OPEN ACCESS Freely available online

PLOS ONE

Role of Alternative Polyadenylation during Adipogenic Differentiation: An *In Silico* Approach

Lucía Spangenberg[1], Alejandro Correa[2], Bruno Dallagiovanna[2], Hugo Naya[1,3]*

1 Bioinformatics Unit, Institut Pasteur Montevideo, Montevideo, Uruguay, **2** Instituto Carlos Chagas, Fiocruz-Paraná, Curitiba, Paraná, Brazil, **3** Departamento de Producción Animal y Pasturas, Facultad de Agronomía, Universidad de la República

Abstract

Post-transcriptional regulation of stem cell differentiation is far from being completely understood. Changes in protein levels are not fully correlated with corresponding changes in mRNAs; the observed differences might be partially explained by post-transcriptional regulation mechanisms, such as alternative polyadenylation. This would involve changes in protein binding, transcript usage, miRNAs and other non-coding RNAs. In the present work we analyzed the distribution of alternative transcripts during adipogenic differentiation and the potential role of miRNAs in post-transcriptional regulation. Our *in silico* analysis suggests a modest, consistent, bias in 3′UTR lengths during differentiation enabling a fine-tuned transcript regulation via small non-coding RNAs. Including these effects in the analyses partially accounts for the observed discrepancies in relative abundance of protein and mRNA.

Citation: Spangenberg L, Correa A, Dallagiovanna B, Naya H (2013) Role of Alternative Polyadenylation during Adipogenic Differentiation: An *In Silico* Approach. PLoS ONE 8(10): e75578. doi:10.1371/journal.pone.0075578

Editor: Qiong Wu, Harbin Institute of Technology, China

Received July 2, 2013; **Accepted** August 14, 2013; **Published** October 15, 2013

Funding: This work was supported by grants from Ministério da Saúde and Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, FIOCRUZ-Pasteur Research Program and Fundação Araucária. Lucía Spangenberg received a fellowship from ANII (Agencia Nacional de Investigación e Innovación, Uruguay); Bruno Dallagiovanna was supported by CNPq, Hugo Naya by FIOCRUZ-Pasteur and Alejandro Correa received a fellowship from Fundação Araucária. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: naya@pasteur.edu.uy

Introduction

Mesenchymal stem cells (MSCs) are able to differentiate to mutiple cell types including those in bone, ligament, muscle and connective tissue [1] among others and are thus the focus of stem cell-based therapies. Tissue engineering [2], therapy for degenerative and autoimmune diseases [3,4] and cardiac tissue repair [5,6] are some of the areas of focus in adult stem cell research. Although much progress has been made, the regulatory processes controlling MSC differentiation remains poorly understood. Adipose derived human MSCs are easily isolated from pools of cells resident in vascular stroma of adipose tissue. Since adipose tissue is ubiquitous and easily accessible with minimally invasive procedures [7], it is an ideal resource for research and development of cell-based therapy. Understanding MSC commitment to differentiation to a specific cell type is essential for the successfully repair or regeneration of injured tissues. The switch from self-renewal to differentiation is regulated by many factors including cytokines, growth factors and extracellular matrix components present in a given microenvironment [8]. Nevertheless, the transcriptional and post-transcriptional regulatory processes remain not fully understood.

Gene expression analysis has provided great insights into the regulatory networks determining self-renewal and differentiation processes [9,10]. Deep sequencing techniques have also played a key role in clarifying the complex mechanisms involved. Regulation is at both the transcriptional [11] and post-transcriptional [12,13] levels. Also non-coding elements are involved [14] in the regulatory machinery [15]. In order to address post-transcriptional regulation, many groups are focusing on sequencing mRNAs associated to translating polysomes and comparing them with total RNA [12,13,16].

Expression analysis with deep sequencing methods enables the distinction of alternative transcripts of the same gene. In this context, the focus is shifted from analyzing genes as an entity (represented by a single canonical transcript) towards an alternative transcript usage model, non-coding RNAs (e.g., miRNAs), alternative splicing, 3′UTR switching, polyadenylation [17,18], etc. Alternative polyadenylation (APA) results in subpopulations of transcripts differing in 3′UTR length, which makes them more or less susceptible to the regulation by miRNAs (shorter 3′UTR might have fewer miRNA binding sites) [19,20]. A recent study has shown a role for APA in muscle stem cell development. The Pax3 protein represses differentiation in that transcripts can be targeted by mir-206. Boutet *et al.* [18] showed that different muscle tissues process Pax3 transcripts differently through APA, in which transcripts were differentially targeted by miR-206 based on 3′UTR length. In turn, different Pax3 protein levels result in functional changes in muscle stem cell behavior. Other groups assessed this type of mechanism in a global way, analyzing 3′UTRs length patterns of all genes in different scenarios. Sandberg *et al.* showed a global shortening of 3′UTRs in proliferating murine CD4+ T lymphocytes [21], and Kolle *et al* showed human embryonic stem cells to have extended 3′UTRs. The latter study also found alternative gene model usage [13]. In addition, Ji and collaborators reported that mouse genes tend to express longer 3′UTRs during the progression of embryonic development [22].

