

Article

# The Circular RNA Landscape of Non-Small Cell Lung Cancer Cells

Nele Van Der Steen <sup>1,2,†</sup>, Yanhong Lyu <sup>3,†</sup>, Anne K. Hitzler <sup>1,†</sup>, Andrea C. Becker <sup>1</sup>,  
Jeanette Seiler <sup>3</sup> and Sven Diederichs <sup>1,2,3,\*</sup> 

<sup>1</sup> Division of Cancer Research, Department of Thoracic Surgery, Medical Center—University of Freiburg, Faculty of Medicine, University of Freiburg, 79106 Freiburg, Germany; nele.vandersteen@uniklinik-freiburg.de (N.V.D.S.); anne.hitzler@uniklinik-freiburg.de (A.K.H.); a-c-becker@gmx.de (A.C.B.)

<sup>2</sup> German Cancer Consortium (DKTK), Partner Site Freiburg, 79106 Freiburg, Germany

<sup>3</sup> Division of RNA Biology & Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; y.lyu@dkfz.de (Y.L.); jeanette.seiler@dkfz.de (J.S.)

\* Correspondence: s.diederichs@dkfz.de

† Authors contributed equally.

Received: 30 March 2020; Accepted: 23 April 2020; Published: 28 April 2020



**Abstract:** The class of circular RNA (circRNA) is characterized by head-to-tail bonds between exons formed by backsplicing. Here, we provide a resource of circRNA expression in a comprehensive panel of 60 lung cancer and non-transformed cell lines (FL3C dataset). RNA sequencing after depletion of ribosomal RNA quantified the expression of circRNA and linear RNA. We detected 148,811 circular RNAs quantified by 2.8 million backsplicing reads originating from 12,251 genes. The number of identified circRNAs was markedly higher using rRNA depletion compared to public polyA-enriched RNA-seq datasets. CircRNAs almost never started in the first exon nor ended in the last exon and started more frequently in earlier exons. Most circRNAs showed high cell line specificity and correlated positively with their linear RNA counterpart. Known cancer genes produced more circRNAs than non-cancer genes. Subsets of circRNAs correlated with cell proliferation, histological subtype or genotype. *CircTNFRSF21* was translated crossing the backsplice site in two different reading frames. Overexpression of *circPVT1*, *circERBB2*, *circHIPK3*, *circCCNB1*, *circSMAD2*, *circTNFRSF21* and *circKIF5B* significantly increased colony formation. In conclusion, our data provide a comprehensive map of circRNA expression in lung cancer cells and global patterns of circRNA production as a useful resource for future research into lung cancer circRNAs.

**Keywords:** circular RNA; circRNA; NSCLC; lung cancer; lung adenocarcinoma

## 1. Introduction

Circular RNAs (circRNAs) were first described in eukaryotes in the early 1990s and were originally considered to be formed by splicing defects [1–3]. Nowadays, numerous circRNAs have been identified by analyzing RNA sequencing data [4]. In contrast to mature messenger RNAs (mRNAs), circRNAs do not possess a polyA tail and a 5′-cap, but are circularized head-to-tail with a 5′–3′ covalent bond [3]. CircRNAs can be identified by mapping RNA sequencing data to a reference genome and identifying sequences that have a reverse orientation [4]. CircRNAs are more stable than their linear counterparts, likely because of protection from exonucleolytic degradation by their covalent head-to-tail bond [5,6].

Backsplicing is dependent on the splicing machinery and uses canonical splice sites [4,6–9]. Although the details of the process have not been fully elucidated, several rules about the regulation of circRNA formation have been reported. First, a high transcription rate of RNA polymerase II leads to

an increase in backsplicing events [10]. Second, there is a competition between circRNA biogenesis and pre-mRNA splicing [8]. Third, backsplicing seems to be facilitated by inverted repeats in the flanking introns [11], e.g., Alu elements [12]. Fourth, RNA binding proteins might promote the looping of RNA through dimerization upon binding of their respective motifs, e.g., Quaking (QKI) [13] and FUS [14] or reduce looping by reducing complementarity, e.g., ADAR1 [11]. In general, cellular proliferation leads to a downregulation in circRNA expression [15], although a small subset of circRNAs shows an increase upon proliferation [16]. While many circRNAs contain in silico predicted open reading frames (ORFs) with internal ribosomal entry sites (IRES), only a small number of these proteins/peptides have been detected or experimentally verified [17–22].

Currently, the majority of studies linking circRNAs to cancer focus on their function as miRNA sponges [23,24], thus influencing oncogenic pathways [25]. Additionally, peptides arising from circRNAs can function as decoys and thus influence cancer growth, e.g., the peptide translated from *circ $\beta$ -catenin* protects full-length  $\beta$ -catenin from phosphorylation by GSK3 $\beta$  and subsequent degradation [26]. Finally, circRNAs can influence cell proliferation by protein scaffolding, e.g., the *circFOXO3* RNA forms a complex with CDK2 and p21 to prevent cell cycle entry [27].

Lung cancer, representing 11.8% of all cancer diagnoses, is the most commonly diagnosed cancer type worldwide [28]. It is also the leading cause of cancer-related deaths worldwide, with 1.8 million deaths per year, which represents 18.4% of all cancer-related deaths [28]. The most common type of lung cancer is non-small cell lung cancer (NSCLC), representing 85% of lung cancers. NSCLC can be further divided into adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) subtypes [29]. While many pathways have been linked to lung tumorigenesis like EGFR or KRAS [30], the underlying mechanisms remain unknown in many cases with non-coding RNAs emerging as additional players in carcinogenesis and tumor progression like *MALAT1* [31], *PVT1* [32] or *linc00673* [33].

Due to their high stability, circRNAs are considered as good candidates for new biomarkers [34]. A specific example for lung cancer are the circRNAs that originate from the EML4-ALK fusion gene, F-circEA, which can be detected in plasma samples of these patients [35,36]. Moreover, circRNAs might serve as good predictive biomarkers for response to therapy [37–39].

Here, we describe the circRNA landscape in non-small cell lung cancer cell lines. After assembling a platform of 60 lung cell lines (57 lung cancer cell lines and 3 non-transformed lung cell lines), we used deep sequencing of rRNA-depleted RNA for profiling the exonic circRNAs and the linear RNA transcriptome. We describe the general characteristics of this dataset taking into account differences between the gene level (all circRNAs of one gene were grouped during analysis) and the backsplice level (all circRNAs were considered separately during analysis). Furthermore, we link circRNAs to specific phenotypes and genotypes in non-small cell lung cancer.

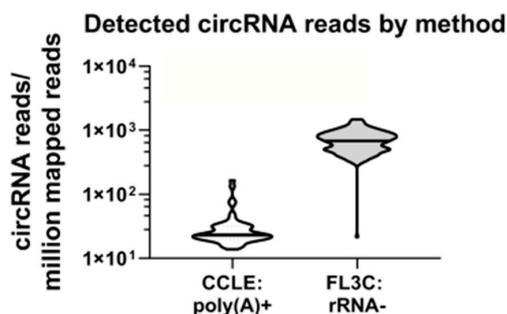
## 2. Results

### 2.1. circRNA Detection in Lung Cancer Cells after rRNA Depletion

We assembled a lung cell line panel of 60 lung cell lines, consisting of 50 adenocarcinoma cell lines, seven other NSCLC cell lines and three non-transformed cell lines (Supplementary Table S1), which we named the Freiburg Lung Cancer Cell Collection (FL3C). After total RNA isolation, the rRNA was depleted and RNA of all cell lines was sequenced in replicate ( $n = 175$  with two or three replicates per cell line) and mapped to a reference genome to generate the linear RNA dataset. Next, we identified circRNAs by identifying reverse mapped reads resulting from backsplicing and constructed a separate circRNA dataset. In total, we found 2.8 million backsplicing reads compared to 3.8 billion reads mapping linearly to the genome.

Overall, we found on average 731 circRNA reads per million reads in our dataset based on rRNA depletion prior to RNA sequencing. At the gene level, we detected circRNAs for 12,251 genes and provide the full dataset for 60 cell lines in Supplementary Table S2. At the backsplice level, we identified 148,811 individual circRNAs and provide the full dataset in Supplementary Table S3.

We compared our dataset to a publically available dataset of the Cancer Cell Line Encyclopedia (CCLE) [40,41] from which we retrieved RNA sequencing data after polyA-enrichment from 54 cell lines (single replicate) overlapping with our panel. Notably, these data contained 25-fold less circRNA reads (Figure 1).



**Figure 1.** Detected circRNA reads by method. This violin plot compares the detected circRNA reads per million mapped reads in the CCLE and the FL3C database.

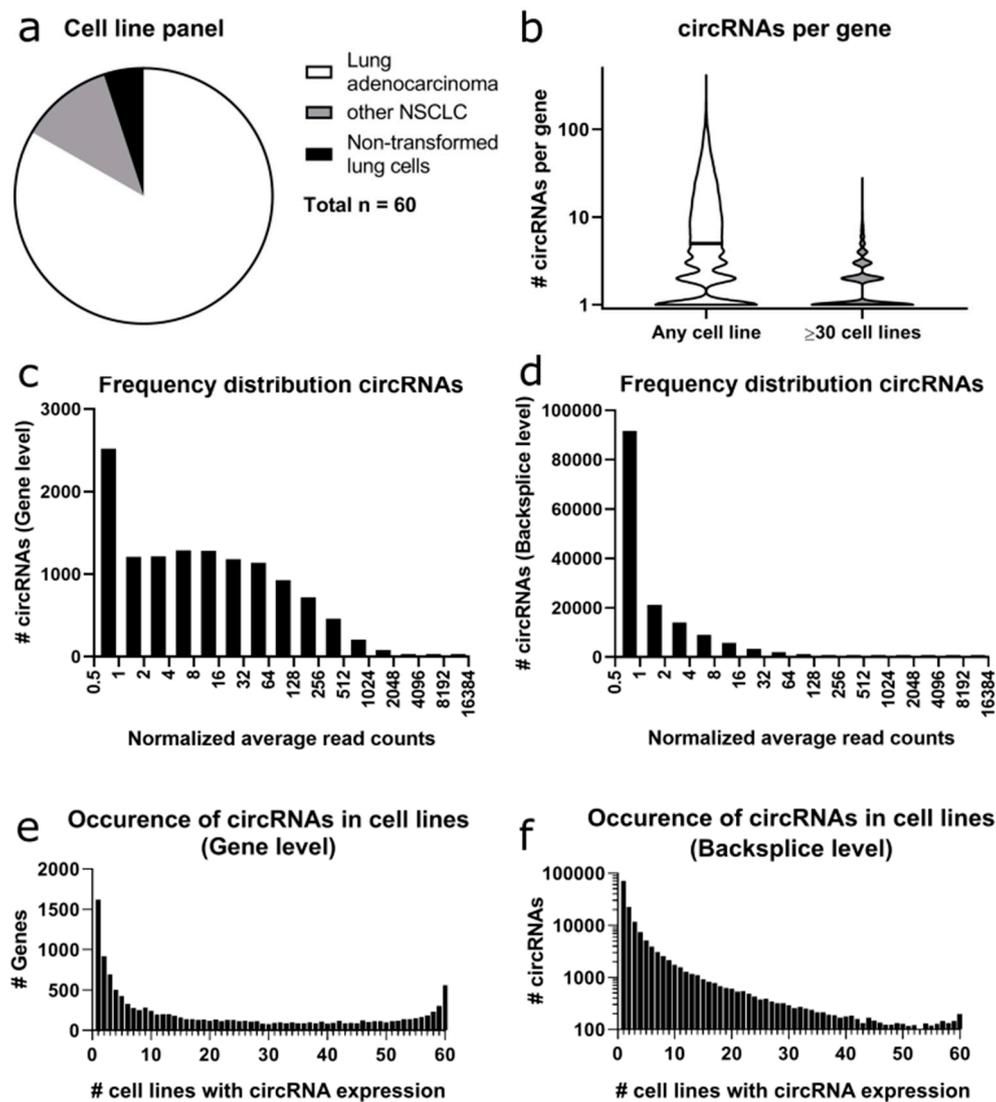
Next, we looked at the enrichment in polyA stretches between the CCLE and the FL3C datasets. In the CCLE dataset, 11,441 circRNAs were identified, of which 5587 were overlapping with the FL3C dataset, which contained in total 148,811 circRNAs. When we compared the top 100 most strongly expressed circRNAs, 15 showed no overlap and 85 were shared between the datasets. Of the shared circRNAs, 69% contained polyA stretches of 5 or more consecutive As, versus only 33% of the circRNAs that were uniquely identified in the FL3C dataset. In conclusion, there may be a trend towards enrichment in polyA stretches in circRNAs that were identified by sequencing after polyA selection. In brief, efficient circRNA detection was feasible after rRNA depletion and subsequent RNA sequencing while the circRNA content detected in polyA-enriched RNA was significantly lower.

## 2.2. Lung Cell Line Transcriptomics and the circRNA Landscape

After quantifying the linear and circular RNA read counts in the FL3C dataset (Figure 2a), we calculated the number of different backsplicing sites per gene representing different circRNAs derived from the same gene (Figure 2b). For the entire dataset, we found on average 12 and at median five different backsplice sites per gene, with a minimum of one (in 2506 out of the 12,251 genes with circRNAs detected) and a maximum of 418 for *BIRC6*. Next, we repeated the analysis only for the circRNAs that were expressed in at least 30 out of the 60 cell lines focusing on more broadly detectable circRNAs. This filter resulted in 4124 circRNAs at the gene level and 5232 circRNAs at the backsplice level. In this subset, we found on average two and at median only one circRNA per gene with *PTK2* expressing the maximum number of circRNAs with 28 unique backsplice sites. For 1170 genes, more than one circRNA was expressed in the majority of cell lines. The normalized circular read counts (average of 2–3 replicates, normalized to the library size) at the gene level ranged from 14,487 to 0.17, with an average of 70.51 circular reads at gene level in all cell lines (Figure 2c). The genes with the highest circRNA read counts were *CYP24A1* (14,487) and *ASPH* (13,783).

This result is mirrored at the backsplice level (Figure 2d), with the circRNA *circCYP24A1* with a backsplice site linking exons 11 to 3 having 14,417 counts and circRNA *circASPH* (backsplice site: exon 3-2) having 10,493 counts. Overall, the normalized counts at the backsplice level ranged from 14,417 to 0.17, with an average of 5.8 reads per unique circRNA.

When comparing the expression of circRNAs across all cell lines tested, there were major differences between the analysis at the gene level (Figure 2e) and the backsplice level (Figure 2f). At the gene level, 13.2% of genes expressed circRNAs only in a single cell line, while at the backsplice level, 47.3% of the circRNAs were only expressed in a single cell line. At the gene level, 14.1% of the genes expressed circRNAs in  $\geq 90\%$  of the cell lines, while at the backsplice level, only 0.7% of the circRNAs were expressed in  $\geq 90\%$  of cell lines.

