASEH2A* at the protein expression level using the STRING protein–protein association network, we looked at 10 transcripts having high correlation with *RNASEH2A* (including *RNASEH2A*) that also have at least 20 known and characterized co-expressed proteins in the STRING network (full list of proteins in Table S3a). Out of the list of 200 proteins (Table S3a), 97 were unique (Table S3b), out of which 87 have Pearson’s correlation coefficient, $R > 0.696$ falling in the top five percentile (Table S3c), and 68 have Pearson’s correlation coefficient, $R > 0.842$ falling in the top one percentile (Table S3c) of correlation coefficient distribution previously represented in Figure 1.

Analysis of the top 2% co-expressed correlated genes with *RNASEH1* revealed functional network of genes that are involved in RNA binding (Table S4 for functional network analysis and Table S5 for complete list of *RNASEH1* co-expressed genes). After filtering top 5% correlated genes, we could identify processes such as regulation of telomerase RNA and telomeres maintenance (Table S4), which were previously reported [38,39]. When we analyzed the top 2% co-expressed correlated genes with *RNASEH2B*, we found the functional network of genes involved in DNA replication (Table S6 for functional network

analysis and Table S7 for complete list of *RNASEH2B* co-expressed genes), as expected for this RNase H2 subunit. For *RNASEH2C*, we could not find any functional network either in the 2% or the 5% analyses, as there were no GO terms with an enrichment p -value below the specified value $p < 0.001$ (data not shown).

3.2. *RNASEH2A* Expression Is Increased in Actively Cycling Cells and Tissues

Next, we studied the expression levels of *RNASEH* genes using the GTEx v.7 database, which allowed us to compare the expression levels of *RNASEH1*, *RNASEH2A*, *RNASEH2B*, and *RNASEH2C* genes in 53 different human tissues, including two cell lines (Table 2). We noticed that tissues that are part of the reproduction system (such as testis, uterus, and cervix) as well as transformed lymphocytes and transformed fibroblasts tend to have a higher expression level of *RNASEH* genes, while low levels were found mainly in tissues with low proliferation and cell turnover capacity, such as most regions of the brain, the kidney, and the heart (Table 2).

Table 2. *RNASEH2A* is highly expressed in human proliferative tissue. Expression of *RNASEH* genes in different 53 human tissues. Data obtained from the Genotype-Tissue Expression portal (GTEx) consist of the median transcripts per million (TPM) from >90 individuals for each tissue. Blue color indicates low expression levels and red color indicates high expression levels.

	<i>RNASEH1</i>	<i>RNASEH2A</i>	<i>RNASEH2B</i>	<i>RNASEH2C</i>		<i>RNASEH1</i>	<i>RNASEH2A</i>	<i>RNASEH2B</i>	<i>RNASEH2C</i>
Liver	5.193	3.947	3.878	11.57	Adipose-Subcutaneous	13.845	15.96	16.585	30.145
Muscle-Skeletal	11.72	5.9535	5.51	9.215	Lung	12.92	14.51	19.34	30.99
Heart-Left Ventricle	7.251	4.262	8.466	12.71	Breast-Mammary Tissue	12.93	16.39	16.65	33.895
Pancreas	5.3175	6.44	10.175	11.48	Pituitary	8.271	11.02	14.5	52.95
Whole Blood	2.726	7.958	6.471	16.33	Vagina	12.08	20.33	16.85	39.11
Brain-Putamen (basal ganglia)	4.6985	5.5965	7.875	15.635	Esophagus-Muscularis	14.91	8.687	22.84	46.885
Brain-Caudate (basal ganglia)	5.1025	6.1935	9.2325	17.175	Artery-Tibial	16.91	8.638	17.8	50.75
Brain-Substantia nigra	6.0725	7.4295	8.123	16.355	Thyroid	12.615	16.935	24.905	43.205
Brain-Amygdala	5.0275	7.4755	8.2985	17.285	Esophagus-Gastroesophageal Junction	13.75	8.3635	22.05	53.92
Brain-Hippocampus	5.907	6.655	7.567	18.76	Bladder	15.4	15.95	23.68	43.32
Brain-Anterior cingulate cortex (BA24)	5.921	6.326	9.554	19.13	Artery-Coronary	13.65	10.5	18.58	56.4
Brain-Hypothalamus	6.903	7.451	8.639	18.47	Artery-Aorta	14.63	9.235	17.31	58.05
Kidney-Cortex	6.303	6.632	6.614	22.18	Prostate	11.11	11.475	20.27	57.625
Brain-Nucleus accumbens (basal ganglia)	6.557	7.162	12.03	20.17	Nerve-Tibial	15.18	19.24	19.55	48.22
Heart-Atrial Appendage	8.609	5.969	9.644	22.33	Fallopian Tube	14.57	13.31	25.31	49.21
Brain-Cortex	5.5355	6.8145	11.095	25.555	Colon-Sigmoid	13.24	10	24.46	55.42
Brain-Frontal Cortex (BA9)	7.597	6.766	13.7	23.46	Cells-Transformed fibroblasts	29.98	34.04	17.95	25.17
Stomach	8.19	9.396	12.55	24.365	Brain-Cerebellum	9.384	16.21	12.41	70.97
Brain-Spinal cord (cervical c-1)	9.183	13.54	10.54	22.75	Spleen	10.935	25.695	24.555	49.11
Skin-Sun Exposed (Lower leg)	11.62	18.75	10.87	19.18	Ovary	12.4	19.46	20.07	59.51
Skin-Not Sun Exposed (Suprapubic)	10.72	18.3	11.7	19.98	Cervix-Ectocervix	14.765	24.625	23.54	51.435
Minor Salivary Gland	10.33	13.21	13.03	24.15	Brain-Cerebellar Hemisphere	13.325	17.52	14.98	68.75
Esophagus-Mucosa	10.58	24.62	11.92	18.2	Cervix-Endocervix	15.15	19	26.62	55
Colon-Transverse	9.484	12.52	14.885	28.96	Testis	14.12	49.84	23.42	31.66
Adrenal Gland	10.045	9.7865	12.39	35.965	Uterus	16.6	21.33	30.43	66.03
Adipose-Visceral (Omentum)	12.25	12.14	14.71	32.23	Cells-Epstein-Barr virus (EBV)-transformed lymphocytes	27.94	79.945	41.36	32.195
Small Intestine-Terminal Ileum	9.633	14.12	18.32	30.46					

We then examined the expression level of the *RNASEH* genes in cancer using data obtained from Cancer RNA-seq Nexus (CRN) [27] in 29 different randomly selected tumors

in different tissues at different stages of cancer progression compared to their healthy controls (complete list of tumors in Table S8 and raw data in Table S9). To validate our approach, we also examined the expression level of *MYBL2* and *SCARA5* genes, which are reported to be upregulated and downregulated in cancer, respectively [40]. As expected, *MYBL2* was mostly upregulated in cancers such as triple negative breast cancer, bladder urothelial carcinoma stages 2 and 4, and lung adenocarcinoma stages 2 and 4, while *SCARA5* was downregulated in cancers such as bladder urothelial carcinoma stages 2 and 4; thyroid carcinoma stage 1, and rectum adenocarcinoma stage 2A (Figure 2). *RNASEH2A* was the only *RNASEH* gene that showed increased expression levels in all 29 tumors relative to the control tissues (Figure 2).