In the present work, we focus on post-transcriptional regulation during adipogenesis, specifically analyzing transcript usage differ-

ences based on 3′UTR length. We analyze data previously obtained using RNAseq [12] to study the initial phases of adipocyte differentiation of adipose-derived human mesenchymal stem cells (hASCs). Total mRNA (total) and mRNAs associated with translating ribosomes (polysomal fraction) were sequenced at two time points: 0 and 3 days after induction. We found that 3′UTRs tended to be longer after cells were induced, thereby potentially providing more miRNA binding sites. A mean difference of 18 bases in transcript length was found in induced versus control conditions. In our previous study, based on a subset of the proteomic data of Molina *et al.* [23], we found a low correlation between protein and corresponding mRNA changes. Standard linear models predicting changes in protein levels based only on mRNA changes were inaccurate. Here, we propose linear models that incorporate the effect of miRNAs on protein changes, which substantially improve the correlation between protein and mRNA change. Furthermore, our linear models indicate several miRNAs that could potentially be involved in post-transcriptional regulation of genes relevant for adipogenesis. Moreover, we also observed that genes previously described as involved in the differentiation process (Plurinet genes [24]) are enriched in longer 3′UTR in the induced condition.

Results

1 Global analysis of differential transcript usage

Previous studies have shown that the use of alternative polyadenylation sites, which generates transcripts with varying 3′UTR length (shorter or longer), are associated with cells having higher proliferation rates [21,22] (those generally having shorter 3′UTR), with cells undergoing differentiation [13] (longer 3′UTR) and with post-transcriptional regulation events in general. We determined alternative transcript usage by comparing the proportions of FPKM of each transcript for IN (induced samples, differentiating cells) vs. CT (control samples, undifferentiated cells). Analysis was done with total and polysomal fractions (see 2), however, total RNA was analyzed in greater detail to more accurately recover all alternative transcripts. Transcripts destabilized by miRNA are not expected to be associated with polysomes.

A preliminary global analysis of our data showed that the average 3′UTR length, weighted by the proportion of transcripts used for each gene, differed under IN compared with CT conditions. The mean difference was 18 bases, and 11 bases when outliers were excluded. In this context, we defined outliers as 3′UTRs with an average difference between conditions (IN–CT) longer than 1 kb. We excluded extreme values to avoid a bias in the determination of the mean (only for these calculations). Both lengths (18 and 11) are sufficient for generation of an additional miRNA binding site (see Discussion). Extension of 3′UTR regions was found in 6608 genes ($IN-CT>0$, weighted by the proportion of transcripts), whereas 5931 had shorter 3′UTRs ($IN-CT<0$). As such, we observed a tendency for longer 3′UTR under IN conditions compared with CT ($p<1*10^{-8}$, Wilcoxon test). We tested our data using the Cochran-Mantel-Haenszel (CMH) statistic, as in Fu *et al.* [25] to assess the significant of the differences observed. Since several genes have more than two transcripts and the length of the 3′UTR is a quantitative variable, the linear trend alternative to independence test [26] is more accurate than a standard χ^2 test. CMH determines a trend value for each gene, based on a Pearson correlation, with a corresponding p-value. In our setup, a positive correlation is observed if there is a tendency for longer 3′UTRs under IN conditions and a negative correlation for longer 3′UTR in CT. From the 16832 genes tested, 5952 displayed a negative trend, 6675 a positive trend and 4205 showed no trend. Tendencies are based on the calculated correlation values needed for the CMH test. Furthermore, 182 genes were significant at an FDR <0.01. Of the significant 182 genes, 114 had a positive correlation value and 68 a negative one. This difference is again significant ($p<1*10^{-3}$, Wilcoxon test). In summary, we found that there is a modest but consistent tendency to use alternative transcripts with longer 3′UTR under IN conditions compared with CT in our dataset.

Trends observed in polysomal fractions were similar to those in total RNA fractions, however, the number of genes were smaller: 5340 genes had a negative trend (length $CT>IN$), 6152, a positive trend (length IN>CT) and 4210 no trend ($p<3.6*10^{-14}$). These trend results are also based on the correlation values used for the CMH test. Differences in the distribution of gene trends for total and polysomal fractions were significant ($p<5*10^{-4}$), but were relatively small considering the large numbers compared. Of 92 significant genes at FDR <0.01, 51 had positive correlation values and 41 negative values. A number of significant genes, each having at least 20 nucleotides of 3′UTR length difference between conditions, were found in both total and polysomal fractions (positives and negatives). The overlap list of negative genes includes: ARL6IP5, COL1A2, RPL23, CD59, THBS1, TMED9, SPARC and MFAP5, and the positive list includes: DCN, BRK1, OSTC, PEBP1, BNIP3L, SAR1A and LSM6.

The observed mean difference in whole transcript lengths between conditions was 20 bases, considering all 3′UTRs, and 12 bases without outliers (defined as before). Interestingly, the correlation between trend statistic for total and polysomal fractions was very low, $r=0.06$ ($p<1.6*10^{-13}$), pointing towards important differences in post-transcriptional regulation.