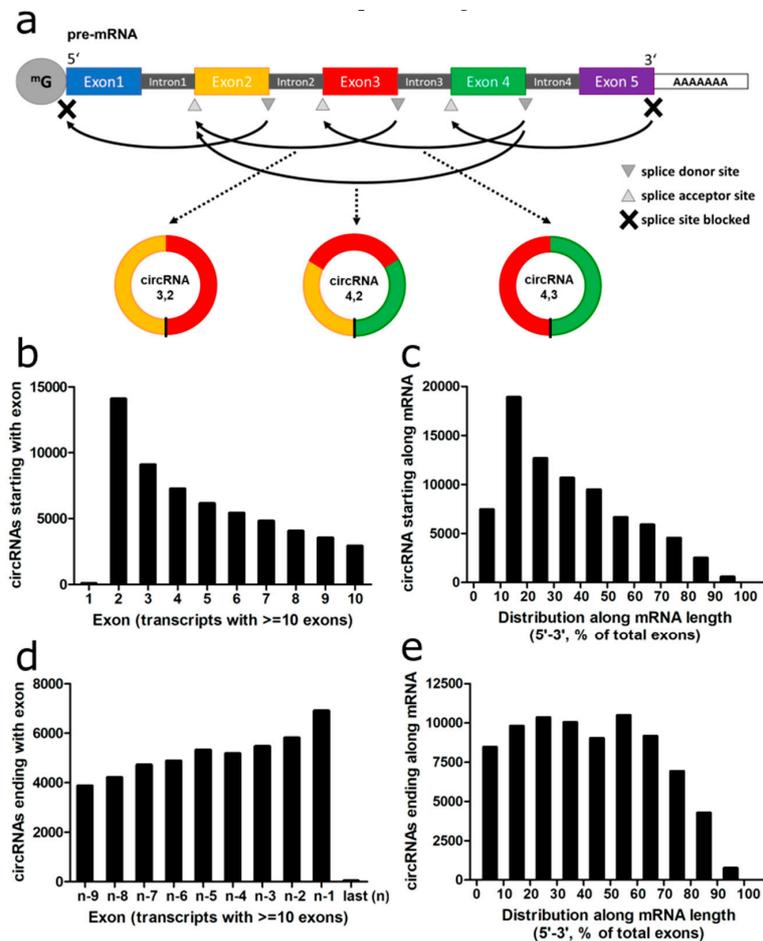


**Figure 2.** General dataset characteristics. (a) The FL3C cell line panel consists of 60 human lung cell lines (50 cell lines have been described as adenocarcinoma, 7 cell lines as non-adenocarcinoma NSCLC and 3 cell lines as non-transformed lung cells). (b) circRNAs per gene. The violin plot depicts the number of circRNAs (different backsplice sites) per gene in the FL3C dataset before and after selection of the circRNAs that are expressed in at least half of the cell lines. The black line depicts the median. (c) Frequency distribution of circRNAs at the gene level. Bar graph depicting the distribution of the sum of the circRNA reads found in one gene in the FL3C dataset, i.e., how many genes (y-axis) show a total number of circRNA reads indicated on the x-axis. (d) Frequency distribution of circRNAs at the backsplice level (as in (c)). (e) Occurrence of circRNAs in cell lines at the gene level. The graph depicts the distribution of genes according to the number of cell lines with occurrence of a circRNA for this gene. (f) Occurrence of circRNAs in cell lines at the backsplice level. The graph depicts the distribution of circRNAs according to the number of cell lines with occurrence of the circRNA.

For six selected circRNAs, we verified their expression in lung cancer cell lines using RT-PCR in comparison to the total RNA of the respective gene (Supplementary Figure S1a). To confirm the circularization of the detected amplicons, we treated the RNA with the exonuclease RNase R. The ratio of circular to total RNA markedly differed between the genes with *PVT1* having the largest circular fraction of its total RNA and *CCNB1* having the lowest circular fraction of its total RNA (Supplementary Figure S1b).

### 2.3. The First and Last Exon Are Depleted From circRNAs

We then analyzed the representation of the exons along the mRNA within circRNAs. First, we analyzed the acceptor exon of the circRNAs, i.e., the “start” exon appearing first in the linear RNA to which a later exon, the donor exon, was backspliced (Figure 3a).



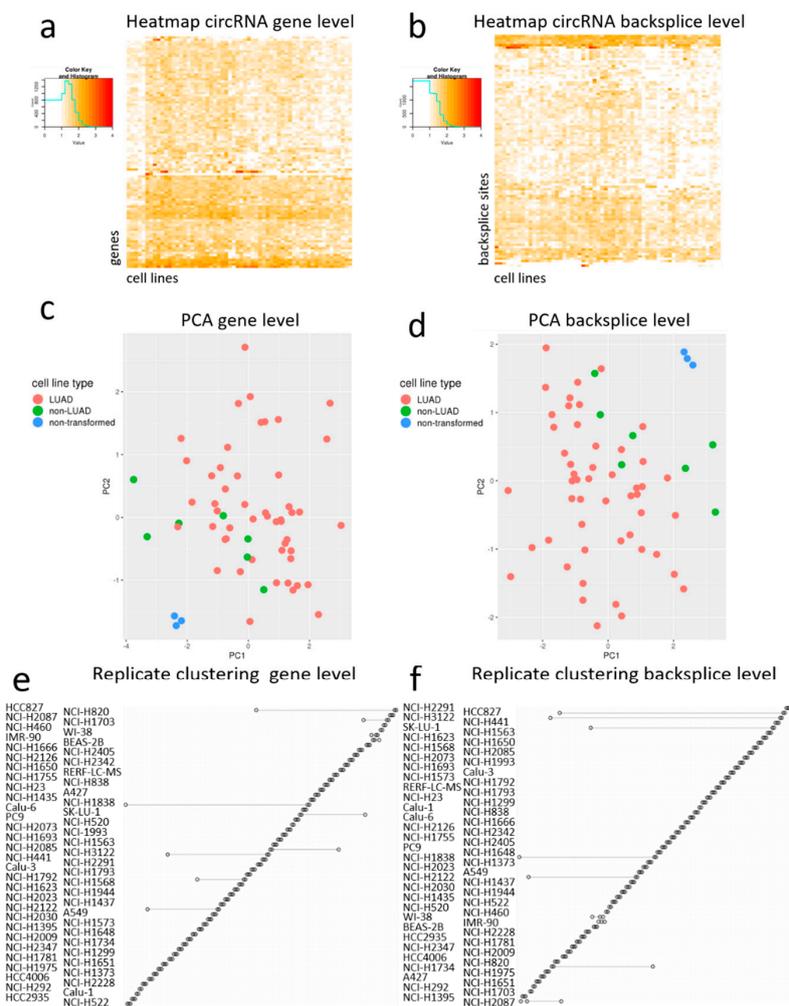
**Figure 3.** Prevalence of exons in circRNAs. (a) Schematic depiction of alternative backsplicing events. (b) Number of detected circRNAs that use an exon as acceptor exon. (c) Number of detected circRNAs that use a percentile of mRNA as acceptor exon. (d) Number of detected circRNAs that use an exon as donor exon. (e) Number of detected circRNAs that use a percentile of mRNA as donor exon.

The second exon of the transcript was more frequently serving as acceptor exon than the first exon with a 78-fold difference for all circRNAs and a 126-fold difference for circRNAs detected with at least two reads and derived from transcripts with at least 10 exons (Figure 3b). When looking along the entire mRNA, the first 10% of exons were 2.5-fold less often serving as acceptor exons in comparison to the second 10%. The second exon was the most prevalent acceptor exon for circularization, while this steadily decreased over the subsequent exons (Figure 3c).

We repeated the analysis for the donor exon of the circRNA. The last exon of the transcript was less often serving as donor exon in comparison to the second to last exon with a 62-fold difference across all circRNAs and a 1778-fold difference for transcripts of  $\geq 10$  exons and at least two circRNA reads (Figure 3d). The last 10% of the exons were 1.4-fold less often donor exons in comparison to the second to last increment (Figure 3e). While the rate of circRNAs using an exon as donor was declining with decreasing exon number similar to the acceptor exons along the mRNA, this effect was much less pronounced than for the acceptor exons. In summary, the first and last exons very rarely were included in circRNAs, while the second exon was most frequently serving as acceptor exon.

#### 2.4. circRNAs Separate Lung Cancer Cells from Non-Transformed Lung Cells

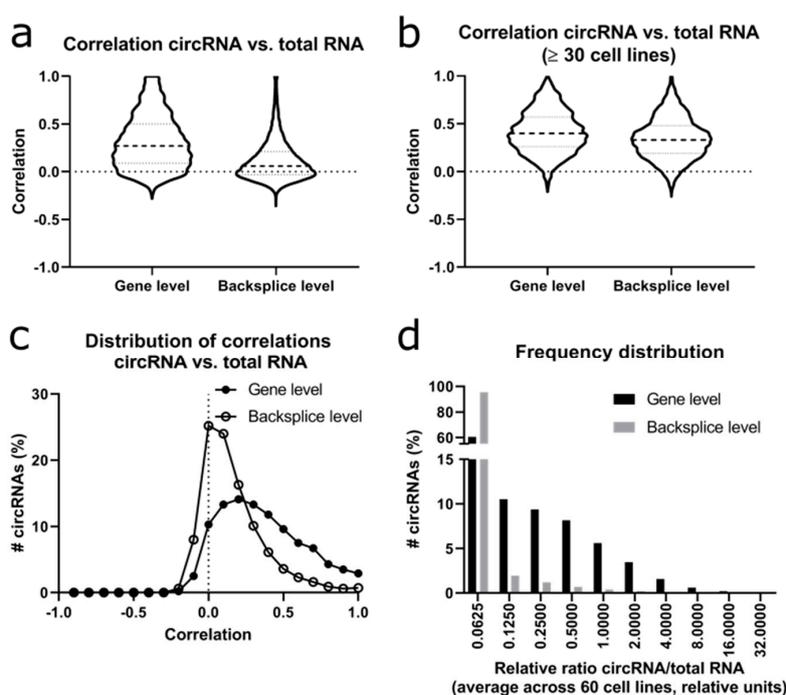
Next, we constructed heatmaps, based on the top 100 genes producing most circRNAs (Figure 4a) and on the top 100 most highly expressed circRNAs (Figure 4b). Both heatmaps (gene level and backsplice level) showed the high abundance of circRNAs derived from *CYP24A1* and *ASPH*. Moreover, the heatmap for individual circRNAs also showed a high expression, e.g., of *circERBB2* specifically in the Calu-3 cell line, which contained an *ERBB2* amplification, thus showing a link with the genotype of the cell line. When performing unsupervised clustering for the cell lines, the non-transformed cell lines clustered together, both at the gene and the backsplice level. The principal component analysis also depicted the non-transformed cell lines clustering separately from the adenocarcinoma and non-adenocarcinoma NSCLC cell lines (Figure 4c,d). To a lesser extent, the non-adenocarcinoma cell lines also clustered together at both levels. The replicate clustering showed a very good clustering reflecting high reproducibility for the large majority of the cell line replicates both at the gene and the backsplice level (Figure 4e,f).



**Figure 4.** Clustering and reproducibility of circRNA expression at the gene or backsplice level. (a) Heatmap illustrating clustering of the cell lines by the top 100 highest expressed circRNAs grouped per gene. (b) Heatmap illustrating clustering of the cell lines by the top 100 highest expressed unique circRNAs. (c) Principal component analysis (PCA) of the cell lines based on the top 100 highest expressed circRNAs grouped per gene. (d) Principal component analysis of the cell lines based on the top 100 highest expressed unique circRNAs. (e) Clustering of the cell line replicates at gene level. (f) Clustering of the cell line replicates at backsplice level.

### 2.5. Positive Correlation between Circular and Linear RNA Expression

Since circRNAs are formed from the linear transcript, we calculated the correlation distribution between circRNA and “linear” RNA derived from the same gene. We refer to the signals derived from the linear RNA analysis as “total RNA” since all reads of the circRNA outside of the backsplice region would be indistinguishable from linear reads and would thus be included into this analysis of linear RNAs. When considering the total dataset, there was a positive correlation at gene level with a median correlation of 0.27 (Figure 5a), whereas this correlation was much lower at the backsplice level when considering each of the circRNAs separately with a median correlation of only 0.06. When repeating this analysis for circRNAs that were expressed in  $\geq 30$  cell lines, the difference between gene level and backsplice level decreased (Figure 5b), with a median correlation of 0.40 at the gene level and of 0.27 at the backsplice level. So, the majority of the well expressed circRNAs showed a positive correlation to the underlying total (linear) RNA, while only very few examples showed any negative correlation. At the gene level, 86.9% of circRNAs showed a positive correlation with their co-derived total (linear) RNA, versus 66.2% of circRNAs at backsplice level (Figure 5c). Next, we calculated the relative ratio of circRNA/total RNA (Figure 5d).



**Figure 5.** Correlation between circRNA and total (linear) RNA. (a) Distribution of correlation coefficients between circRNA and total RNA (detected by analysis of linear reads, which may, outside of the backsplice region, also be derived from circRNAs) at gene level (circRNA grouped per gene) and backsplice level (unique circRNAs) for the whole dataset. Dashed lines depict the median, dotted lines depict the first and third quartiles. (b) Distribution of correlation between circRNA and total RNA at gene and backsplice level of RNAs that are expressed in  $\geq 30$  cell lines. Dashed line depicts the median, dotted lines depict the first and third quartiles. (c) Distribution of correlation between circRNA and total RNA at gene and backsplice level. (d) Relative circRNA/total RNA ratio, both at gene and backsplice level (rel. units).

Importantly, the calculated ratios of circRNA/total RNA did not reflect an absolute ratio of molecules in the cell but a relative number dividing the number of circRNA reads (only backsplice reads, normalized only for library size) to the FPKM of the linear reads (all reads along the transcript, normalized for library size and gene length). When considering the average ratio of circRNA/total RNA over the 60 cell lines, the median ratio was 0.03 at the gene level versus  $<0.01$  at the backsplice level.

Overall, the circRNA/total RNA ratios showed a broad distribution with a range from <0.01 to >23 for *MUC16* indicating that circRNAs can have largely varying ratios to the total RNAs derived from the same locus. These values may be used to select highly efficiently circularized transcripts.