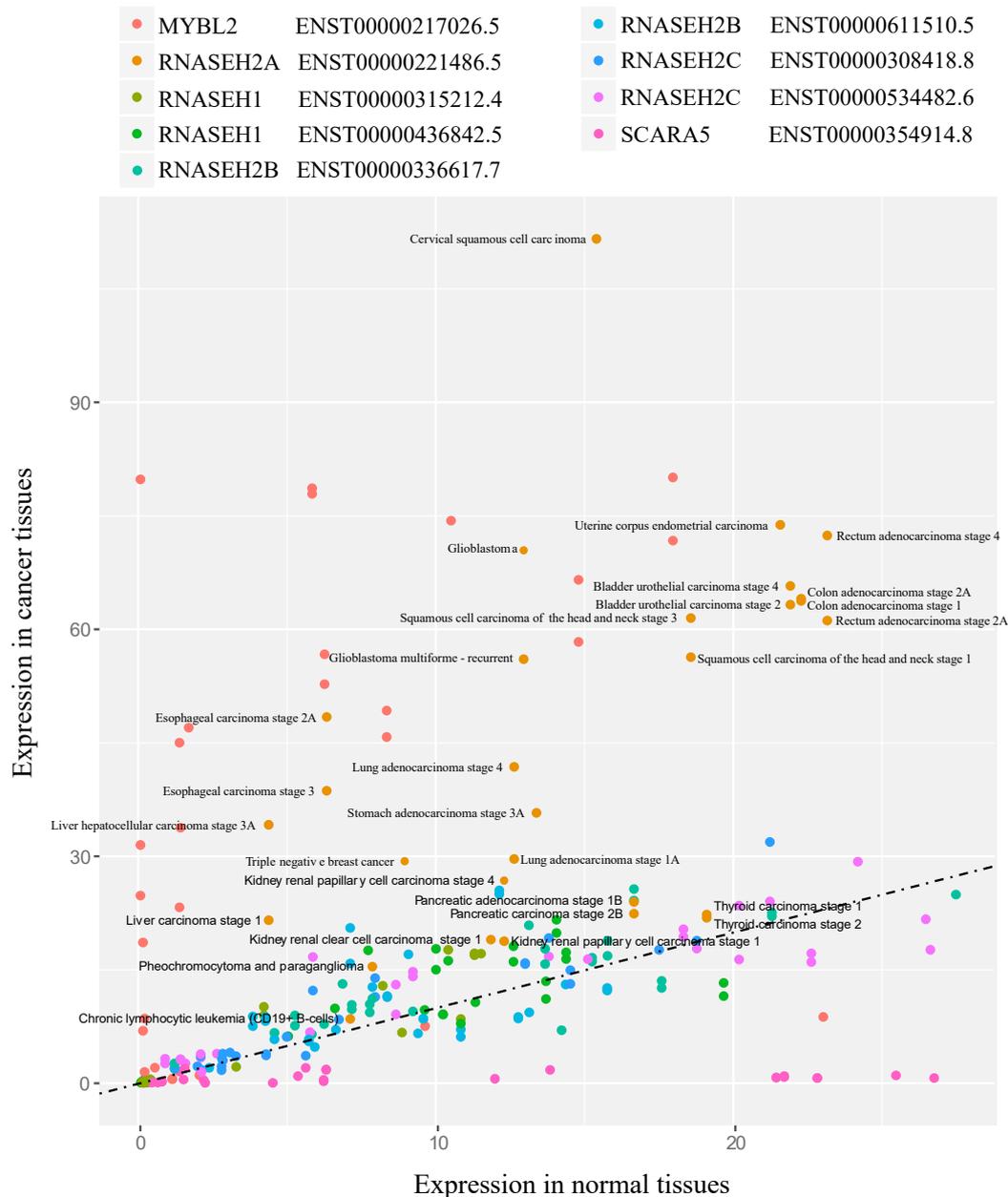


Figure 2. *RNASEH2A* gene expression in cancer tissues. Expression levels in TPM and only for triple negative breast cancer and chronic lymphocytic leukemia in fragments per kilobase million (FPKM) in 29 different cancer tissues compared to non-cancerous tissue controls. Data obtained from Cancer RNA-seq Nexus (CRN). Regression line represents equal expression levels in cancer and normal tissues.

To study the levels of *RNASEH* genes throughout the progression of cancer, we examined the expression levels of the genes at different stages of lung squamous adenocarcinoma, breast invasive carcinoma, and bladder urothelial carcinoma. In all these cancers, the expression of *RNASEH2A* changed drastically from a normal stage to early cancer stages and remained high with the progression of the disease (Figure 3 and Table S10 and raw data in Table S11). No other gene beside *MYBL2* was upregulated at the different stages studied. These results imply that *RNASEH2A* might have a role in cancer etiology or transformation from normal tissue to cancer tissue, that is dependent on its expression level.

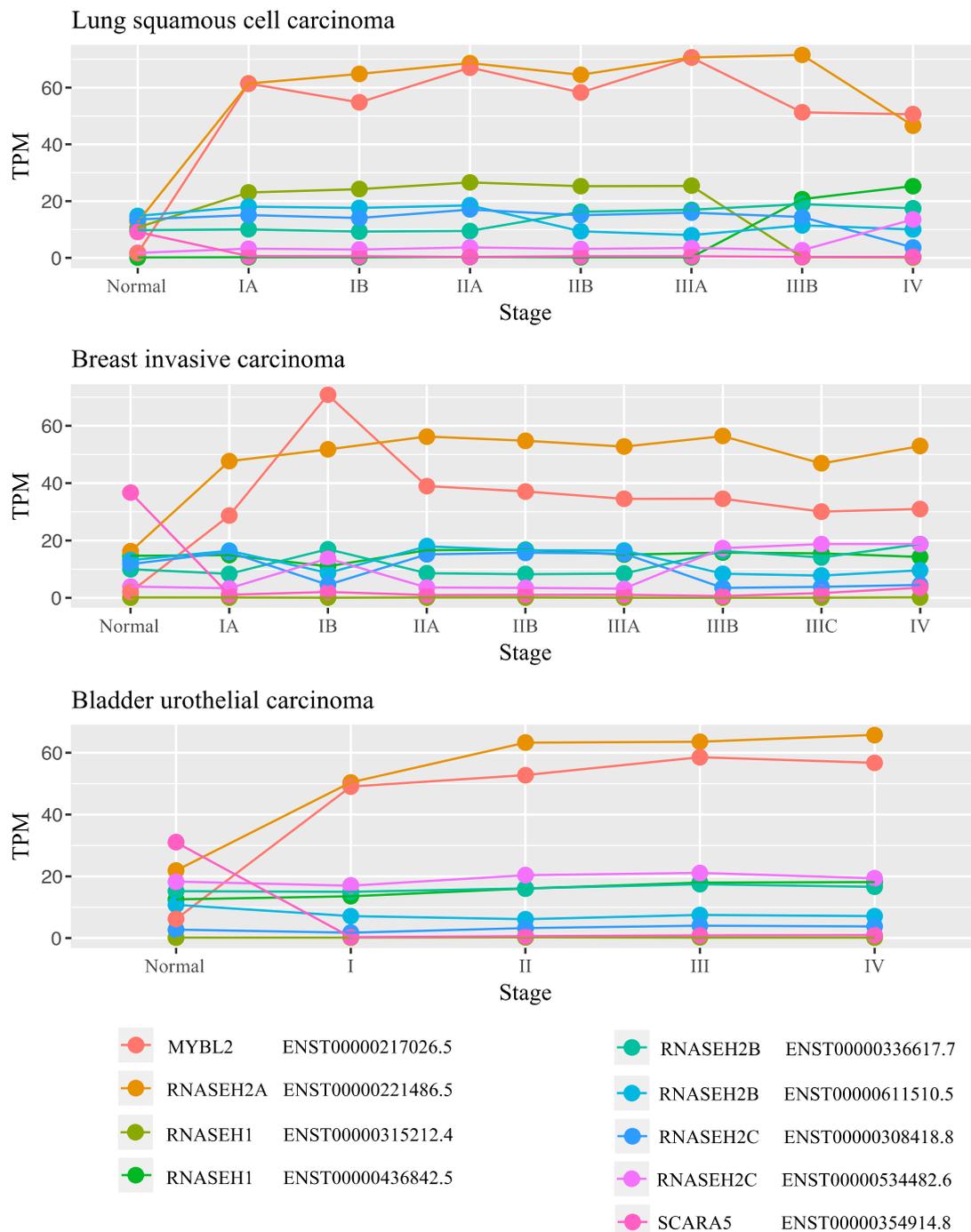


Figure 3. *RNASEH2A* levels in normal and at different stages of lung squamous cell carcinoma, breast invasive carcinoma, bladder urothelial carcinoma (top to bottom, respectively). Expression levels of different transcripts of *RNASEH* genes, including *MYBL2* and *SCARA5*, were examined in in TPM.