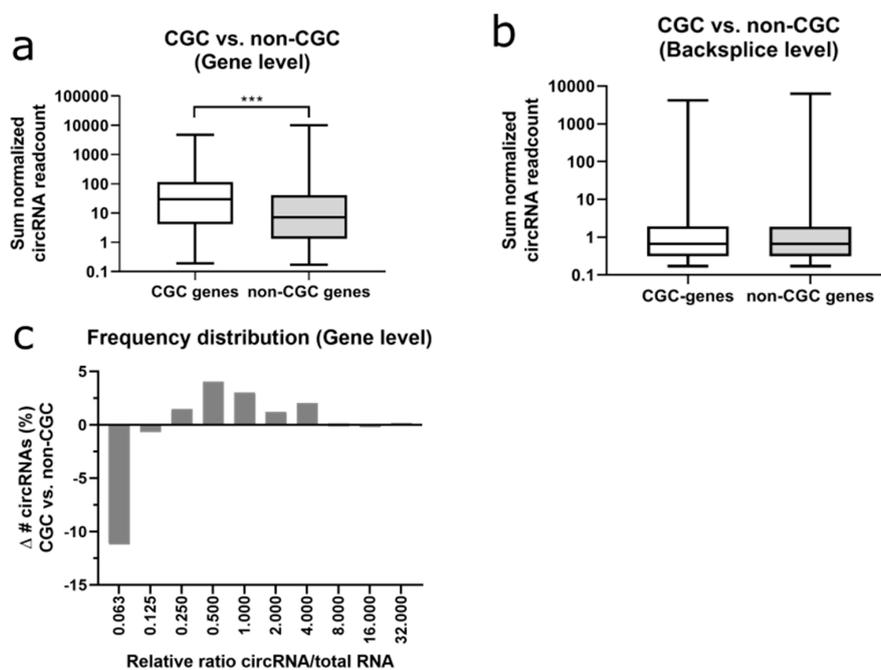
When comparing the ratios in individual cell lines, the highest ratio of 561.3 at gene level was identified for the *EFCAB12* (EF-Hand Calcium Binding Domain 12) gene, but this circRNA was only expressed in 5/60 cell lines. The second highest ratio of 427.8 was reached by the *OTUD7A* (OTU Deubiquitinase 7A) gene, which was expressed in 10/60 cell lines. When a cut-off of expression in  $\geq 30$  cell lines was used, *MUC16* showed the highest ratio of 248.8 with an expression in 40 cell lines. Focusing on individual circRNAs at the backsplice level, *OTUD7A* (backsplice site: 4–3) had the highest circRNA/total RNA ratio of 427.8 but was only expressed in three cell lines. *MUC16* (backsplice site: 69–68) was in the top 10 with a ratio of 248.8 and was expressed in 27 cell lines.

In summary, circRNAs showed a positive correlation with the total linear RNA derived from the same gene, but the ratios varied largely when comparing individual cell lines, genes or circRNAs.

## 2.6. Cancer Genes Produce More Circular RNAs

To assess the circRNA subset derived from known cancer genes, we divided the FL3C dataset into cancer genes and non-cancer genes according to the Cancer Gene Census (CGC) of COSMIC (version 90) [42]. At the gene level, 4.6% of circRNAs originated from CGC-genes (Figure 6a), while CGC genes represent only 2.8% of all coding genes. Average normalized read counts were significantly higher for cancer genes (CGC) as compared to non-CGC-listed genes (2.1-fold,  $T$ -test  $p = 1.8 \times 10^{-5}$ ). At the backsplice level, even 8.4% of circRNAs originated from cancer genes (Figure 6b). In contrast to the gene level, there was only a statistically insignificant 1.1-fold change in average normalized read counts between both categories at the backsplice level.

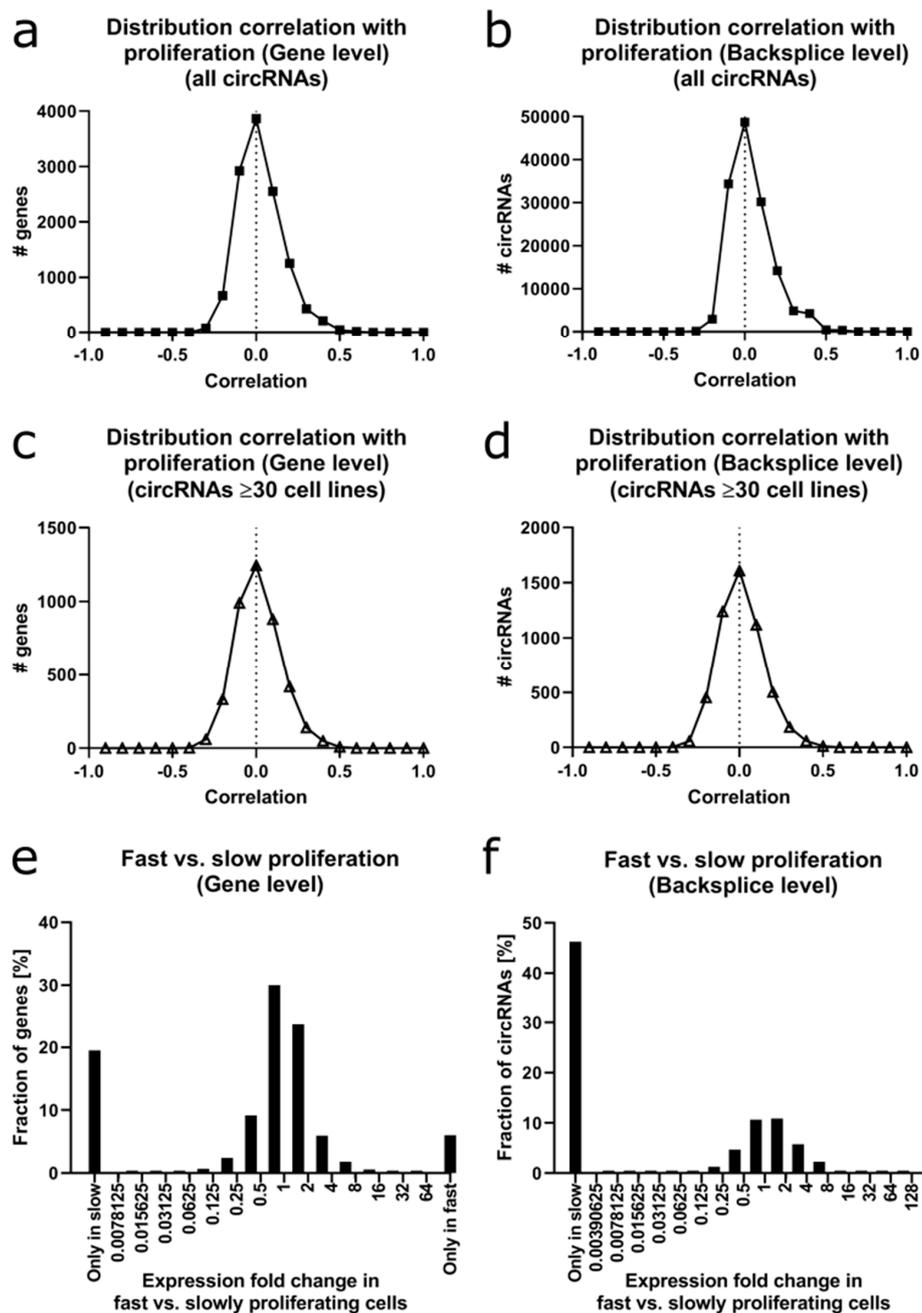
CircRNAs with a low circRNA/total RNA ratio were less frequently originating from cancer genes (Figure 6c). In summary, cancer genes were generating significantly more circRNAs as compared to non-CGC-listed genes, but this difference was not found at the individual circRNA level.



**Figure 6.** CircRNAs in Cancer Genes (CGC, v90). (a) Sum of normalized circRNA read counts in Cancer Gene Census (CGC) genes vs. non-CGC genes at the gene level.  $T$ -test:  $p$ -value < 0.001 (\*\*\*). (b) idem at backsplice level. (c) Frequency distribution of the difference in number of circRNAs between CGC and non-CGC genes for the relative circRNA/total RNA ratios.

### 2.7. A Subset of Circular RNAs Correlates with Proliferation

We determined the correlation of circRNAs with the proliferation rate of the FL3C cell lines, which we had experimentally determined before (see Supplementary Table S1). At the gene level, the correlation coefficient ranged from +0.58 to −0.47 for the whole dataset (Figure 7a).



**Figure 7.** Correlation of circRNAs with cell proliferation. (a) Distribution of the correlation of circRNAs with proliferation rate at gene level (circRNAs grouped per gene) across the whole dataset. (b) Distribution of the correlation of circRNAs with proliferation rate at backsplice level (unique circRNAs) across the whole dataset. (c) Same as (a) with a cut-off of circRNA detection in  $\geq 30$  cell lines. (d) Same as (b) with a cut-off of circRNA detection in  $\geq 30$  cell lines. (e) Distribution of the fold change of circRNAs between fast and slow proliferating cell lines at the gene level. (f) Same as (e) at the backsplice level.

For circRNAs expressed in  $\geq 30$  cell lines, the correlation ranged from +0.52 to  $-0.45$  (Figure 7c). The majority of the circRNAs showed as expected little to no correlation with the proliferation rate.

At the backsplice level, the correlation ranged from +0.65 to  $-0.45$  for the whole dataset (Figure 7b), and +0.49 to  $-0.45$  for circRNAs expressed in  $\geq 30$  cell lines (Figure 7d).

As an orthogonal analysis to the correlation analysis, we divided the cell lines into two groups based on their proliferation rate. Cell lines with  $\geq 5$ -fold increase in cell count within 72 h were considered to be fast proliferating (11 cell lines). Cell lines with  $\leq 3$ -fold cell count increase within 72 h were considered as slowly proliferating (21 cell lines). At the gene level (Figure 7e), circRNAs from 17.8% of the 12,251 genes were only expressed in slowly proliferating cell lines versus 5.5% only expressed in fast proliferating cells. At the backsplice level (Figure 7f), 33.8% of the unique circRNAs were only expressed in slowly proliferating cells versus 12.7% of the circRNAs only expressed in fast proliferating cells. Focusing on the fold change of expression between fast and slowly proliferating cells, 49.0% of the circRNAs had a fold change between 0.5 and 2.0 when considering the whole dataset versus 15.7% at the backsplice level. These percentages rose to 88.5% and 84.5%, respectively, when taking only the circRNAs in consideration that were expressed in  $\geq 30$  cell lines.

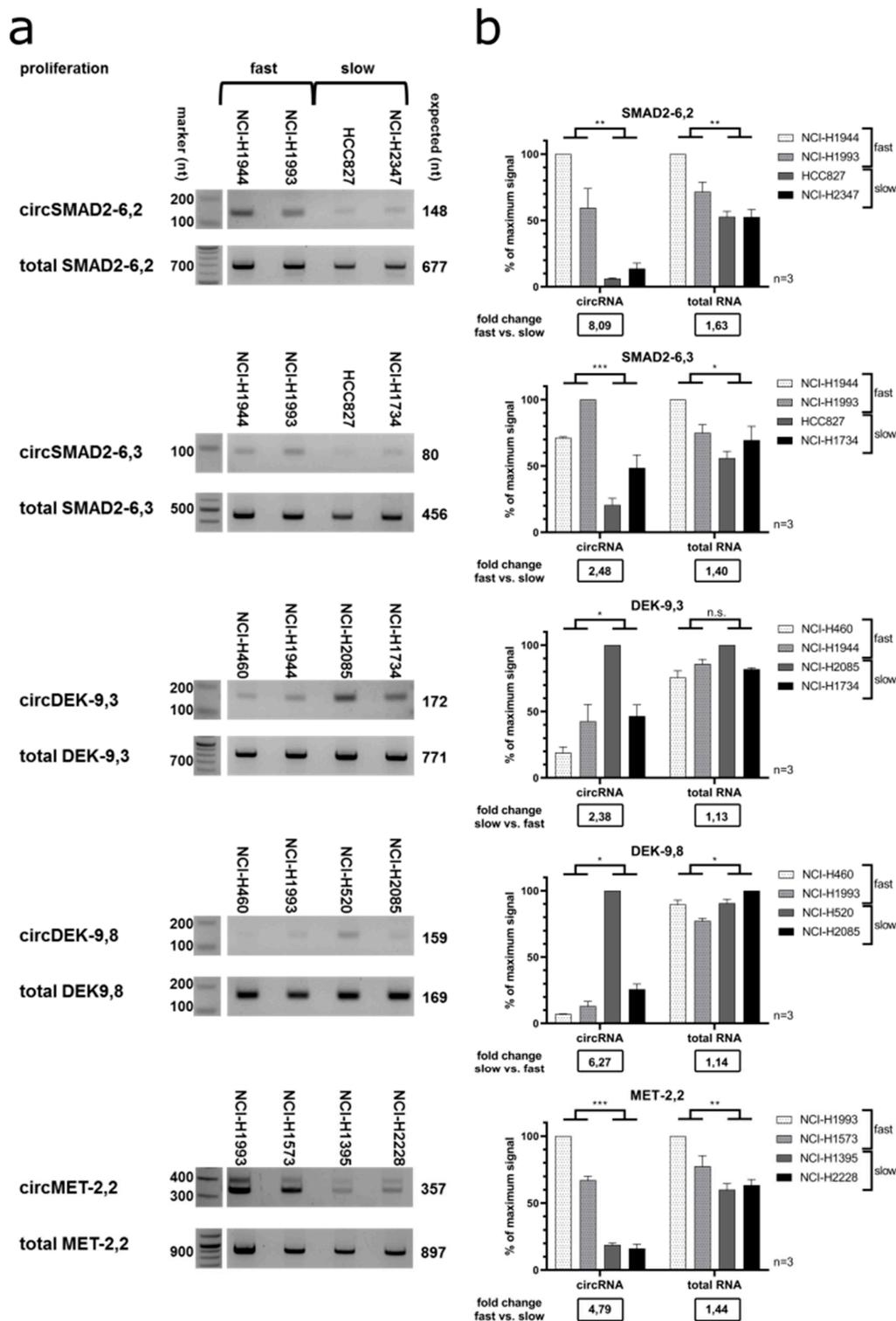
Next, we selected circRNAs derived from CGC genes which were expressed in  $\geq 30$  cell lines and showed a strong positive and/or negative correlation between circRNA expression and proliferation.

At the gene level, three genes drew our attention: *SMAD2*, *MET* and *DEK*. *SMAD2* and *MET* showed a positive correlation between proliferation and circRNA expression (+0.38 and +0.32, respectively) and between proliferation and total RNA expression (+0.44 and +0.23, respectively). In contrast, *DEK* showed a negative correlation between its circRNA expression and proliferation ( $-0.34$ ), while there was notably no correlation between its total (linear) RNA expression and proliferation ( $-0.01$ ).

At the backsplice level, two isoforms of *circSMAD2* were identified with a high correlation to proliferation: backsplice site exon 6–exon 2 (+0.37) and exon 6–exon 3 (+0.35). *MET*, (backsplice site exon 2–exon 2) showed a correlation of +0.32 with proliferation. For *DEK*, also two circRNAs were identified with a large negative correlation to proliferation: backsplice site exon 9–exon 8 ( $-0.32$ ) and exon 9–exon 3 ( $-0.31$ ). For the *KIF5B* gene, two different circRNAs showed opposite correlations to proliferation: backsplice site exon 24–exon 20 (+0.29) and backsplice site exon 20–exon 18 ( $-0.29$ ), while the total RNA showed a weaker negative correlation to proliferation of  $-0.16$ .

We validated these results with semi-quantitative RT-PCR in a subset of cell lines with either fast or slow proliferation (Figure 8). *CircSMAD2* (backsplice site exon 6–exon 2) was upregulated in fast proliferating cell lines and showed a significant 8.1-fold change, versus a significant 1.6-fold change for the total *SMAD2* RNA. The stronger differential expression of *circSMAD2-6,2* than the total *SMAD2* mRNA was also verified by RT-qPCR (Supplementary Figure S2). This difference was less pronounced for *circSMAD2* (backsplice site exon 6–exon 3) resulting in a 2.5-fold change versus a 1.4-fold for the total RNA. For the *circDEK* isoforms, backsplice site exon 9–exon 8 was most strongly downregulated in fast proliferating cell lines, with a significant 6.3-fold change versus a significant only 1.1-fold change for the total RNA. *CircDEK* backsplice site exon 9–exon 3 had a significant 2.4-fold change versus a non-significant 1.1-fold change for the total RNA. *CircMET* showed a significantly higher expression in fast proliferating cell lines with a 4.8-fold change versus a significant only 1.4-fold change for the total RNA (Figure 8a,b).

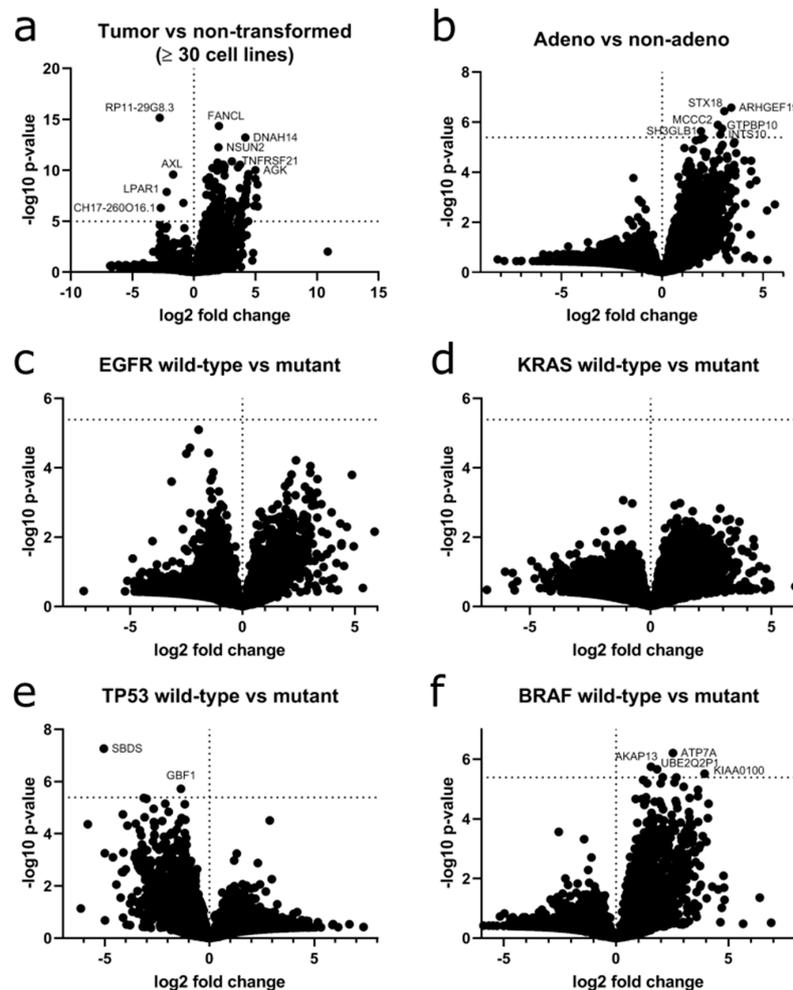
Overall, only a small subset of circRNAs correlated positively with cell proliferation or was differentially expressed between fast and slowly proliferating cells. Interestingly, circRNAs and linear RNAs or different circRNAs originating from the same gene may demonstrate differences in correlation with cell proliferation.



**Figure 8.** Validation of proliferation-correlated circRNAs in fast vs. slowly proliferating cell lines. (a) Representative images after gel electrophoresis. Per row the first two cell lines were fast proliferating ( $\geq 5$ -fold increase in cell count in 72 h), the last two cell lines were slowly proliferating ( $\leq 3$ -fold cell count increase in 72 h). Numbers depict marker size (left) and expected fragment sizes (right). (b) Signal strength of the bands after gel electrophoresis was quantified ( $n = 3$ ). The relative signal (% of highest signal in each group) is depicted for circRNA and total RNA, respectively. *T*-test: not significant (n.s.),  $p$ -value  $< 0.05$  (\*),  $p$ -value  $< 0.01$  (\*\*),  $p$ -value  $< 0.001$  (\*\*\*). Below the graph, the fold change of average relative signal between the groups is shown.

### 2.8. Circular RNAs Correlate with Histological or Genetic Classification

Next, we calculated the difference in circRNA expression according to transformation state, histological subtype and genetic background. When grouping cell lines according to their transformation status (NSCLC vs. non-transformed), the difference in circRNA expression of 128 genes reached statistical significance after correction for multiple testing, e.g., *circTNFRSF21* (Figure 9a).



**Figure 9.** Volcano plots of circRNAs at gene level according to transformation status, histological subtype or genotype depicting fold change and statistical significance. (a) Differential circRNA expression in tumor-derived cell lines versus non-transformed cell lines. (b) Differential circRNA expression in adenocarcinoma cell lines versus non-adenocarcinoma NSCLC cell lines. (c) Differential circRNA expression in EGFR wild-type cell lines versus EGFR mutant cell lines. (d) Differential circRNA expression in TP53 wild-type cell lines versus TP53 oncogenic mutant cell lines. (e) Differential circRNA expression in KRAS wild-type cell lines versus KRAS mutant cell lines. (f) Differential circRNA expression in BRAF wild-type cell lines versus BRAF mutant cell lines. Vertical dotted line depicts a fold change of zero. Horizontal dotted line depicts  $p$ -value threshold for significance corrected for multiple testing. Gene names of genes with a significant fold change in circRNA expression are depicted where possible.

Next, we grouped the tumor cell lines according to their histology (adenocarcinoma vs. non-adenocarcinoma NSCLC, Figure 9b). A generally higher circRNA expression was found in adenocarcinoma. For 22 genes including *TALDO1*, *TMC5*, *MVP* and *FNDC3B*, this change in circRNA expression was significant. Lastly, we grouped the tumor cell lines according to their genotype as determined in the COSMIC cell line project [42]. We focused on prevalent mutations in lung cancer and

compared mutant to wildtype cell lines. Cell lines harboring mutations in *EGFR* (Figure 9c) and *KRAS* (Figure 9e) did not show any significant changes. *TP53* mutant cells showed a higher expression of circRNAs with significance for *circSBDS*, *circGBF1*, *circATP6V1A* and *circVANGL1* (Figure 9d). Cell lines harboring *BRAF* mutations showed an overall lower circRNA expression with 10 genes reaching significance (Figure 9f). All differentially expressed circRNAs are listed in Supplementary Table S4. Overall, selected circRNAs correlate with the transformation status, the histology or the genotype of the lung cells in our study making them candidates for future evaluation of biomarkers due to their high stability.

### 2.9. circRNAs Were Efficiently Overexpressed in NSCLC Cell Lines with Laccase2 Introns

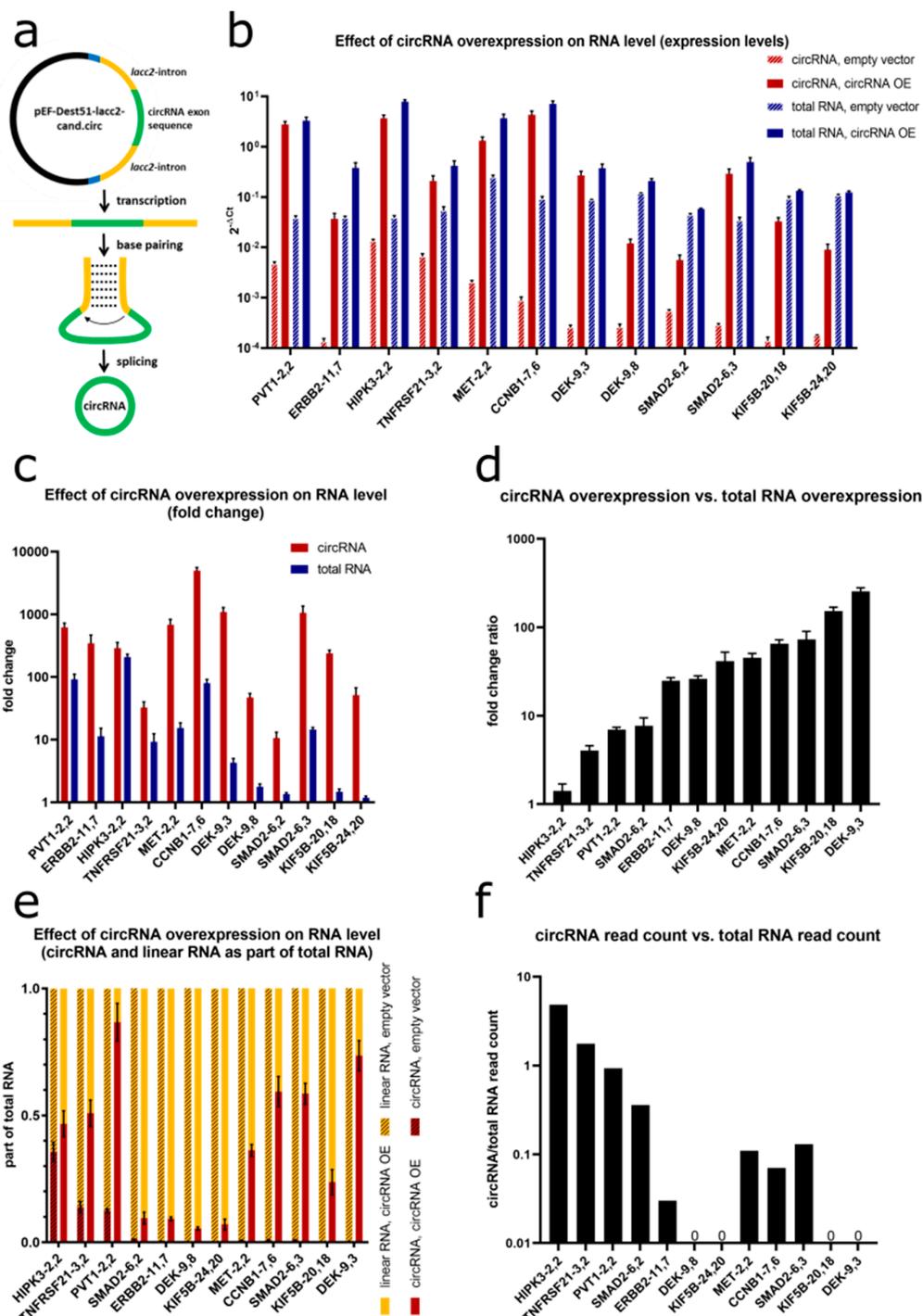
To further investigate the phenotype of circRNAs in NSCLC cell lines, overexpression vectors for 12 candidate circRNAs were designed. We used the pcDNA3.1-Laccase 2 designed by Kramer et al. [43], which contained a multiple cloning site flanked by the minimal circularization inducing sequences of the laccase2 introns [43]. Due to base-pairing between reverse complementary sequences within the introns, backsplicing was efficiently facilitated (Figure 10a). By combining gateway and restriction site cloning, we constructed the circRNA expression vector pEF-Dest51-lacc2-cand.circ with cand.circ representing the exonic sequence of the respective candidate circRNA. We transfected cell lines with low endogenous circRNA expression with circRNA overexpression vector or empty vector control. RT-qPCR confirmed the successful overexpression of all 12 candidate circRNAs. Since total RNA comprised not only linear RNA but also circRNA, an increase in total RNA after circRNA overexpression was accordingly observed (Figure 10b). Overexpression levels for circRNAs ranged between 11-fold for *circSMAD2* (backsplice site connecting exons 6-2) and 4996-fold for *circCCNB1* (backsplice site 7-6) (Figure 10c). For all candidates, the fold changes for circRNA were higher than for total RNA derived from the gene. Hence, the fraction of circRNA in total RNA increased after overexpression (Figure 10d). The ratio between circRNA overexpression and total RNA overexpression was quite consistent for each circRNA between the replicates, but heterogeneous between the individual circRNAs. The highest ratio was found for *circDEK* (backsplice 9-3) increasing 1083-fold whereas total *DEK-9,3* only increased 4-fold leading to a circRNA/total RNA fold change ratio of 255. At the other end, *HIPK3-2,2* showed the lowest ratio with 290-fold change for the circRNA vs. 208-fold change for the total RNA, resulting in an overall increase in the circular to total ratio of only 1.4. Due to the high variability in initial circRNA levels between the respective candidates and the different overexpression properties, the proportions of circRNA relative to total RNA after overexpression differed substantially among the circRNAs ranging between 6% for *circDEK* (backsplice site 9-8) and 86% for *circPVT1* (backsplice site 2-2) (Figure 10e).

Next, we analyzed whether the extent of the circRNA overexpression ratio (Figure 10d) correlated with the initial ratio of circRNA relative to total RNA (Figure 10f). Comparing the RNA-Seq read counts for the respective circRNAs and total RNAs in the FL3C dataset, we indeed found a negative correlation ( $R = -0,38$ ) between the endogenous circRNA/total RNA read count ratio and the circRNA/total RNA ratio change upon overexpression. This indicated that for circRNAs which were present in a high fraction of the total RNA already endogenously, also the overexpression of circRNA and total RNA remained rather parallel (e.g., *circHIPK3-2,2*). For circRNAs which made up only a very small fraction of the endogenous total RNA, the overexpression largely increased the fraction of circRNA relative to total RNA (e.g., *circDEK-9,3*).