3.3. RNASEH2A Gene Amplifications Has Higher Prevalence in Multiple Cancer Types

The analysis of *RNASEH2A* copy number alterations and mRNA expression in various types of cancer (prevalence of *RNASEH2A* CNAs in Table 3, expression summary for CNA types and raw data in Table S12) revealed that the average expression of *RNASEH2A* is higher in patients with amplifications of the *RNASEH2A* gene and lower in patients with deep deletions when compared to normal diploid samples. This correlates with previous findings in mRNA versus CNA type expression levels in the TCGA Pan Cancer dataset [41]. We also observed that, in 16 out of 35 cancer types/subtypes, percentage of patients with amplifications was higher than percentage of patients with deep deletions, with 10% cancer types having >1% amplifications as copy number alteration. The maximum percentage of patients having *RNASEH2A* amplifications in a given type of cancer was 8.14% for ovarian epithelial tumor, followed by 3.29% in endometrial and 2.63% in adrenocortical carcinoma (Table 3 and raw data in Table S12).

Table 3. Percentage of patients with *RNASEH2A* copy number alterations from Genomic Identification of Significant Targets in Cancer (GISTIC) in The Cancer Genome Atlas (TCGA) Pan Cancer cohorts.

Cancer Type	Percentage of Patients with Copy Number Alterations				
	Deep Deletion	Shallow Deletion	Diploid	Gain	Amplification
Ovarian Epithelial Tumor	0.34%	29.15%	26.10%	36.27%	8.14%
Endometrial Carcinoma	-	12.13%	69.67%	14.90%	3.29%
Adrenocortical Carcinoma	-	1.32%	34.21%	61.84%	2.63%
Pleural Mesothelioma	-	8.05%	73.56%	16.09%	2.30%
Esophageal Squamous Cell Carcinoma	-	34.04%	44.68%	19.15%	2.13%
Cervical Squamous Cell Carcinoma	0.41%	26.23%	57.79%	13.52%	2.05%
Diffuse Glioma	0.20%	3.33%	73.92%	20.78%	1.76%
Sarcoma	-	10.36%	50.20%	37.85%	1.59%
Invasive Breast Carcinoma	0.09%	21.25%	60.11%	17.23%	1.31%
Ocular Melanoma	-	3.75%	92.50%	2.50%	1.25%
Thymic Epithelial Tumor	-	2.52%	94.12%	2.52%	0.84%
Esophagogastric Adenocarcinoma	0.62%	33.40%	57.73%	7.42%	0.82%
Head and Neck Squamous Cell Carcinoma	0.20%	19.45%	66.99%	12.57%	0.79%
Glioblastoma	-	8.11%	58.11%	33.11%	0.68%
Non-Small Cell Lung Cancer	0.30%	44.10%	43.39%	11.71%	0.50%
Bladder Urothelial Carcinoma	-	29.21%	53.71%	16.58%	0.50%
Prostate Adenocarcinoma	0.20%	5.74%	92.21%	1.64%	0.20%
Cervical Adenocarcinoma	2.17%	26.09%	65.22%	6.52%	-
Cholangiocarcinoma	-	13.89%	72.22%	13.89%	-
Colorectal Adenocarcinoma	0.51%	11.02%	71.86%	16.61%	-
Encapsulated Glioma	-	-	100.00%	-	-
Fibrolamellar Carcinoma	-	-	100.00%	-	-
Hepatocellular Carcinoma	0.28%	21.51%	64.25%	13.97%	-
Leukemia	-	1.81%	95.18%	3.01%	-
Mature B-Cell Neoplasms	-	2.08%	91.67%	6.25%	-
Melanoma	-	22.62%	61.31%	16.08%	-
Miscellaneous Neuroepithelial Tumor	-	-	96.30%	3.70%	-
Non-Seminomatous Germ Cell Tumor	-	52.33%	33.72%	13.95%	-
Pancreatic Adenocarcinoma	-	14.77%	77.27%	7.95%	-
Pheochromocytoma	-	0.75%	84.33%	14.93%	-
Renal Clear Cell Carcinoma	-	2.17%	88.56%	9.27%	-

Table 3. Cont.

Cancer Type	Percentage of Patients with Copy Number Alterations				
	Deep Deletion	Shallow Deletion	Diploid	Gain	Amplification
Renal Non-Clear Cell Carcinoma	-	5.17%	87.36%	7.47%	-
Seminoma	-	1.59%	55.56%	42.86%	-
Undifferentiated Stomach Adenocarcinoma	-	25.00%	58.33%	16.67%	-
Well-Differentiated Thyroid Cancer	-	1.21%	97.37%	1.41%	-
All Cancer types	0.17%	17.69%	66.83%	14.30%	1.01%

CNAs are categorized as: 1. Deep deletion (−2) indicating a deep loss/homozygous deletion; 2. Shallow deletion (−1) indicates shallow loss/heterozygous deletion; 3. Diploid (0) indicates homozygous genes; 4. Gain indicates a low-level gain (a few additional copies, often broad); 5. Amplification (2) indicate a high level amplification (more copies, often focal).

3.4. Mass Spectrometry Analysis Identified RNASEH2A Binding Partners Involved in Mitosis Regulation

To identify proteins that interact with *RNASEH2A*, we over-expressed the control eGFP (Figure 4a, lanes 1–3) and the eGFP-*RNASEH2A* (Figure 4a, lanes 4–6) proteins in HEK293 cells. We performed co-immunoprecipitation (Co-IP) using an anti-GFP antibody (Figure 4a). We also confirmed that both *RNASEH2A* (Figure 4b, lanes 1–2) and GFP (Figure 4b, lanes 3–4) antibodies identify the same GFP-*RNASEH2A* protein (Figure 4b). After confirming the successful pull-down of *RNASEH2A* protein, we analyzed the eluted proteins that were interacting with *RNASEH2A* via mass spectrometry. The analysis revealed a short list of five human proteins with at least three peptide spectrum matches (PSMs) interacting with *RNASEH2A* (Table 4 and full data in Table S13a,b). *RNASEH2B* and *RNASEH2C* were interacting with *RNASEH2A* confirming the validity of the Co-IP experiment. Interestingly, among the proteins interacting with *RNASEH2A*, we identified the T-complex protein 1 subunits theta Chaperonin Containing TCP1 Subunit 8 (CCT8) and DnaJ homologue subfamily A member 1 (DNAJA1) and apoptosis-inducing factor mitochondrion-associated 1 (AIFM1) proteins with a gene correlation coefficient with *RNASEH2A* of 0.84, 0.585, and 0.378 respectively. These three potential *RNASEH2A* binding partners have shown to play a role in cell cycle regulation and mitosis [42–45].