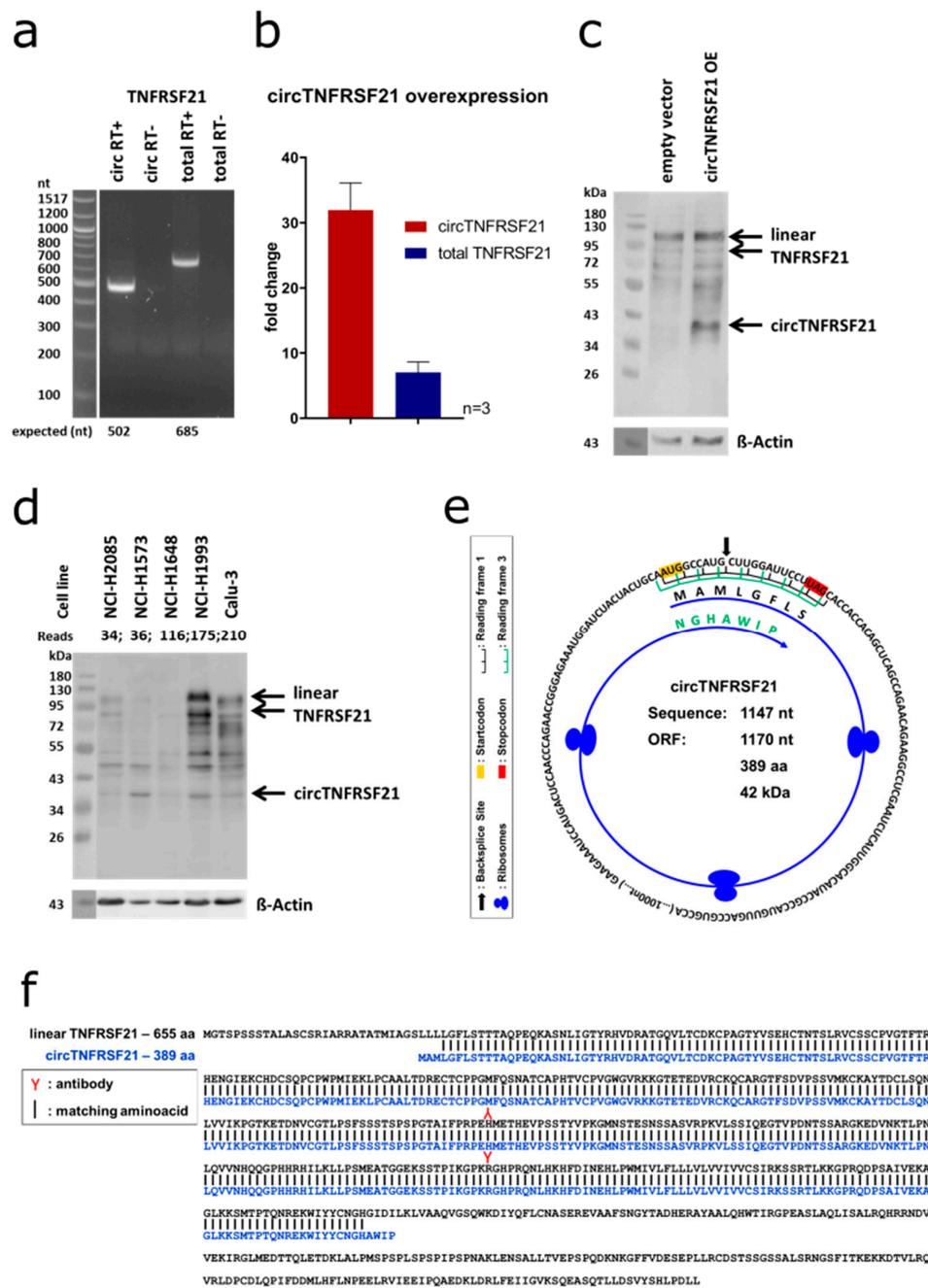
Overall, we successfully constructed vectors that efficiently and specifically overexpressed circRNAs at base-pair resolution.

### 2.10. circTNFRSF21 is Translated in Two Different Reading Frames Crossing the Backsplice Site Twice

One circRNA with high abundance and high differential expression in tumor cell lines versus non-transformed cell lines was *circTNFRSF21*. We validated the expression of *circTNFRSF21* backsplice site linking exon 2 to exon 3 (2-3) by PCR (Figure 11a,b).



**Figure 10.** CircRNA overexpression (a) vector: pEF-Dest51-lacc2-cand.circ, the candidate circRNA exonic sequence (cand.circ) is flanked by laccase 2 introns (lacc2). Base-pairing between laccase2 introns enhanced backsplicing of the circRNA exonic sequence. (b–e) Effect of circRNA overexpression on circRNA and total RNA levels measured by RT-qPCR normalized to Cyclophilin A. Cell lines were transfected with the circRNA overexpression vector or empty vector as control.  $n \geq 3$ . (b) Expression levels of circRNA and total RNA in controls and after overexpression depicted as  $2^{-\Delta Ct}$ . (c) Fold changes of circRNA and total RNA after overexpression compared to control. (d) Ratio between fold change of circRNA and fold change of total RNA after overexpression compared to control. (e) CircRNA and linear RNA as part of total RNA in controls and after overexpression. (f) Ratio between average circRNA and average total RNA read counts in RNA-Seq in the respective cell line used for overexpression. circRNA read count of zero in this cell line is depicted as 0.



**Figure 11.** Translation of *circTNFRSF21*. (a) Validation of *circTNFRSF21* expression (circ) and total *TNFRSF21* expression (total). Test samples: “RT+”, negative controls: “RT-” lacking reverse transcriptase. Numbers represent expected fragment sizes in number of nucleotides. (b) Signal strength of bands after gel electrophoresis quantification ( $n = 3$ ). (c) Effect of *circTNFRSF21* overexpression on protein synthesis. Representative western blot image after transfection.  $\beta$ -Actin was used as loading control, 30  $\mu$ g of total protein was loaded. (d) Endogenous protein expression of *TNFRSF21*. Representative western blot image.  $\beta$ -Actin was used as loading control, 30  $\mu$ g of total protein was loaded. (e) *circTNFRSF21* nucleotide sequence of 1147 nt and the predicted open reading frame (ORF) with 1170 nt. Start codon is marked in yellow, stop codon is marked in red. The frame crossing the backsplice site first is marked in black (reading frame +1), the frame after the second crossing of the backsplice site is marked in green (reading frame +3). (f) Amino acid sequence of *TNFRSF21* derived from the linear mRNA (655 aa, black) and derived from the *circTNFRSF21* (389 aa, blue). The antibody used for western blot was raised against a region around Histidin-235 (red Y) which is included in the sequence of both proteins.

This circRNA showed no significant fold change after RNase R treatment, in contrast to a significant decrease in the total *TNFRSF21* RNA (linear and circular) and *Cyclophilin A* RNA expression levels (Supplementary Figure S1). Next, we overexpressed *circTNFRSF21* (Figure 11b) and performed western blotting for TNFRSF21. The full-length TNFRSF21 protein gave rise to two bands of respectively 80 and 120 kDa as expected and detected in both overexpressed and control samples in comparable amounts indicating unaltered endogenous TNFRSF21 protein expression.

Upon *circTNFRSF21* overexpression, we noticed a prominent additional band at approx. 42 kDa (Figure 11c). This band was also present endogenously in several cell lines which showed endogenous *circTNFRSF21* expression at the RNA level (Figure 11d). When assessing the coding potential of *circTNFRSF21* with the NCBI ORFfinder, we identified an open reading frame in *circTNFRSF21*, that crosses the backsplice site twice in two different reading frames (a schematic representation can be found in Figure 11e) and results in a protein of an estimated molecular weight of 42 kDa (Figure 11f). In conclusion, *circTNFRSF21* could be translated into a 42 kDa protein.

### 2.11. *circRNA* Overexpression influences Colony Formation in NSCLC Cell Lines

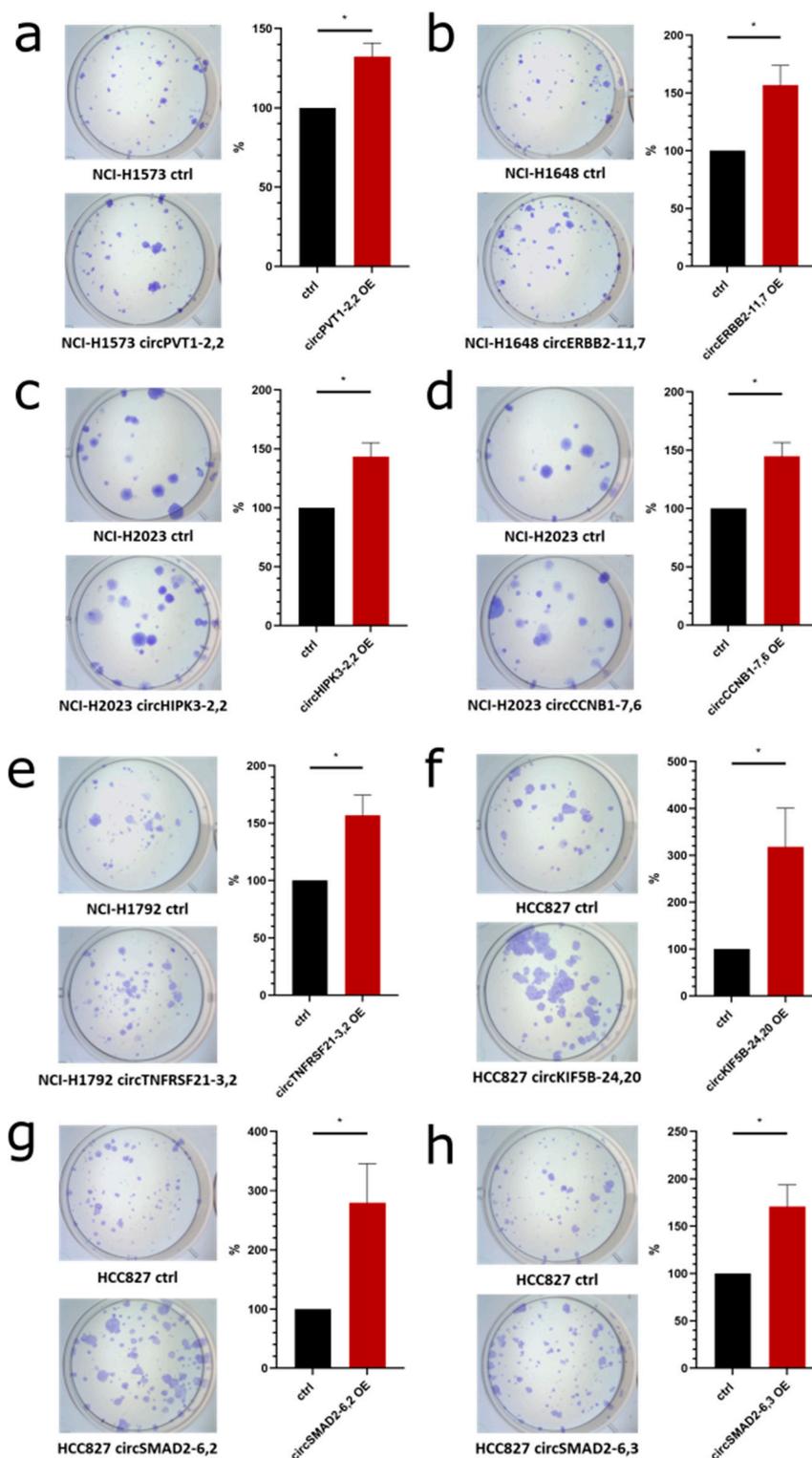
To determine the influence of circRNAs on the colony forming capacity of lung cancer cells, we overexpressed several circRNAs in cell lines with low endogenous expression of the respective circRNA and compared their impact on colony formation to empty vector controls.

We started with the overexpression of *circPVT1* (backsplice site 2–2) (Figure 12a), which is a known oncogenic circRNA [44–46], which resulted in a significant increase in colony formation as expected (Figure 12a). Furthermore, overexpression of *circERBB2* (backsplice site 11–7) which emanated from a known oncogenic driver gene in lung cancer [47] led to a significant increase in colony formation (Figure 12b).

As an example of a gene that had not been connected to cancer before (non-CGC gene), we selected *circHIPK3* which was highly expressed and showed a very high ratio of circular/total RNA expression. Overexpression of *circHIPK3* (backsplice site 2–2) also significantly increased colony formation (Figure 12c). Since *circCCNB1* (backsplice site 7–6) was derived from a gene involved in cell cycle control [48], we selected it and found a significant increase in colony formation upon overexpression (Figure 12d). Also for the overexpression of *circTNFRSF21* (backsplice site 3–2), which was differentially expressed in tumor vs. non-transformed cell lines (Figure 9a), we observed a significant increase in colony formation (Figure 12e).

Also, we analyzed the correlation of circRNA expression with the proliferation capacity which we had determined for all lung cell lines in our FL3C panel (Figure 8). *CircKIF5B* (backsplice site 24–20) showed a positive correlation of 0.29 with proliferation. In the colony formation assay, the overexpression of *circKIF5B* (backsplice site 24–20) caused a significant increase (Figure 12f). Moreover, we identified a highly positive correlation of two circRNAs generated from the *SMAD2* gene, a known cancer gene listed in CGC. Overexpression of *circSMAD2* (backsplice site 6–2,  $R = 0.37$ , Figure 12g) and *circSMAD2* (backsplice site 6–3,  $R = 0.35$ , Figure 12h) also showed a significantly increased colony formation.

In summary, these results demonstrate the functional impact of selected circRNAs identified in our resource validating its usefulness to find circRNAs for future functional studies to unravel their role in lung cancer.



**Figure 12.** CircRNA overexpression impact on colony formation in lung cancer cells. Left: representative images of colonies after staining with crystal violet. Right: area covered by colonies after overexpression depicted in % of control,  $n \geq 3$ . Cells were transfected with overexpression (OE) vector or empty vector as control (ctrl). (a) NCI-H1573 cells transfected with *circPVT1-2,2* OE vector. (b) NCI-H1648 cells transfected with *circERBB2-11,7* OE vector. (c) NCI-H2023 cells transfected with *circHIPK3-2,2* OE vector. (d) NCI-H2023 cells transfected with *circCCNB1-7,6* OE vector. (e) NCI-H1792 cells transfected with *circTNFRSF21-3,2* OE vector. (f) HCC827 cells transfected with *circKIF5B-24,20* OE vector. (g) HCC827 cells transfected with *circSMAD2-6,2* OE vector. (h) HCC827 cells transfected with *circSMAD2-6,3* OE vector. T-test: not significant (n.s.),  $p$ -value < 0.05 (\*).

### 3. Discussion

In our study, we created a comprehensive resource of the circRNA landscape in lung cells from 175 transcriptomes of rRNA-depleted RNA with 2.8 million backsplicing reads identifying and quantifying circRNAs. The depletion of rRNA allowed us to sequence simultaneously both the circRNA and the linear RNA transcriptome across the cell lines. In comparison, we found significantly and strongly decreased circRNA read counts in public polyA-enriched RNA-Seq datasets, which was not surprising given that circRNAs do not contain a polyA tail at their 3'-end. The frequently used TCGA and CCLE databases are both prepared using polyA selection [30,40]. Hence, while these are highly valuable tools for the analysis of polyadenylated RNAs (e.g., mRNAs), these should be used with caution for circRNAs. Another approach is the use of the exonuclease RNase R to digest most of the linear RNA by exonucleolytic degradation. This method will specifically enrich circRNA reads, but prevents a comparison between circRNA and total (linear) RNA since the linear RNA is depleted. Finally, exome capture was designed to detect both circRNA and linear RNA in samples with poor quality RNA [49]. Although this method resulted in a high amount of reads for both circRNA and linear RNA [16], the usage of probes might create a sequence bias and novel transcripts lacking complementarity to the probes cannot be detected.

We analyzed the data both at the gene level (grouping all the circRNAs arising from the same gene) and at the backsplice level (for every unique circRNA, i.e., every backsplice site defined by two linked exons). Strikingly, 14.1% of the genes formed circRNAs in  $\geq 90\%$  of the 60 cell lines, whereas only 0.7% of individual circRNAs were found in  $\geq 90\%$  of the cells. A similar pattern emerges from the analysis of the correlation between the circRNA relative to the total (linear) reads from the same locus: the correlation was much higher at the gene level (median 0.27) than at the backsplice level (median 0.06). This observation could point to a hypothesis that in general, the formation of circRNAs from a gene is more "important" and hence more homogeneous and more frequently correlated than forming a specific circRNA from the respective gene—which may be subject to future studies to unravel the functional importance of different circRNAs produced from the same locus.

Our data also illustrates a high cell line specificity for most circRNAs. This is corroborated by a comparison with MiOncoCirc [16], which reported a subset of 589 out of 12397 genes (4.8%) giving rise to circRNAs in  $>90\%$  of the cell lines tested. The higher fraction (14.1%) in our FL3C dataset is likely explained by the inclusion of different cancer entities in MiOncoCirc, while our dataset contains only lung cells and is hence derived from a more homogeneous tissue background.

When looking at the acceptor and donor exons, we noticed that the first and last exons were almost excluded from circRNA formation in comparison to the second and second-to-last exons, respectively. The predominance of the second exon as an acceptor exon is in accordance with previous data by Ragan et al. [50]. Mechanistically, the lack of acceptor exons in first exons may point towards the post-transcriptional nature of backsplicing [10], which is likely blocked by co-transcriptional 5'-capping. The last exon cannot act as donor exon since its 3'-end is cut and polyadenylated rather than spliced.

Focusing on the number of circRNAs originating from a single gene, *BIRC6* showed the highest number of circRNA isoforms in the FL3C dataset (418), which was in agreement with MiOncoCirc reporting more than 500 isoforms of *circBIRC6* [16]. The gene with the highest circRNA read counts in our data was *CYP24A1*, which has also previously been identified as the most highly expressed circRNA in the single A549 cell line [51].

Unsupervised hierarchical clustering of the top 100 most highly expressed circRNAs both at the gene level and at the backsplice level resulted in a separate cluster for the non-transformed cell lines, whereas this separation was less prominent between adenocarcinoma and non-adenocarcinoma NSCLC cell lines. Hence, circRNA expression levels could distinguish different cell types, which is in agreement with a previous study in hematopoietic cells [52].

The overall correlation between circRNA expression and total (linear) RNA expression was neutral or positive, with a stronger positive correlation at the gene level than at the backsplice level, while negative correlations were very rare and only weak. Given that the linear primary

RNA transcript is the precursor of both the mature mRNA and the circRNA, this trend to a positive correlation may not be unexpected. Several previous studies reported a lack of a significant correlation between circular and linear RNA. This difference may be explained largely by multiple reasons: first, several of these studies compared circular and linear RNA ratios in different heterogeneous tissues or different developmental stages [51,53], while our study analyzed the correlation of circular and linear RNA in a more homogeneous of cell lines derived from the same organ and most of them from the same tumor subtype—lung adenocarcinoma. Hence, there may be differential efficiency of RNA circularization between different tissues, but this may be less pronounced in the same tissue. Second, another study based the correlation analysis of circular and linear RNAs on only highly abundant circRNAs (279 out of 15,223) [54], while our study included 148,811 circRNAs (all) or 5232 circRNAs with expression in more than half of the cell lines. Lastly, multiple previous studies were based on RNA-seq of polyA-enriched RNA, while we used rRNA-depleted RNA which may also cause a different representation of the circRNAs. Nonetheless, also our data states that the correlation between circular and total (linear) RNA is weak in most cases.

The differences in correlation for the different circRNAs from the same parental mRNA might be explained by the different elongation rates of RNA polymerase II (reviewed in [55]). The production rate of circRNAs is positively correlated with the transcription elongation rate [10]. A second factor influencing the circRNA biogenesis is the availability of the splicing machinery [56] and the efficiency of canonical splicing [8]. When splicing is less efficient, the RNA has more opportunities to circularize upon itself which eventually leads to backsplicing. This circularization might be influenced by flanking introns containing complementary sequences [11], RNA binding proteins [13,14] or RNA editing [11]. Taken together, both the transcription elongation rate and splicing contribute to the formation of circRNA, and might explain the differences in circRNA expression levels between circRNAs originating from the same mRNA.

The known cancer genes (CGC) [42] produced on average more circRNAs in comparison to non-CGC genes. There are two possible explanations: either these CGC genes are more abundantly transcribed in comparison to non-CGC genes and thus there are more opportunities for circRNAs to be formed as a by-product from canonical splicing. Or these circRNAs are produced to fulfill an oncogenic function in the cell—which will be a relevant topic for elucidation in the future.

When correlating circRNA expression to the proliferation rate of the cell lines, there was a larger subset of circRNAs that were only expressed in the slow and a smaller subset of circRNAs that were only expressed in fast proliferating cells. This result was in accordance with previous reports [15,16] and warrants future research into a functional role for the differentially regulated circRNAs. So far, there is no verified explanation for the general decrease in circRNA expression in fast proliferating cells, but possibly the stable circRNAs could be diluted by cell division [15]. Another possibility is a deregulation of the splicing machinery thus influencing the production of circRNAs. However, more research is needed to elucidate the exact regulatory mechanism in highly proliferating cells.

Adenocarcinoma cell lines showed a higher expression of circRNAs in comparison with non-adenocarcinoma NSCLC cell lines. This may in part be due to a difference in the average proliferation rate of the cells [15]. The seven non-LUAD cell lines have an average proliferation rate of a 6.0-fold increase in cell count after 72 h, in comparison to a 3.3-fold increase for the 50 LUAD cell lines (significant:  $T$ -test  $p = 0.0014$ ). Among the genes with a significant change in circRNA expression, some had been previously identified as markers for LUAD: *TALDO1* [57], *TMC5* [58], *MVP* [59] and *FNDC3B* [60]. This confirms that our cell line panel is able to recapitulate biomarker patterns at the circRNA level which were previously established at the linear mRNA level for different NSCLC histological subtypes. When comparing *TP53* wild-type versus *TP53* mutant cell lines, there were two circRNA-producing genes that showed a significant change: *SBDS* and *GBF1*, with *SBDS* having previously been linked to *TP53* [61].

Comparing tumor cells to non-transformed cell lines identified a number of significantly differentially expressed circRNAs—including *circTNFRSF21*. Overexpression of this circRNA gives rise to an additional band matching in size to an open reading frame (ORF) in *circTNFRSF21* which crosses the backsplice site twice in two different reading frames. The translation of circRNAs has been described previously [62]. Given the lack of a 5'-cap structure, mechanisms like an IRES sequence [63] or m6A methylation could initiate translation [64]. Eukaryotic cells are capable of translating circRNAs according to a rolling circle mechanism [65], which is fitting with the ORF in *circTNFRSF21* crossing the backsplice site twice. Further research is warranted to find out the translation initiation mechanism of *circTNFRSF21* and its functional role in lung cancer.

We showed that overexpression of circRNAs is possible, using an adapted method that was originally described by Kramer et al. [43], using the highly complementary sequence of the *Laccase2* introns from *Drosophila melanogaster*. The base-pairing between the intronic sequences brings the donor and acceptor exon into close proximity to each other, thus creating favorable conditions for circularization [43,66]. However, the efficiency in overexpressing circRNAs was heterogeneous. This reflects in part the endogenous ratio of circular to total RNA, but might also be explained by the role of other factors like the different transcription efficiency [10].

Overexpression of circRNAs allowed us to study their phenotype and influence on colony formation. First, circRNAs originating from known cancer genes (CGC-listed genes) also showed an oncogenic effect and their overexpression in low endogenously expressing cell lines led to increased colony formation, e.g., for *circPVT1*, which corroborates previous studies [45,46]. Second, we identified circRNAs from known cancer genes and from non-cancer genes to influence lung cancer cell colony formation. Altogether, these data indicate that further functional studies are warranted to unravel the oncogenic potential of these circRNAs.

In summary, our study provides a deep insight into the expression of circRNAs in lung cell lines. Next to our findings on global circRNA properties like correlation with linear RNA, exon distribution, distinction between non-transformed and tumor cell lines by PCA, enrichment in cancer genes and association with cell proliferation and genotypes, we provide individual examples of circRNAs governing the colony formation potential of lung cancer cells as well as *circTNFRSF21* being translated with two reading frames across the backsplice site. For the future, our resource can be widely used to identify new circRNAs in lung cancer, to find cell lines with endogenous expression of specific circRNAs for further experiments, to search for genes with a particularly high or low circularization rate, to select circRNAs correlated with cell proliferation for mechanistic studies or to identify circRNAs associated with certain genotypes in lung adenocarcinoma.

## 4. Materials and Methods

### 4.1. Cell Culture

All cell lines were purchased from American Type Culture Collection (ATCC, Manassas, VA, USA) and were maintained in ATCC-modified RPMI-1640 medium (cat#A1049101, Life Technologies, Carlsbad, CA, USA) at 37 °C and 5% CO<sub>2</sub> in cell culture dishes (TPP, Trasadingen, Switzerland) [67]. At maximum 80% confluence, cells were detached using 0.25% trypsin (Life Technologies, Carlsbad, CA, USA) and splitted 1:3–1:5 dependent on the cell line. Cells were regularly tested for mycoplasma contamination. The total FL3C database contains 50 lung adenocarcinoma cell lines (LUAD), seven non-LUAD NSCLC cell lines and three non-transformed cell lines (Supplementary Table S1). Genotype data was downloaded from the COSMIC Cell line project, v90 [42].

The proliferative capacity of the cell lines was determined as follows: 150,000 cells per well were seeded into a 6-well plate and trypsinized after 24 h, 48 h and 72 h, respectively (2-wells per timepoint). At each timepoint, the cells were counted in duplicate. Ratios between the starting amount of cells (150,000/well) and cell count after 24, 48 and 72 h were determined. The increase in ratios represented the proliferative capacity of the cell line and is provided in Supplementary Table S1 (72 h).

#### 4.2. RNA-Seq Library Construction and circRNA Analysis

RNA was isolated with the RNeasy mini kit (Qiagen, Hilden, Germany). An on-column DNase I digest was performed with the DNase I RNase-free DNase set (Qiagen). From 5 µg total RNA, ribosomal RNA (rRNA) was depleted with the RiboZero Gold Kit H/R/M (Illumina, San Diego, CA, USA). The RNA library was constructed with the SureSelect Strand-Specific RNA library prep for the Illumina Multiplexed Sequencing Protocol (Illumina) and sequenced with the HiSeq4000 paired end 100 bp kit (Illumina). All kits were used according to manufacturers' instructions.

Raw reads were mapped using Tophat2 [68,69] with the following parameters: -a 6 -m 2 -g 1 -p 16. Reads that were not aligned with Tophat2 were filtered and aligned by Tophat2 fusion using the following parameters: - fusion search - p 15 - keep fasta order - bowtie1 - no coverage search. The circExplorer2 [70] pipeline was used to identify circRNAs using standard parameters. CircRNAs were length-normalized using the fragment-length as determined by DESeq [71]. All reads were mapped to the Ensembl GRCh38 gene set.

The number of reads mapping to the backsplice side were used to calculate the circRNA expression level, while the sum of the total linear reads mapping to the parental gene was used to measure the total (linear) RNA expression level. The circular/total ratio was calculated by dividing the normalized backsplice reads (normalized to library size) by the linear FPKM. All statistical analyses were performed in R and the R-packages described (v3.6.0) [72].

#### 4.3. RNA Isolation for RT-(q)PCR

Cells were washed with 1x PBS (Sigma-Aldrich, St. Louis, MO, USA) and lysed in Tri-reagent (Sigma-Aldrich). RNA was isolated according to the manufacturer's instructions. DNA was digested with DNase I (Roche, Mannheim, Germany) for 30 min at 37 °C. Reverse transcription was performed on 1 µg of RNA with RevertAid Reverse Transcriptase (ThermoFisher, Schwerte, Germany) and random hexamer primers (ThermoFisher).

#### 4.4. PCR and Quantitative Real-Time PCR

cDNA was amplified using the Q5 high fidelity polymerase (New England Biolabs, Ipswich, MA, USA) following the manufacturer's instructions according to the following program: an initial incubation of 30 s at 98 °C, followed by 30 cycles of 10 s 98 °C, 30 s 58 °C and 45 s 72 °C, followed by a final incubation of 5 min at 72 °C in a Bio-Rad T100 Thermal Cycler. Next, the PCR-reaction was mixed with DNA loading dye (New England Biolabs) and separated on a 1% or 2% agarose gel containing SybrSafe (Thermo Fisher) and run in 1× TAE (40 mM Tris Base, 20 mM acetic acid, 1.27 mM EDTA, pH 8.3). The gel was imaged with the Gel iX20 imager (Intas, Göttingen, Germany). Band quantification was performed with Image Studio Lite software (Licor, Lincoln, NE, USA).

The expression of the circRNA and its corresponding total RNA were determined via RT-qPCR using 2× Power SybrGreen Master Mix (Applied Biosystems, Foster City, CA, USA). Reactions were run in triplicate on the Applied Biosystems StepOnePlus cycler using the following program: 10 min incubation at 95 °C prior to 40 cycles of 15 s at 95 °C and 30 s at 60 °C. *Cyclophilin A* was used as housekeeper gene for normalization. The relative expression was determined by the  $2^{-\Delta\Delta C_t}$  method using average  $C_t$  values [73]. All primer sequences for PCR, RT-qPCR and cloning are listed in Supplementary Table S5.

#### 4.5. RNase R Treatment

One µg of total RNA was treated with 10 U of RNase R (Lucigen, Middleton, WI, USA) in 10 µL of total volume with 0.2 µL 50 mM MgCl<sub>2</sub> in PCR Single Cap 8er-SoftStrips reaction tubes for 60 min at 37 °C in a T100 Thermal cycler (Bio-Rad, Feldkirchen, Germany) according to manufacturer's instructions. The reaction was prepared in parallel in duplicates. After incubation, one reaction mix was

used for the RT-PCR and the second reaction served as non-enzyme control (no reverse transcriptase was added to this reaction). Protocols were further executed as described in Sections 4.3 and 4.4.