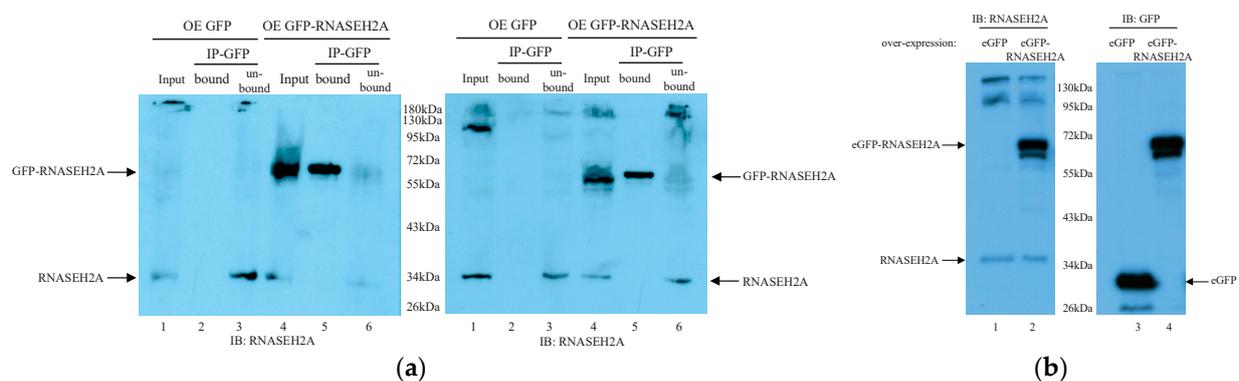


Figure 4. Co-immunoprecipitation and antibodies verification: (a) co-immunoprecipitation of *RNASEH2A*. Western Blot of the GFP co-immunoprecipitation. The two panels indicate two independent experiments. HEK293 cells were overexpressed (OE) with the eGFP vector (lanes 1–3 in each panel) or with the eGFP-*RNASEH2A* vector (lanes 4–6 in each panel). IP: anti-GFP, IB: anti-*RNASEH2A*. Arrows indicate endogenous *RNASEH2A* (lower arrow) and exogenous eGFP-*RNASEH2A* (upper row); (b) *RNASEH2A* and GFP antibodies verification. Western Blot of the anti-*RNASEH2A* (left panel) and the anti-GFP (right panel). In the two panels, HEK293 cells were OE with eGFP vector (lanes 1, 3) or eGFP-*RNASEH2A* vector (lanes 2, 4). In the left panel, arrows indicate endogenous *RNASEH2A* (lower arrow) and exogenous eGFP-*RNASEH2A* (upper row). In the right panel, arrows indicate eGFP (lower arrow) and eGFP-*RNASEH2A* (upper row).

Table 4. List of proteins identified as binding partners of *RNASEH2A* by mass spectrometry analysis. The two top proteins, *RNASEH2B* and *RNASEH2C*, confirm the authenticity of the results.

Protein	Gene	Description
Q5TBB1	<i>RNASEH2B</i>	Ribonuclease H2 subunit B
E9PN81	<i>RNASEH2C</i>	Ribonuclease H2 subunit C
P31689	<i>DNAJA1</i>	DnaJ homolog subfamily A member 1
O95831	<i>AIFM1</i>	Apoptosis-inducing factor 1, mitochondrial
P50990	<i>CCT8</i>	T-complex protein 1 subunit theta

3.5. *RNASEH2A* Expression Positively Correlates with Cancer Proliferation Markers and Cell Cycle Genes

To further explore the involvement of *RNASEH2A* in cancer, we examined the correlations between *RNASEH2A* and 39 additional genes in CCLE and TCGA Pan Cancer datasets (Table 5 and Table S14a,b). The correlation analysis uncovered a positive correlation of *RNASEH2A* with genes up-regulated in cancer and a negative correlation with genes that are instead down-regulated. In addition, *RNASEH2A* showed a higher correlation with cluster of genes known to play a role in cancer proliferation as compared to *RNASEH2B*, *RNASE2C*, and *RNASEH1*. We also observed a similar trend for the three putative binding partners of *RNASEH2A* that we identified by mass spectrometry, showing a correlation comparable with that found in the GTEx dataset. The analysis of cell cycle related genes showed a high correlation between *RNASEH2A* and genes involved in Gap 1/Synthesis (G1/S), Gap 2 (G2), G2/Mitosis (M), M, and M/G1 phases and a comparatively lower or no correlation with genes unique to G1 and S phases (Figure 5 and raw data in Figure S15a,b).

Table 5. List of genes related to cancer proliferation and cell cycle phases used for correlation analysis in Cancer Cell Line Encyclopedia (CCLE) and TCGA Pan Cancer datasets.

Gene Name (s)	Function/Role	References
<i>CDKN2A, CDKN2B, CCND1</i>	G1 Cell Cycle Phase	A. Subramanian et al. Gene Set Enrichment Analysis (GSEA Database) [46]
<i>DHFR, CCNE1</i>	G1/S Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>AKT1-3, E2F4-5</i>	S Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>CDKN2D, MDM2</i>	G2 Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>CCNB2, TOPBP1</i>	G2/M Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>APC, BUB1</i>	M Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>E2F2, E2F3, CCNB1</i>	M/G1 Cell Cycle Phase	A. Subramanian et al. (GSEA Database) [46]
<i>MYBL2, FOXM1, BUB1, AURKA, AURKB</i>	Upregulated in cancer	M. Li et al. [40]
<i>SCARA5, MYO1</i>	Downregulated in cancer	M. Li et al. [40]
<i>PCNA, MKI67(Ki67), MCM2–MCM6, E2F1</i>	Proliferative markers in cancer	M.L. Whitfield et al. [47]
<i>CCNE1, CCND1, CCNB1</i>	Cell cycle markers associated with cancer (G1/S, G2, and M)	M.L. Whitfield et al. [47]
<i>RNASEH2A, RNASEH2B, RNASEH2C, RNASEH1</i>	Target genes in this study	This study
<i>CCT8, DNAJA1, AIFM1</i>	Predicted Binding partners of <i>RNASEH2A</i>	This study

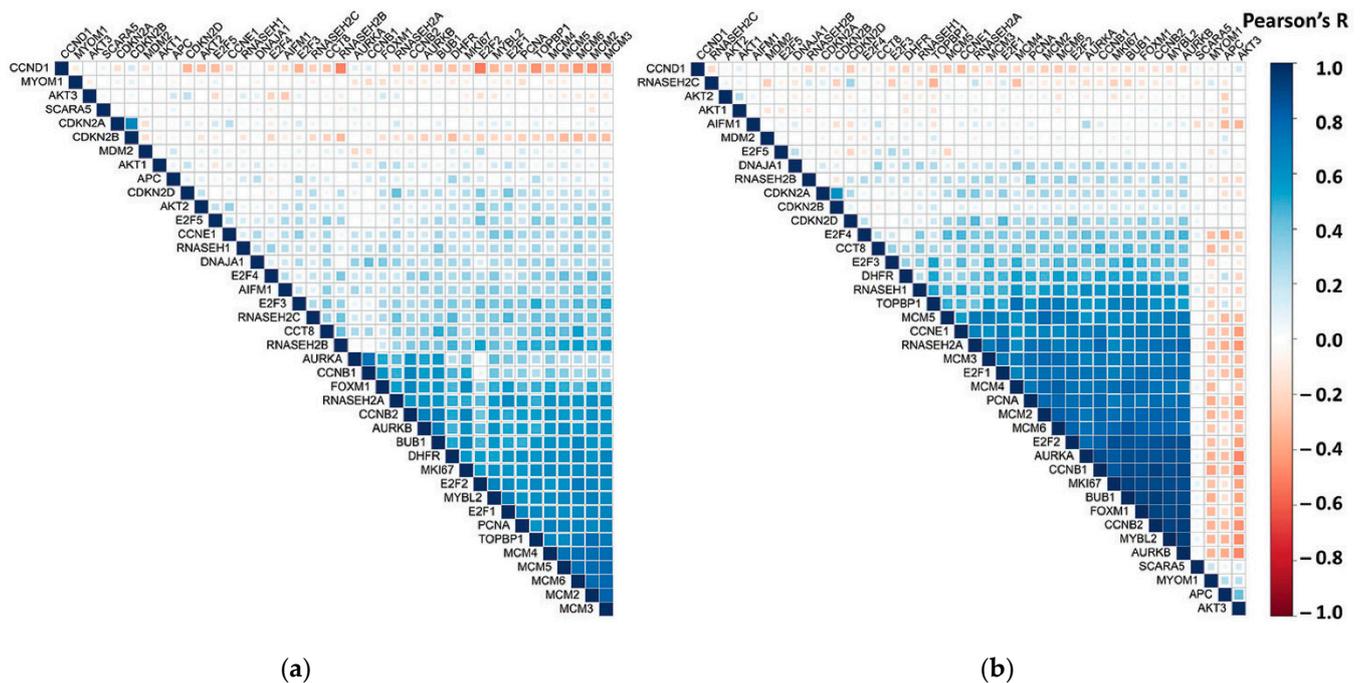


Figure 5. *RNAES2A* correlation with 39 other genes previously listed in CCLE and TCGA Pan Cancer datasets: (a) correlation coefficient plot generated from RNA-Seq expression levels in 1019 cancer cell lines from CCLE database measured in RPKM ($\log_2(\text{value} + 1)$); (b) correlation coefficient plot generated from RNA-Seq expression levels in 10,071 cancer tissue samples from the Cancer Genome Atlas (TCGA) Pan Cancer studies measured in RSEM (Bach normalized from Illumina HiSeq_RNASeqV2($\log_2(\text{value} + 1)$)). The proportion of shaded area in each cell corresponds to the Pearson's R.

4. Discussion

Co-expression correlation analysis of genes is a simple approach to suggest functional network of genes. We anticipate that this approach is more accurate for analyzing genes sharing a higher correlation value. In addition, we assume that the genes analyzed should have a functional network involved in a process that is dependent on a spatio-temporal gene expression profile. Cell cycle regulation is a good example for such a process. In fact, the analysis of *CDK1*, a regulator of cyclin B implicated in cell cycle control, shows enrichment of genes involved in mitotic cell cycle regulation, microtubule cytoskeleton organization, and regulation of G1/S transition of mitotic cell cycle (Table S16 and complete list of *CDK1* co-expressed genes in Table S17). This approach can be applied to other processes such as, for example, RNA regulation in stress granules, which is a well-orchestrated process dependent on time and stress conditions. In this context, *G3BP1*, a RNA helicase that is one of the key assemblers of the stress granules, shows enrichment of co-expressed genes involved in molecular function of RNA binding and RNA helicase activity in a ribonucleoprotein cellular component (Table S16 and complete list of *G3BP1* co-expressed genes in Table S18).

Among the three *RNASEH* genes, we found that only *RNASEH2A* was involved in the functional network of mitotic cell cycle regulation. One possibility is that *RNASEH2A* has a function independent of *RNASEH2B* and *RNASEH2C*, as its distribution in the cell was shown not to be limited to the interaction with *RNASEH2B* and *RNASEH2C* [6]. This notion was also suggested by others showing that knocking out *RNASEH2A* in cervical cancer HeLa cells results in an increased sensitivity to ataxia telangiectasia and Rad3-related (ATR) inhibitors compared to knocking out *RNASEH2B* [48]. We propose that *RNASEH2A* levels constitute an additional layer of regulation of *RNASEH2A* activity, which is not dictated by *RNASEH2B* or *RNASEH2C* levels, and this is why the co-expression correlation analysis did not show a functional network of mitotic cell regulation for *RNASEH2B* and *RNASEH2C* genes. Following this notion, it would be intriguing to examine the levels of the

three RNase H2 genes throughout the cell cycle in human cells and study how perturbing their levels affects RNase H2 function and/or cell cycle progression. Similar studies have been performed on yeast showing a cell cycle dependent expression of *RNH201* (orthologous gene of *RNASEH2A*), peaking at S and G2 phases [23,49], and being recruited to chromatin/telomeres at the G2 phase [49].

Rejins et al. [50], using an anti-mouse antibody for all proteins of the RNase H2 complex, demonstrated an increase in the expression level of RNase H2 in the mouse blastocyst in all three embryonic layers during gastrulation and showed, in newborns and adults, that the expression becomes restricted to highly proliferative tissues such as intestinal crypt epithelium and testes [50]. Moreover, they reported that RNase H2 levels correlate with the proliferation marker Ki67 [50]. These data are consistent with our findings showing an increase in *RNASEH2A* level in actively cycling tissues as well as different cancer tissues compared to normal ones.

Using our co-expression correlation approach, we identified three main functional networks in which *RNASEH2A* is involved: DNA replication, DNA damage repair, and regulation of chromosome segregation in mitosis. Hyper-proliferation of cells results in excessive addition of RNA Okazaki fragments into the genome during DNA replication, increases the number of events when replicative DNA polymerases insert ribonucleotides instead of deoxyribonucleotides into the genomic DNA [51,52], and increases the rate of chromosomes reduction from 4N to 2N by symmetric segregation into two new daughter cells. All these events correspond to the functional networks identified in this study for *RNASEH2A* and link *RNASEH2A* activity to cell cycle regulation. In support of this notion, we identified several cellular pathways among the highest *RNASEH2A* co-expressed genes such as the minichromosome maintenance (MCM) complex, a replicative eukaryotic helicase that, in yeast, has been demonstrated to be activated upon accumulation of RNA-DNA hybrids at the G2-M checkpoint. This finding suggests that targeting the *RNASEH2A* level and/or activity could prevent the DNA damage occurring during replication, which leads to mitotic catastrophe and cell death [53,54]. The other identified complexes, such as the *NDC80* complex, the *CENP* family, and the *KIF* family, have all been shown to function in cell cycle regulation through regulation of the microtubule filaments with the kinetochore [55–57].

Our analyses of expression levels focused on the *RNASEH* genes (*RNASEH1*, *RNASEH2A*, *RNASEH2B*, and *RNASEH2C*) using the GTEx v.7 database and revealed a marked contrast in *RNASEH2A* expression levels between low and high proliferative tissues. *RNASEH2A* showed low expression levels in low proliferative tissues such as brain, kidney (cortex), and heart and high expression levels in highly proliferative tissues such as skin, esophagus, small intestine, testis, cervix, as well as in transformed fibroblasts and transformed lymphocytes. Moreover, in line with the initial study by Flanagan et al. [18], showing elevated expression levels of *RNASEH2A* in transformed mesenchymal stem cells, our analysis of 29 different randomly selected tumors in different tissues compared to their healthy controls revealed that *RNASEH2A* is the only *RNASEH* gene displaying increased expression levels. Further analyses of *RNASEH* expression levels throughout different stages of lung squamous adenocarcinoma, breast invasive carcinoma, and bladder urothelial carcinoma again highlighted specific upregulation of *RNASEH2A* from normal tissues to early stages and more advanced stages of these cancers. It is interesting to note that Dai et al. reported on a possible role of *RNASEH2A* in glioma cell proliferation. Their findings suggested a role of *RNASEH2A* upregulation in cell growth and apoptosis, contributing to glioma-genesis and cancer progression [58]. More recently, *RNASEH2A* overexpression was associated with cancer cell resistance to chemotherapy in vitro and with aggressiveness and poor outcomes in breast cancers of estrogen receptor (ER)-positive subtypes [59]. The analysis of copy number alterations of the *RNASEH2A* gene that we performed in our study further supports the role of *RNASEH2A* in cancer development. We demonstrated that *RNASEH2A* gene amplifications have higher incidence in multiple cancer types when compared to deep deletions and normal diploid samples suggesting *RNASEH2A* as a target for cancer diagnosis and therapy. Interestingly,

cancers from tissues with high cell turnover, such as the reproductive system, showed a maximum percentage of patients with *RNASEH2A* amplification, suggesting a possible role of adult stem cells in the overexpression of *RNASEH2A*. In fact, deregulation of the adult stem cell niche is considered a key event in the etiology of cancer [60].