#### 4.6. Cloning

First, the minimal Laccase2 intronic sequence, including a multiple cloning site (MCS), from the pcDNA3.1(+) Laccase2 MCS Exon Vector (Addgene, Cambridge, MA, USA) [43] was amplified and inserted into the pDONR221 vector by BP Gateway cloning (ThermoFisher).

Second, the complete sequence of the circRNA was amplified by PCR (see above) followed by sequencing to assess which exons were included. The amplicons were separated on an agarose gel, the target band was purified using the GeneJET Gel extraction kit (ThermoFisher) and the bands were sequenced with SupremeRun Tube Sanger Sequencing (GATC Eurofins, Ebersberg, Germany).

Third, the target sequence was amplified by Q5-PCR (see above), separated on an agarose gel and purified. The circRNA sequence was inserted into the MCS using restriction site cloning with either PacI and AscI or PstI-HF and AscI (all from New England Biolabs). Next, the entry clones were recombined into the pEF-DEST51 Gateway expression vector (ThermoFisher) using the LR Clonase II enzyme mix (ThermoFisher) according to manufacturer's instructions.

Lastly and importantly, to delete the remaining sequences of the MCS at both ends of the circRNA target sequences, a two-step PCR protocol was used. With a first PCR, the circRNA target sequence was amplified in two parts using overlapping primer pairs lacking the MCS sequence. In a second step, the outer primers were used to create the full target sequence. This was inserted into the expression vector by restriction enzyme cloning.

#### 4.7. Colony Formation Assay

Cells were seeded in a 6 cm cell culture dish (TPP) and a 12-well plate (Greiner Bio-One, Kremsmünster, Austria) and allowed to attach overnight. A 1000–2000/well cells were seeded in the 12-well plate for colony formation assay, 250,000 cells were seeded in the 6 cm dish for RNA isolation. Next day, the cells were transfected with 1 µg (12-well) or 5 µg (6 cm culture dish) of plasmid using 1 mg/mL polyethylenimine (PEI) (Polysciences Inc., Warrington, PA USA) in a 1:3 ratio in OPTI-MEM (Gibco, ThermoFisher, Carlsbad, CA, USA) according to the manufacturer's instructions. The empty vector was used as negative control.

After 48 h, cells from the 6 cm culture dish were lysed and RNA was isolated as previously described. Reverse transcription and quantitative real-time PCR was used to determine the level of overexpression of both the circRNA and the corresponding total (linear) RNA.

For the colony formation assay, cells in the 12-well plate were allowed to grow during 8–17 days, depending on the proliferation rate of the cell line (NCI-H1573: 10 days, NCI-H2023: 15–17 days, HCC827 and NCI-H460: 8–10 days). The medium was changed every 4 days during this period. Afterwards, cells were washed with 1× ice-cold PBS (Sigma) and fixed with 100% methanol (Carl Roth, Karlsruhe, Germany) during 10 min at −20 °C. Cells were stained with 0.5% crystal violet (Carl Roth) dissolved in 25% methanol in ddH<sub>2</sub>O during 10 min at room temperature. Remaining stain was removed by washing with ddH<sub>2</sub>O and the plate was air-dried overnight. The wells were imaged with a stereo Discovery V8 microscope (Carl Zeiss AG, Oberkochen, Germany) equipped with a 0.3× objective and 1.0× zoom and colony areas were determined using ImageJ [74].

#### 4.8. Western Blotting

Cells were seeded in a 6 cm dish and transfected as described in 4.7. After 48 h, cells were lysed on ice in 1× RIPA buffer (150 mM NaCl, 1% Triton X-100, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris pH8, freshly supplemented with Complete protease inhibitor cocktail (cat#04693159001, Roche, Basel, Switzerland) and spun down for 15 min at 12,000× g at 4 °C. The lysate was transferred to a new tube. The protein concentration was measured with the BCA assay (Sigma, Darmstadt, Germany). Thirty µg of total protein was loaded on a 10% SDS-PAGE gel and run for 1 h at 120V. Next, proteins

were semi-dry blotted onto a nitrocellulose membrane (1 h 20 min at 0.07 A). The membrane was blocked in 5% skimmed milk during 1 h at room temperature. The membrane was incubated overnight at 4 °C with the primary antibody for TNFRSF21 (DR6, clone E8D2I Rabbit mAb, #93026, Cell Signaling, Leiden, The Netherlands) in a 1:1000 dilution in 5% BSA in 1× TBS (24.7 mM Tris base, 137 mM NaCl, 2.68 mM KCl, pH 7.4) with 0.1% Tween-20. After washing with 1× TBS with 0.1% Tween-20, the membrane was incubated with the secondary antibody (Peroxidase-AffiniPure Goat Anti-Rabbit IgG, Dianova, Hamburg, Germany, 1:5000 dilution in 5% BSA in 1× TBS with 0.1% Tween-20) during 2 h at room temperature. After washing, bands were detected with SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) and visualized with the Intas ECL Chemocam Imager (Intas). As loading control, membranes were stained for β-Actin (clone AC74, β-actin mouse mAb, Sigma) in a 1:10000 dilution in 5% BSA in 1× TBS with 0.1% Tween-20, incubated for 2 h at room temperature. After washing, the membrane was incubated with the secondary antibody for 2 h at room temperature (Peroxidase-AffiniPure Goat Anti-Mouse IgG, Dianova, 1:10,000 dilution in 5% BSA in 1× TBS with 0.1% Tween-20).

#### 4.9. Statistics

Statistics were performed in Excel (Microsoft, Redmond, WA, USA) and GraphPad Prism (version 8, GraphPad Software, San Diego, CA, USA). Statistical significance was assessed with a two-sided *T*-test after performing an *F*-test to test scedasticity. Experiments were always performed at least in biological triplicates. The RT-qPCR was performed in three and colony assays were performed in three to six technical replicates per biological replicate. Colony assay replicates with a cell colony number of zero were excluded from further calculations. The correlations per histology or genotype were performed by multiple *T*-tests (*T*-test per row) with Bonferroni correction for multiple testing. Groups with a mean value of zero were excluded from statistical analysis.

## 5. Conclusions

In this study, we present a comprehensive map of the circRNA landscape in the FL3C panel of 60 lung cell lines based on 2.8 million backsplice reads derived from deep sequencing of rRNA-depleted RNA. Most circRNAs show high cell line specificity and generally a positive correlation between circRNA and total (linear) RNA expression levels. Known cancer genes tend to produce more circRNAs. CircRNA patterns can be used for classification and clustering of cell lines or reflect tumor histology, genotype and proliferation rate. Correlation of cell proliferation with circRNA expression can identify circRNAs whose overexpression has a significant impact on colony formation of lung cancer cells. Moreover, *CircTNFRSF21* can be translated into a 42 kDa protein crossing the backsplice site in two different reading frames. In conclusion, this lung cell line dataset advances our understanding of circRNA expression and will enable future research on the function and relevance of circRNAs in lung cancer.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-6694/12/5/1091/s1>, Supplementary Table S1: Overview of the used cell lines; Supplementary Table S2: Full circRNA dataset at gene level; Supplementary Table S3: Full circRNA dataset at backsplice level; Supplementary Table S4: Differentially expressed circRNAs; Supplementary Table S5: Primer sequences (subtables for primers used for PCR, RT-qPCR, cloning and mutagenesis, all sequences in 5' to 3' direction). Supplementary Figure S1: Resistance of candidate circRNAs to exonucleolytic degradation by RNase R; Supplementary Figure S2: Validation of proliferation-correlated circSMAD2-6,2 in fast vs. slowly proliferating cell lines with RT-qPCR. The first two cell lines (NCI-H1944, NCI-H1993) were fast proliferating (≥5-fold increase in cell count in 72 h), the last two cell lines (HCC827, NCI-H2347) were slowly proliferating (≤3-fold cell count increase in 72 h). Expression levels of circSMAD2-6,2 and total SMAD2-6,2 are depicted as % of maximum expression level (in NCI-H1993). *T*-test: not significant (n.s.), *p*-value < 0.05 (\*), *p*-value < 0.01 (\*\*), *p*-value < 0.001 (\*\*). Below the graph, the fold change of average relative expression level between the groups is shown.

**Author Contributions:** Conceptualization, S.D. and N.V.D.S.; RNA isolation and sequencing, A.C.B. and J.S. software and sequence analysis, Y.L.; data analysis, S.D. and N.V.D.S.; validation, N.V.D.S. and A.K.H.; functional analysis, A.K.H.; writing—original draft preparation, N.V.D.S. and S.D.; writing—review and editing, all authors;

supervision & funding acquisition, S.D.; All authors have read and agreed to the published version of the manuscript. Parts of this manuscript will be part of the Ph.D. thesis of Y.L. and the M.D. thesis of A.K.H.

**Funding:** Lung cancer research in our lab is supported by the German Cancer Consortium DKTK—FR04 and the German Cancer Aid (Deutsche Krebshilfe), grant number 70113349.

**Acknowledgments:** The authors would like to thank the DKFZ Genomics and Proteomics Core Facility for support, expert advice and RNA sequencing and the DKFZ HUSAR server platform for providing the storage space and computing capacity to execute this project. The authors are grateful to Jana Seiler for her help in setting up the lung cell line panel as well as to all members of the Diederichs labs for helpful discussions. Lastly, we acknowledge the Lighthouse Core Facility from the University Hospital Freiburg for their help with taking high quality pictures. Flanking sequences for circRNA overexpression were provided from the Jeremy Wilusz lab via Addgene (#69893) and recloned.

**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. S.D. is co-owner of siTOOLS Biotech GmbH in Martinsried, Germany, with no relation to this study.

## References

1. Cocquerelle, C.; Daubersies, P.; Majérus, M.A.; Kerckaert, J.P.; Bailleul, B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* **1992**, *11*, 1095–1098. [[CrossRef](#)] [[PubMed](#)]
2. Cocquerelle, C.; Mascrez, B.; Héтуin, D.; Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* **1993**, *7*, 155–160. [[CrossRef](#)] [[PubMed](#)]
3. Nigro, J.M.; Cho, K.R.; Fearon, E.R.; Kern, S.E.; Ruppert, J.M.; Oliner, J.D.; Kinzler, K.W.; Vogelstein, B. Scrambled exons. *Cell* **1991**, *64*, 607–613. [[CrossRef](#)]
4. Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S.D.; Gregersen, L.H.; Munschauer, M.; et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **2013**, *495*, 333–338. [[CrossRef](#)]
5. Enuka, Y.; Lauriola, M.; Feldman, M.E.; Sas-Chen, A.; Ulitsky, I.; Yarden, Y. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.* **2016**, *44*, 1370–1383. [[CrossRef](#)]
6. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **2013**, *19*, 141–157. [[CrossRef](#)]
7. Salzman, J.; Gawad, C.; Wang, P.L.; Lacayo, N.; Brown, P.O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **2012**, *7*. [[CrossRef](#)]
8. Ashwal-Fluss, R.; Meyer, M.; Pamudurti, N.R.; Ivanov, A.; Bartok, O.; Hanan, M.; Evantal, N.; Memczak, S.; Rajewsky, N.; Kadener, S. CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol. Cell* **2014**, *56*, 55–66. [[CrossRef](#)]
9. Starke, S.; Jost, I.; Rossbach, O.; Schneider, T.; Schreiner, S.; Hung, L.H.; Bindereif, A. Exon circularization requires canonical splice signals. *Cell Rep.* **2015**, *10*, 103–111. [[CrossRef](#)]
10. Zhang, Y.; Xue, W.; Li, X.; Zhang, J.; Chen, S.; Zhang, J.L.; Yang, L.; Chen, L.L. The Biogenesis of Nascent Circular RNAs. *Cell Rep.* **2016**, *15*, 611–624. [[CrossRef](#)]
11. Ivanov, A.; Memczak, S.; Wyler, E.; Torti, F.; Porath, H.T.; Orejuela, M.R.; Piechotta, M.; Levanon, E.Y.; Landthaler, M.; Dieterich, C.; et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* **2015**, *10*, 170–177. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, X.O.; Wang, H.B.; Zhang, Y.; Lu, X.; Chen, L.L.; Yang, L. Complementary sequence-mediated exon circularization. *Cell* **2014**, *159*, 134–147. [[CrossRef](#)] [[PubMed](#)]
13. Conn, S.J.; Pillman, K.A.; Toubia, J.; Conn, V.M.; Salamanidis, M.; Phillips, C.A.; Roslan, S.; Schreiber, A.W.; Gregory, P.A.; Goodall, G.J. The RNA binding protein quaking regulates formation of circRNAs. *Cell* **2015**, *160*, 1125–1134. [[CrossRef](#)] [[PubMed](#)]
14. Errichelli, L.; Dini Modigliani, S.; Laneve, P.; Colantoni, A.; Legnini, I.; Capauto, D.; Rosa, A.; De Santis, R.; Scarfò, R.; Peruzzi, G.; et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **2017**, *8*, 1–11. [[CrossRef](#)] [[PubMed](#)]