Our mass spectrometry analysis shows five protein candidates that specifically interact with *RNASEH2A*. The top two proteins observed were *RNASEH2B* and *RNASEH2C*, confirming the authenticity of the results. Additional proteins that we identified in the context of *RNASEH2A* biology and mitosis regulation were *CCT8*, *DNAJA1*, and *AIFM1*. *CCT8* is part of the *CCT/TRiC* complex that was reported to regulate telomerase function by mediating its trafficking from the cytoplasm to the telomeres [36]. In addition, it was shown to promote chromosome segregation by disassembling checkpoint complex from sister chromatids in the initial phase of mitosis [37], supporting the *RNASEH2A* functional network in mitosis regulation. *DNAJA1* belongs to the family of DnaJ heat shock protein family (Hsp40); it has been shown to be activated by E2F transcription factor 1 (E2F1) and to promote cell cycle by stabilizing cell division protein 45 (CDC45) [38]. Of interest, the induction of *RNASEH2A* by E2F1 has been also reported in human papillomavirus cervical cancers [19]. *AIFM1* is a well known factor that plays a key role in the apoptosis signaling pathways. A recent study has uncovered a new role of *AIFM1* on the proliferation of hepatoma cells and cell cycle arrest [39].

The high correlation of *RNASEH2A* expression with genes that play a role in cancer development and cell cycle progression validate our study, showing that *RNASEH2A* plays a role in cancer transformation and progression. In addition, our final analysis supports the mass spectrometry findings and indicates *CCT8*, *DNAJA1*, and *AIFM1* as binding partners of *RNASEH2A*.

The analysis and the experimental approaches used in this work have some limitations. In this respect, we made use of the Cancer RNA-Seq Nexus database for which patient information is not available. Factors such as age, gender, and medical treatment versus surgery could have influenced the results of our in-silico analysis. In addition, although mass spectrometry is a highly specific and sensitive technique, further investigation is needed to confirm the specific interactions between the *RNASEH2A* subunit and the candidate proteins found in our experiments. In our in vitro experiments, we used the HEK293 cell line. Depending on a number of factors, this cell line in culture can either show the phenotype of a normal or of a cancer cell line [61]. Although we were able to support our in vitro finding by in-silico analysis, the characterization of the biological role of *CCT8*, *DNAJA1*, and *AIFM1* in concert with *RNASEH2A* requires a more appropriate model system.

5. Conclusions

Overall, our study suggests an emerging role of *RNASEH2A* in cell cycle regulation that appears independent from its function as part of the RNase H2 whole enzyme. The presented findings stimulate new research directions such as investigating the differential expression of *RNASEH2A* during cell cycle phases to better understand and characterize the function of *RNASEH2A* in cell proliferation in healthy, adult stem cells and cancer cells and to support further exploring *RNASEH2A* as a target for cancer diagnosis and therapy.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2079-7737/10/3/221/s1>. Table S1: List of *RNASEH2A* co-expressed correlated genes. Table S2: Gene ontology of the 2% co-expressed correlated *RNASEH2A* genes using the GOrilla analysis tool. Table S3: Protein network association data obtained from STRING protein–protein association network applied to *RNASEH2A* and its top co-expressed genes (3a–c). Table S4: Gene ontology of the 2% and the 5% co-expressed correlated *RNASEH1* genes using the GOrilla analysis tool showing a functional network of genes involved in RNA binding and processes of regulation of telomerase RNA and telomeres maintenance. Table S5: List of *RNASEH1* co-expressed correlated genes. Table S6: Gene ontology of the 2% co-expressed correlated *RNASEH2B* genes using the GOrilla analysis tool showing

functional network of genes involved in DNA replication. Table S7: List of *RNASEH2B* co-expressed correlated genes. Table S8: Expression level of the *RNASEH* genes in cancer using data obtained from Cancer RNA-seq Nexus (CNR) in 29 different randomly selected tumors. Table S9: Expression level of the *RNASEH* genes in cancer using data obtained from Cancer RNA-seq Nexus (CNR) in different tissues at different stage of cancer progression. Table S10: Level of the *RNASEH* genes throughout the progression of cancer showing constant upregulation of *RNASEH2A* in lung, breast, and bladder tumors. Table S11: Level of the *RNASEH* genes at different stages of cancer. Table S12: (a) Average expression of patients with different *RNASEH2A* copy-number alterations from GISTIC in TCGA Pan Cancer studies; (b) raw data for *RNASEH2A* mRNA expression with copy number alteration type in TCGA Pan Cancer studies. Table S13: Mass spectrometry results of first run (4a) and second run (4b) of immunoprecipitation experiments. Table S14: (a) Count and percentage of Cancer Cell line types in CCLE database; (b) count and percentage of cancer tissue types/subtypes in TCGA Pan Cancer Dataset. Table S15: (a) Correlation Matrix (Pearson's R and p-value) for *RNASEH2A* and 39 subset list of genes in CCLE; (b) correlation Matrix (Pearson's R and p-value) for *RNASEH2A* and 39 subset list of genes in TCGA. Table S16: Gene ontology of the 2% co-expressed correlated *CDK1* genes and *G3BP1* using the GOrilla analysis tool showing enrichment of genes involved in cell cycle regulation and in molecular binding of RNA. Table S17: List of *CDK1* co-expressed correlated genes. Table S18: List of *G3BP1* co-expressed correlated genes.

Author Contributions: Conceptualization, A.T. and F.S.; methodology, A.T. and S.M.; software, A.T., S.M. and D.K.; validation, A.T. and S.M.; formal analysis, A.T., S.M. and D.K.; investigation, A.T. and S.M.; resources, F.S.; data curation, A.T., S.M. and D.K.; writing—original draft preparation, A.T.; writing—review and editing, S.M., D.K. and F.S.; visualization, A.T. and S.M.; supervision, F.S.; project administration, F.S.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Health, NIH, NIGMS R01 GM115927 (F.S.); NIH NIEHS R01ES026243 (F.S.); the National Science Foundation, NSF, MCB-1615335 (F.S.) and the Howard Hughes Medical Institute Faculty Scholar grant 55108574 (F.S.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data of the study are presented within the Supplementary Tables and are also available from the corresponding author.