15. Bachmayr-Heyda, A.; Reiner, A.T.; Auer, K.; Sukhbaatar, N.; Aust, S.; Bachleitner-Hofmann, T.; Mesteri, I.; Grunt, T.W.; Zeillinger, R.; Pils, D. Correlation of circular RNA abundance with proliferation—Exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.* **2015**, *5*, 8057. [[CrossRef](#)] [[PubMed](#)]
16. Vo, J.N.; Cieslik, M.; Zhang, Y.; Shukla, S.; Xiao, L.; Zhang, Y.; Wu, Y.M.; Dhanasekaran, S.M.; Engelke, C.G.; Cao, X.; et al. The Landscape of Circular RNA in Cancer. *Cell* **2019**, *176*, 869–881.e13. [[CrossRef](#)]
17. Legnini, I.; Di Timoteo, G.; Rossi, F.; Morlando, M.; Briganti, F.; Sthandier, O.; Fatica, A.; Santini, T.; Andronache, A.; Wade, M.; et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **2017**, *66*, 22–37.e9. [[CrossRef](#)]
18. Pamudurti, N.R.; Bartok, O.; Jens, M.; Ashwal-Fluss, R.; Stottmeister, C.; Ruhe, L.; Hanan, M.; Wyler, E.; Perez-Hernandez, D.; Ramberger, E.; et al. Translation of CircRNAs. *Mol. Cell* **2017**, *66*, 9–21.e7. [[CrossRef](#)]
19. Yang, Y.; Fan, X.; Mao, M.; Song, X.; Wu, P.; Zhang, Y.; Jin, Y.; Yang, Y.; Chen, L.L.; Wang, Y.; et al. Extensive translation of circular RNAs driven by N<sup>6</sup>-methyladenosine. *Cell Res.* **2017**, *27*, 626–641. [[CrossRef](#)]
20. Yang, Y.; Gao, X.; Zhang, M.; Yan, S.; Sun, C.; Xiao, F.; Huang, N.; Yang, X.; Zhao, K.; Zhou, H.; et al. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* **2018**, *110*, 304–315. [[CrossRef](#)]
21. Zhang, M.; Huang, N.; Yang, X.; Luo, J.; Yan, S.; Xiao, F.; Chen, W.; Gao, X.; Zhao, K.; Zhou, H.; et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* **2018**, *37*, 1805–1814. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, M.; Zhao, K.; Xu, X.; Yang, Y.; Yan, S.; Wei, P.; Liu, H.; Xu, J.; Xiao, F.; Zhou, H.; et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
23. Hansen, T.B.; Jensen, T.I.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* **2013**, *495*, 384–388. [[CrossRef](#)] [[PubMed](#)]
24. Hu, W.; Bi, Z.Y.; Chen, Z.L.; Liu, C.; Li, L.L.; Zhang, F.; Zhou, Q.; Zhu, W.; Song, Y.Y.; Zhan, B.T.; et al. Emerging landscape of circular RNAs in lung cancer. *Cancer Lett.* **2018**, *427*, 18–27. [[CrossRef](#)] [[PubMed](#)]
25. Ng, W.L.; Mohd Mohidin, T.B.; Shukla, K. Functional role of circular RNAs in cancer development and progression. *RNA Biol.* **2018**, *15*, 995–1005. [[CrossRef](#)] [[PubMed](#)]
26. Liang, W.C.; Wong, C.W.; Liang, P.P.; Shi, M.; Cao, Y.; Rao, S.T.; Tsui, S.K.W.; Waye, M.M.Y.; Zhang, Q.; Fu, W.M.; et al. Translation of the circular RNA circ $\beta$ -catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol.* **2019**, *20*, 1–12. [[CrossRef](#)]
27. Du, W.W.; Yang, W.; Liu, E.; Yang, Z.; Dhaliwal, P.; Yang, B.B. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.* **2016**, *44*, 2846–2858. [[CrossRef](#)]
28. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)]
29. Herbst, R.S.; Morgensztern, D.; Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **2018**, *553*, 446–454. [[CrossRef](#)]
30. The Cancer Genome Atlas Research Network Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. [[CrossRef](#)]
31. Gutschner, T.; Hammerle, M.; Eißmann, M.; Hsu, J.; Kim, Y.; Hung, G.; Revenko, A.; Arun, G.; Stentrup, M.; Groß, M.; et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **2013**, *73*, 1180–1189. [[CrossRef](#)] [[PubMed](#)]
32. Olivero, C.E.; Martínez-Terroba, E.; Zimmer, J.; Liao, C.; Tesfaye, E.; Hooshdaran, N.; Schofield, J.A.; Bendor, J.; Fang, D.; Simon, M.D.; et al. p53 Activates the Long Noncoding RNA Pvt1b to Inhibit Myc and Suppress Tumorigenesis. *Mol. Cell* **2020**, *77*, 761–774.e8. [[CrossRef](#)] [[PubMed](#)]
33. Roth, A.; Boulay, K.; Groß, M.; Polycarpou-Schwarz, M.; Mallette, F.A.; Regnier, M.; Bida, O.; Ginsberg, D.; Warth, A.; Schnabel, P.A.; et al. Targeting LINC00673 expression triggers cellular senescence in lung cancer. *RNA Biol.* **2018**, *15*, 1499–1511. [[CrossRef](#)] [[PubMed](#)]
34. De Fraipont, F.; Gazzeri, S.; Cho, W.C.; Eymin, B. Circular RNAs and RNA splice variants as biomarkers for prognosis and therapeutic response in the liquid biopsies of lung cancer patients. *Front. Genet.* **2019**, *10*, 1–15. [[CrossRef](#)] [[PubMed](#)]

35. Tan, S.; Gou, Q.; Pu, W.; Guo, C.; Yang, Y.; Wu, K.; Liu, Y.; Liu, L.; Wei, Y.Q.; Peng, Y. Circular RNA F-circEA produced from EML4-ALK fusion gene as a novel liquid biopsy biomarker for non-small cell lung cancer. *Cell Res.* **2018**, *28*, 693–695. [[CrossRef](#)]
36. Tan, S.; Sun, D.; Pu, W.; Gou, Q.; Guo, C.; Gong, Y.; Li, J.; Wei, Y.Q.; Liu, L.; Zhao, Y.; et al. Circular RNA F-circEA-2a derived from EML4-ALK fusion gene promotes cell migration and invasion in non-small cell lung cancer 06 Biological Sciences 0601 Biochemistry and Cell Biology. *Mol. Cancer* **2018**, *17*, 1–5. [[CrossRef](#)]
37. Dong, Y.; Xu, T.; Zhong, S.; Wang, B.; Zhang, H.; Wang, X.; Wang, P.; Li, G.; Yang, S. Circ\_0076305 regulates cisplatin resistance of non-small cell lung cancer via positively modulating STAT3 by sponging miR-296–5p. *Life Sci.* **2019**, *239*, 116984. [[CrossRef](#)]
38. Yu, W.; Peng, W.; Sha, H.; Li, J. Hsa\_circ\_0003998 Promotes Chemoresistance via Modulation of miR-326 in Lung Adenocarcinoma Cells. *Oncol. Res.* **2019**, *27*, 623–628. [[CrossRef](#)]
39. Zhou, Y.; Zheng, X.; Xu, B.; Chen, L.; Wang, Q.; Deng, H.; Jiang, J. Circular RNA hsa\_circ\_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem. Biophys. Res. Commun.* **2019**, *508*, 527–535. [[CrossRef](#)]
40. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R.; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**. [[CrossRef](#)]
41. The Broad Institute of MIT & Harvard CCLE. Available online: <https://portals.broadinstitute.org/ccle> (accessed on 10 January 2020).
42. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [[CrossRef](#)] [[PubMed](#)]
43. Kramer, M.C.; Liang, D.; Tatomer, D.C.; Gold, B.; March, Z.M.; Cherry, S.; Wilusz, J.E. Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* **2015**, *29*, 2168–2182. [[CrossRef](#)] [[PubMed](#)]
44. Panda, A.C.; Grammatikakis, I.; Kim, K.M.; De, S.; Martindale, J.L.; Munk, R.; Yang, X.; Abdelmohsen, K.; Gorospe, M. Identification of senescence-associated circular RNAs (SAC-RNAs) reveals senescence suppressor CircPVT1. *Nucleic Acids Res.* **2017**, *45*, 4021–4035. [[CrossRef](#)] [[PubMed](#)]
45. Qin, S.; Zhao, Y.; Lim, G.; Lin, H.; Zhang, X.; Zhang, X. Circular RNA PVT1 acts as a competing endogenous RNA for miR-497 in promoting non-small cell lung cancer progression. *Biomed. Pharmacother.* **2019**, *111*, 244–250. [[CrossRef](#)] [[PubMed](#)]
46. Li, X.; Zhang, Z.; Jiang, H.; Li, Q.; Wang, R.; Pan, H.; Niu, Y.; Liu, F.; Gu, H.; Fan, X.; et al. Circular RNA circPVT1 promotes proliferation and invasion through sponging miR-125b and activating E2F2 signaling in non-small cell lung cancer. *Cell. Physiol. Biochem.* **2018**, *51*, 2324–2340. [[CrossRef](#)]
47. Collisson, E.A.; Campbell, J.D.; Brooks, A.N.; Berger, A.H.; Lee, W.; Chmielecki, J.; Beer, D.G.; Cope, L.; Creighton, C.J.; Danilova, L.; et al. Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature* **2014**, *511*, 543–550. [[CrossRef](#)]
48. Zhang, H.; Zhang, X.; Li, X.; Meng, W.B.; Bai, Z.T.; Rui, S.Z.; Wang, Z.F.; Zhou, W.C.; Jin, X. Da Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer. *J. Cell. Physiol.* **2018**, *234*, 619–631. [[CrossRef](#)]
49. Cieslik, M.; Chugh, R.; Wu, Y.-M.; Wu, M.; Brennan, C.; Lonigro, R.; Su, F.; Wang, R.; Siddiqui, J.; Mehra, R.; et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **2015**, *25*, 1372–1381. [[CrossRef](#)]
50. Ragan, C.; Goodall, G.J.; Shirokikh, N.E.; Preiss, T. Insights into the biogenesis and potential functions of exonic circular RNA. *Sci. Rep.* **2019**, *9*, 1–18. [[CrossRef](#)]
51. Salzman, J.; Chen, R.E.; Olsen, M.N.; Wang, P.L.; Brown, P.O. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genet.* **2013**, *9*. [[CrossRef](#)]
52. Nicolet, B.P.; Engels, S.; Agliandolo, F.; Van Den Akker, E.; Von Lindern, M.; Wolkers, M.C. Circular RNA expression in human hematopoietic cells is widespread and cell-type specific. *Nucleic Acids Res.* **2018**, *46*, 8168–8180. [[CrossRef](#)] [[PubMed](#)]
53. Barrett, S.P.; Salzman, J. Circular RNAs: Analysis, expression and potential functions. *Development* **2016**, *143*, 1838–1847. [[CrossRef](#)] [[PubMed](#)]

54. Okholm, T.L.H.; Nielsen, M.M.; Hamilton, M.P.; Christensen, L.L.; Vang, S.; Hedegaard, J.; Hansen, T.B.; Kjems, J.; Dyrskjot, L.; Pedersen, J.S. Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer. *NPJ Genomic Med.* **2017**, *2*. [[CrossRef](#)] [[PubMed](#)]
55. Jonkers, I.; Lis, J.T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 167–177. [[CrossRef](#)] [[PubMed](#)]
56. Liang, D.; Tatomer, D.C.; Luo, Z.; Wu, H.; Yang, L.; Chen, L.L.; Cherry, S.; Wilusz, J.E. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol. Cell* **2017**, *68*, 940–954.e3. [[CrossRef](#)] [[PubMed](#)]
57. Best, S.A.; De Souza, D.P.; Kersbergen, A.; Policheni, A.N.; Dayalan, S.; Tull, D.; Rathi, V.; Gray, D.H.; Ritchie, M.E.; McConville, M.J.; et al. Synergy between the KEAP1/NRF2 and PI3K Pathways Drives Non-Small-Cell Lung Cancer with an Altered Immune Microenvironment. *Cell Metab.* **2018**, *27*, 935–943.e4. [[CrossRef](#)] [[PubMed](#)]
58. Zhan, C.; Yan, L.; Wang, L.; Sun, Y.; Wang, X.; Lin, Z.; Zhang, Y.; Shi, Y.; Jiang, W.; Wang, Q. Identification of immunohistochemical markers for distinguishing lung adenocarcinoma from squamous cell carcinoma. *J. Thorac. Dis.* **2015**, *7*, 1398–1405. [[CrossRef](#)]
59. Bai, H.; Wang, C.; Qi, Y.; Xu, J.; Li, N.; Chen, L.; Jiang, B.; Zhu, X.; Zhang, H.; Li, X.; et al. Major vault protein suppresses lung cancer cell proliferation by inhibiting STAT3 signaling pathway. *BMC Cancer* **2019**, *19*, 1–13. [[CrossRef](#)]
60. Bian, T.; Zheng, L.; Jiang, D.; Liu, J.; Zhang, J.; Feng, J.; Zhang, Q.; Qian, L.; Qiu, H.; Liu, Y.; et al. Overexpression of fibronectin type III domain containing 3B is correlated with epithelial-mesenchymal transition and predicts poor prognosis in lung adenocarcinoma. *Exp. Ther. Med.* **2019**, *17*, 3317–3326. [[CrossRef](#)]
61. Mourad, S.; Bilodeau, M.; Roussy, M.; Laramée, L.; Boulianne, L.; Rouette, A.; Jouan, L.; Gendron, P.; Duval, M.; Teira, P.; et al. IDH1 as a cooperating mutation in aml arising in the context of shwachman-diamond syndrome. *Front. Oncol.* **2019**, *9*, 1–7. [[CrossRef](#)]
62. Lei, M.; Zheng, G.; Ning, Q.; Zheng, J.; Dong, D. Translation and functional roles of circular RNAs in human cancer. *Mol. Cancer* **2020**, *19*, 30. [[CrossRef](#)] [[PubMed](#)]
63. Yang, Y.; Wang, Z. IRES-mediated cap-independent translation, a path leading to hidden proteome. *J. Mol. Cell Biol.* **2019**, *11*, 911–919. [[CrossRef](#)] [[PubMed](#)]
64. Meyer, K.D.; Patil, D.P.; Zhou, J.; Zinoviev, A.; Skabkin, M.A.; Elemento, O.; Pestova, T.V.; Qian, S.B.; Jaffrey, S.R. 5' UTR m6A Promotes Cap-Independent Translation. *Cell* **2015**, *163*, 999–1010. [[CrossRef](#)] [[PubMed](#)]
65. Abe, N.; Matsumoto, K.; Nishihara, M.; Nakano, Y.; Shibata, A.; Maruyama, H.; Shuto, S.; Matsuda, A.; Yoshida, M.; Ito, Y.; et al. Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **2015**, *5*, 1–9. [[CrossRef](#)] [[PubMed](#)]
66. Liang, D.; Wilusz, J.E. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* **2014**, *28*, 2233–2247. [[CrossRef](#)] [[PubMed](#)]
67. Seiler, J.; Breinig, M.; Caudron-Herger, M.; Polycarpou-Schwarz, M.; Boutros, M.; Diederichs, S. The lncRNA VELUCT strongly regulates viability of lung cancer cells despite its extremely low abundance. *Nucleic Acids Res.* **2017**, *45*, 5458–5469. [[CrossRef](#)]
68. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **2013**, *14*, 1–13. [[CrossRef](#)]
69. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, *25*, 1105–1111. [[CrossRef](#)]
70. Zhang, X.O.; Dong, R.; Zhang, Y.; Zhang, J.J.L.J.; Luo, Z.; Zhang, J.J.L.J.; Chen, L.L.; Yang, L. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **2016**, *26*, 1277–1287. [[CrossRef](#)]
71. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 1–21. [[CrossRef](#)]
72. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.

73. Klingenberg, M.; Groß, M.; Goyal, A.; Polycarpou-Schwarz, M.; Miersch, T.; Ernst, A.S.; Leupold, J.; Patil, N.; Warnken, U.; Allgayer, H.; et al. The Long Noncoding RNA Cancer Susceptibility 9 and RNA Binding Protein Heterogeneous Nuclear Ribonucleoprotein L Form a Complex and Coregulate Genes Linked to AKT Signaling. *Hepatology* **2018**, *68*, 1817–1832. [[CrossRef](#)] [[PubMed](#)]
74. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).