Acknowledgments: We thank Georgia Institute of Technology's Parker H. Petit Institute for Bioengineering and Biosciences including the Systems Mass Spectrometry Core Facility for supporting this work, C. Meers for critical reading of the manuscript, and all the people of the Storici laboratory for assistance and feedback on this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Burgess, D.J. Gene expression: Principles of gene regulation across tissues. *Nat. Rev. Genet.* **2017**, *18*, 701. [[CrossRef](#)]
2. Sonawane, A.R.; Platig, J.; Fagny, M.; Chen, C.-Y.; Paulson, J.N.; Lopes-Ramos, C.M.; DeMeo, D.L.; Quackenbush, J.; Glass, K.; Kuijjer, M.L. Understanding Tissue-Specific Gene Regulation. *Cell Rep.* **2017**, *21*, 1077–1088. [[CrossRef](#)] [[PubMed](#)]
3. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)] [[PubMed](#)]
4. Cerritelli, S.M.; Crouch, R.J. Ribonuclease H: The enzymes in Eukaryotes. *FEBS J.* **2009**, *276*, 1494–1505. [[CrossRef](#)] [[PubMed](#)]
5. Sparks, J.L.; Chon, H.; Cerritelli, S.M.; Kunkel, T.A.; Johansson, E.; Crouch, R.J.; Burgers, P.M. RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **2012**, *47*, 980–986. [[CrossRef](#)] [[PubMed](#)]
6. Chon, H.; Vassilev, A.; Depamphilis, M.L.; Zhao, Y.; Zhang, J.; Burgers, P.M.; Crouch, R.J.; Cerritelli, S.M. Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. *Nucleic Acids Res.* **2009**, *37*, 96–110. [[CrossRef](#)] [[PubMed](#)]
7. Sollier, J.; Stork, C.T.; García-Rubio, M.L.; Paulsen, R.D.; Aguilera, A.; Cimprich, K.A. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **2014**, *56*, 777–785. [[CrossRef](#)]
8. Hamperl, S.; Bocek, M.J.; Saldivar, J.C.; Swigut, T.; Cimprich, K.A. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **2017**, *170*, 774–786. [[CrossRef](#)] [[PubMed](#)]

9. Meers, C.; Keskin, H.; Storici, F. DNA repair by RNA: Templated, or not templated, that is the question. *DNA Repair* **2016**, *44*, 17–21. [[CrossRef](#)]
10. Wahba, L.; Amon, J.D.; Koshland, D.; Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA: DNA hybrids from generating genome instability. *Mol. Cell* **2011**, *44*, 978–988. [[CrossRef](#)]
11. Cornelio, D.A.; Sedam, H.N.; Ferrarezi, J.A.; Sampaio, N.M.; Argueso, J.L. Both R-loop removal and ribonucleotide excision repair activities of RNase H2 contribute substantially to chromosome stability. *DNA Repair* **2017**, *52*, 110–114. [[CrossRef](#)] [[PubMed](#)]
12. O’Connell, K.; Jinks-Robertson, S.; Petes, T.D. Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. *Genetics* **2015**, *201*, 963–975. [[CrossRef](#)]
13. Zimmer, A.D.; Koshland, D. Differential roles of the RNases H in preventing chromosome instability. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12220–12225. [[CrossRef](#)] [[PubMed](#)]
14. Bartsch, K.; Knittler, K.; Borowski, C.; Rudnik, S.; Damme, M.; Aden, K.; Spehlmann, M.E.; Frey, N.; Saftig, P.; Chalaris, A.; et al. Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. *Hum. Mol. Genet.* **2017**, *26*, 3960–3972. [[CrossRef](#)] [[PubMed](#)]
15. Rice, G.; Patrick, T.; Parmar, R.; Taylor, C.F.; Aeby, A.; Aicardi, J.; Artuch, R.; Montalto, S.A.; Bacino, C.A.; Barroso, B.; et al. Clinical and molecular phenotype of Aicardi-Goutieres syndrome. *Am. J. Hum. Genet.* **2007**, *81*, 713–725. [[CrossRef](#)] [[PubMed](#)]
16. Goutières, F. Aicardi-Goutières syndrome. *Brain Dev.* **2005**, *27*, 201–206. [[CrossRef](#)] [[PubMed](#)]
17. Pulliero, A.; Fazzi, E.; Cartiglia, C.; Orcesi, S.; Balottin, U.; Uggetti, C.; La Piana, R.; Olivieri, I.; Galli, J.; Izzotti, A. The Aicardi-Goutières syndrome. Molecular and clinical features of RNase deficiency and microRNA overload. *Mutat. Res.* **2011**, *717*, 99–108. [[CrossRef](#)]
18. Flanagan, J.M.; Funes, J.M.; Henderson, S.; Wild, L.; Carey, N.; Boshoff, C. Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADAR1 as putative anticancer drug targets. *Mol. Cancer Ther.* **2009**, *8*, 249–260. [[CrossRef](#)]
19. Xu, J.; Liu, H.; Yang, Y.; Wang, X.; Liu, P.; Li, Y.; Meyers, C.; Banerjee, N.S.; Wang, H.-K.; Cam, M.; et al. Genome-Wide Profiling of Cervical RNA-Binding Proteins Identifies Human Papillomavirus Regulation of RNASEH2A Expression by Viral E7 and E2F1. *mBio* **2019**, *10*, e02687-18. [[CrossRef](#)]
20. Williams, K.A.; Lee, M.; Hu, Y.; Andreas, J.; Patel, S.J.; Zhang, S.; Chines, P.; Elkahlon, A.; Chandrasekharappa, S.; Gutkind, J.S.; et al. A systems genetics approach identifies CXCL14, ITGAX, and LPCAT2 as novel aggressive prostate cancer susceptibility genes. *PLoS Genet.* **2014**, *10*, e1004809. [[CrossRef](#)]
21. Zhang, L.; Yang, Y.; Cheng, L.; Cheng, Y.; Zhou, H.H.; Tan, Z.R. Identification of Common Genes Refers to Colorectal Carcinogenesis with Paired Cancer and Noncancer Samples. *Dis. Markers* **2018**, 3452739. [[CrossRef](#)] [[PubMed](#)]
22. Yuan, N.; Zhang, G.; Bie, F.; Ma, M.; Ma, Y.; Jiang, X.; Wang, Y.; Hao, X. Integrative analysis of lncRNAs and miRNAs with coding RNAs associated with ceRNA crosstalk network in triple negative breast cancer. *Onco. Targets Ther.* **2017**, *10*, 5883–5897. [[CrossRef](#)]
23. Lockhart, A.; Pires, V.B.; Bento, F.; Kellner, V.; Luke-Glaser, S.; Yakoub, G.; Ulrich, H.D.; Luke, B. RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep.* **2019**, *29*, 2890–2900.e5. [[CrossRef](#)]
24. Pearson’s Correlation Coefficient. In *Encyclopedia of Public Health*; Kirch, W. (Ed.) Springer: Dordrecht, The Netherlands, 2008. [[CrossRef](#)]
25. Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **2009**, *10*, 48. [[CrossRef](#)]
26. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)] [[PubMed](#)]
27. Li, J.-R.; Sun, C.-H.; Li, W.; Chao, R.-F.; Huang, C.-C.; Zhou, X.J.; Liu, C.-C. Cancer RNA-Seq Nexus: A database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res.* **2016**, *44*, D944–D951. [[CrossRef](#)] [[PubMed](#)]
28. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400–416.e11. [[CrossRef](#)] [[PubMed](#)]
29. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404, Erratum in: *Cancer Discov.* **2012**, *2*, 960. [[CrossRef](#)] [[PubMed](#)]
30. 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*, 11. [[CrossRef](#)] [[PubMed](#)]
31. Garcia, B.A.; Smalley, D.M.; Cho, H.; Shabanowitz, J.; Ley, K.; Hunt, D.F. The platelet microparticle proteome. *J. Proteome Res.* **2005**, *4*, 1516–1521. [[CrossRef](#)]
32. Mukherjee, A.; Stathos, M.E.; Varner, C.; Arsiwala, A.; Frey, S.; Hu, Y.; Smalley, D.M.; Schaffer, D.V.; Kane, R.S. One-pot synthesis of heterodimeric agonists that activate the canonical Wnt signaling pathway. *Chem. Commun.* **2020**, *56*, 3685–3688. [[CrossRef](#)] [[PubMed](#)]
33. Van Ooijen, M.P.; Jong, V.L.; Eijkemans, M.J.C.; Heck, A.J.R.; Andeweg, A.C.; Binai, N.A.; Van den Ham, H.J. Identification of differentially expressed peptides in high-throughput proteomics data. *Brief Bioinform.* **2018**, *19*, 971–981. [[CrossRef](#)] [[PubMed](#)]

34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 25 June 2020).
35. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)] [[PubMed](#)]
36. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R., 3rd; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [[CrossRef](#)] [[PubMed](#)]
37. Harrell, F.E., Jr.; With contributions from Charles Dupont and Many Others. 2020 Hmisc: Harrell Miscellaneous. R Package Version 4.4-1. Available online: <https://CRAN.R-project.org/package=Hmisc> (accessed on 18 December 2020).
38. Arora, R.; Lee, Y.; Wischniewski, H.; Brun, C.M.; Schwarz, T.; Azzalin, C.M. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **2014**, *5*, 5220. [[CrossRef](#)] [[PubMed](#)]
39. Cusanelli, E.; Chartrand, P. Telomeric repeat-containing RNA TERRA: A noncoding RNA connecting telomere biology to genome integrity. *Front. Genet.* **2015**, *6*, 143. [[CrossRef](#)] [[PubMed](#)]
40. Li, M.; Sun, Q.; Wang, X. Transcriptional landscape of human cancers. *Oncotarget* **2017**, *8*, 34534–34551. [[CrossRef](#)]
41. Shao, X.; Lv, N.; Liao, J.; Long, J.; Xue, R.; Ai, N.; Xu, D.; Fan, X. Copy number variation is highly correlated with differential gene expression: A pan-cancer study. *BMC Med. Genet.* **2019**, *20*, 175. [[CrossRef](#)]
42. Freund, A.; Zhong, F.L.; Venteicher, A.S.; Meng, Z.; Veenstra, T.D.; Frydman, J.; Artandi, S.E. Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1. *Cell* **2014**, *159*, 1389–1403. [[CrossRef](#)]
43. Kaisari, S.; Sistry-Shevah, D.; Miniowitz-Shemtov, S.; Teichner, A.; Hershko, A. Role of CCT chaperonin in the disassembly of mitotic checkpoint complexes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 956–961. [[CrossRef](#)]
44. Yang, S.; Ren, X.; Liang, Y.; Yan, Y.; Zhou, Y.; Hu, J.; Wang, Z.; Song, F.; Wang, F.; Liao, W.; et al. KNK437 restricts the growth and metastasis of colorectal cancer via targeting DNAJA1/CDC45 axis. *Oncogene* **2020**, *39*, 249–261. [[CrossRef](#)] [[PubMed](#)]
45. Madungwe, N.B.; Feng, Y.; Lie, M.; Tombo, N.; Liu, L.; Kaya, F.; Bopassa, J.C. Mitochondrial inner membrane protein (mitofilin) knockdown induces cell death by apoptosis via an AIF-PARP-dependent mechanism and cell cycle arrest. *Am. J. Physiol. Cell Physiol.* **2018**, *315*, C28–C43. [[CrossRef](#)] [[PubMed](#)]
46. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
47. Whitfield, M.L.; George, L.K.; Grant, G.D.; Perou, C.M. Common markers of proliferation. *Nat. Rev. Cancer.* **2006**, *2*, 99–106. [[CrossRef](#)] [[PubMed](#)]
48. Wang, C.; Wang, G.; Feng, X.; Shepherd, P.; Zhang, J.; Tang, M.; Chen, Z.; Srivastava, M.; McLaughlin, M.E.; Navone, N.M.; et al. Genome-wide CRISPR screens reveal synthetic lethality of RNASEH2 deficiency and ATR inhibition. *Oncogene* **2019**, *38*, 2451–2463. [[CrossRef](#)]
49. Arudchandran, A.; Cerritelli, S.M.; Narimatsu, S.K.; Itayaa, M.; Shinb, D.-Y.; Shimadac, Y.; Crouch, R.J. The absence of ribonuclease H1 or H2 alters the sensitivity of *Saccharomyces cerevisiae* to hydroxyurea, caffeine and ethyl methanesulphonate: Implications for roles of RNases H in DNA replication and repair. *Genes Cells* **2000**, *5*, 789–802. [[CrossRef](#)]
50. Reijns, M.A.; Rabe, B.; Rigby, R.E.; Mill, P.; Astell, K.R.; Lettice, L.A.; Boyle, S.; Leitch, A.; Keighren, M.; Kilanowski, F.; et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **2012**, *149*, 1008–1022. [[CrossRef](#)]
51. McElhinny, S.A.N.; Watts, B.E.; Kumar, D.; Watt, D.L.; Lundström, E.-B.; Burgers, P.M.J.; Johansson, E.; Chabes, A.; Kunkel, T.A. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4949–4954. [[CrossRef](#)] [[PubMed](#)]
52. Williams, J.S.; Lujan, S.A.; Kunkel, T.A. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 350–363. [[CrossRef](#)]
53. O'Connor, M.J. Targeting the DNA Damage Response in Cancer. *Mol. Cell.* **2015**, *60*, 547–560. [[CrossRef](#)] [[PubMed](#)]
54. Vijayraghavan, S.; Tsai, F.L.; Schwacha, A. A Checkpoint-Related Function of the MCM Replicative Helicase Is Required to Avert Accumulation of RNA:DNA Hybrids during S-phase and Ensuing DSBs during G2/M. *PLoS Genet.* **2016**, *12*, e1006277. [[CrossRef](#)] [[PubMed](#)]
55. Ciferri, C.; Pasqualato, S.; Screpanti, E.; Varetto, G.; Santaguida, S.; Dos Reis, G.; Maiolica, A.; Polka, J.; De Luca, J.G.; De Wulf, P.; et al. Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell* **2008**, *133*, 427–439. [[CrossRef](#)] [[PubMed](#)]
56. Foltz, D.R.; Jansen, L.E.; Black, B.E.; Bailey, A.O.; Yates, J.R., 3rd; Cleveland, D.W. The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **2006**, *8*, 458–469. [[CrossRef](#)]
57. Miki, H.; Setou, M.; Kaneshiro, K.; Hirokawa, N. All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7004–7011. [[CrossRef](#)]
58. Dai, B.; Zhang, P.; Zhang, Y.; Pan, C.; Meng, G.; Xiao, X.; Wu, Z.; Jia, W.; Zhang, J.; Zhang, L. RNaseH2A is involved in human gliomagenesis through the regulation of cell proliferation and apoptosis. *Oncol. Rep.* **2016**, *36*, 173–180. [[CrossRef](#)]
59. Shen, J.; Lin, S.; Liu, L.; Wang, C. Ribonuclease H2 Subunit A impacts invasiveness and chemoresistance resulting in poor survivability of breast cancer in ER dependent manner. *Am. J. Transl. Res.* **2020**, *12*, 2281–2294. [[PubMed](#)]

-
60. Kessler, M.; Zietlow, R.; Meyer, T.F. Stem cells in the reproductive system. In *Adults Stem Cell Niches*; Wislet, S., Ed.; IntechOpen: London, UK, 2014; Chapter 6; pp. 139–169.
 61. Stepanenko, A.A.; Dmitrenko, V.V. HEK293 in cell biology and cancer research: Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **2015**, *569*, 182–190. [[CrossRef](#)]