2 Large fold change differences between mRNA and proteins

Large differences can be observed between mRNA and protein products in eukaryotic cells. This is due to various types of post-transcriptional regulation including tRNA and ribosome availability, regulation by small non-conding RNAs and transcripts nucleotide composition. However, in general a reasonably good agreement (in logarithmic base) is expected [27,28]. We previously correlated protein fold changes (in mouse) determined by SILAC (Molina *et al.* [23]) and our human RNAseq data [12]. We found a relatively high correlation between our RNAseq data and a subset of Molinas data, consisting of a group of secreted proteins. However, we were unable to find a high correlation with the entire dataset, which also included nuclear proteins. Using the same data set, we addressed the reasons behind the low correlations observed between mRNA and protein fold changes. In brief, our RNAseq dataset consists of two sets: RNAseq of total RNA (total) and of polysome associated RNA (polysomal). The samples were hADS cells taken at time point 0 (control; CT) and three days after adipogenesis induction (induced; IN). Molina *et al.* measured 3T3-L1 murine stem cell protein levels at different time points during adipogenesis: day 0, 1, 3, 5 and 7. Ideally, such comparisons would be more appropriate comparing experiments from the same species, however, Molina's dataset was the most suitable available for comparison with our RNAseq analysis (see Materials and Methods). To the best of our knowledge, studies on adipogenesis comparing different species have not been reported. However, embryonic stem cell pluripotency is established and maintained by a largely conserved regulatory network in eutherian mammals [29]. Other studies have shown conserved genes and pathways

involved in mammary gland development in human and mouse similarly governing cell-fate decisions and differentiation processes [30].

A linear model for $\text{logFC}_{protein}$ values (log fold change of protein, e.g. $log(\frac{day5}{day0})$) versus our logFC_{mRNA} values (log fold change values of mRNA, $log(\frac{IN}{CT})$) was fit for each time point in the experiment of Molina *et al.*, and residuals analyzed. Such differences (residuals of the corresponding linear model) were very large for several genes. Fig. 1 shows the differences in logFC for each time point (day 1, 3, 5 and 7) in the secretome dataset (nuclear in Fig. S1). Only those genes with the greatest differences are shown, and both RNA fractions are considered (A polysomal and B total). Genes clustered into two groups: negative differences ($logFC_{protein} \leq logFC_{mRNA}$) are shown at the bottom of Fig. 1 (green) and positive differences above (red). The large differences suggest post-transcriptional regulation of several genes potentially by small non-coding RNAs, especially miRNAs. Linear models were constructed taking into account alternative transcript usage between conditions and characterizing miRNA binding sites involved. We discuss these results in the next subsections.

3 Alternative transcripts and miRNAs help explain protein fold changes

We analyzed the effect of miRNAs targeting 3′UTR of alternative transcripts in the fold change of proteins by linear models. The base model included only logFC_{mRNA} (and the intercept) as predictor variable for the $\text{logFC}_{protein}$. miRNA target sites were then included in order to increase the variance explained by the model. Of the 147 secreted proteins analyzed by Molina *et al.*, 111 genes were represented in our expression dataset (total and polysomal RNA), and of the 280 nuclear proteins, 214 were found in our set. In addition, we determined the relative transcript abundance per gene in our dataset (using *cuffdiff*, see 5). Once we established the miRNAs targeting those transcripts (weighting by transcript usage) and the logFC values for each gene, we predicted the effect of each miRNA on protein level. Hereinafter, when we mention models "including/considering miRNAs", we are referring to models, which incorporate the effect of the differences in miRNA target sites. First, models including each miRNA individually were constructed (694 miRNAs in total), then all combinations of two to five miRNAs were included in models. The best models were selected based on BIC (Bayesian Information Criterion).

Table 1 shows linear model results for secreted and nuclear proteins with polysomal and total RNA fractions. The base model (the effect of logFC_{mRNA} on protein change without considering any miRNAs) is shown, as well as two single miRNA models (per comparison: polysomal/total and secreted/nuclear) and the best model by BIC (including one or two miRNAs).

The variance explained by the models increases substantially when the effect of specific miRNAs is incorporated. For example, for polysomal secreted proteins, the base model explains 15.5% of the variance, while 32.1% is explained by the two-miRNA model. The effect of miR-130b and miR-558 on the logFC_{mRNA} more accurately reflects the observed protein logFC. These miRNAs may have an important regulatory role in adipogenesis. Similar results were obtained with the remaining datasets. In addition, we also found that variances explained by polysomal fraction models (secretome and nuclear) were in general higher than those using total RNA (Table 1). This can be explained by the reduced effect on mRNA destabilization in polysomal mRNAs (they are already associated with polysomes). Finally, Table 2 shows all miRNAs that were significant at an FDR < 0.05 in single miRNA models at day 5, in the different datasets. Several of these miRNAs (underlined in the table) were previously found to be involved in adipogenesis [31]. To assess the possibility that our results were due to random sampling on the miRNA matrix, we performed a

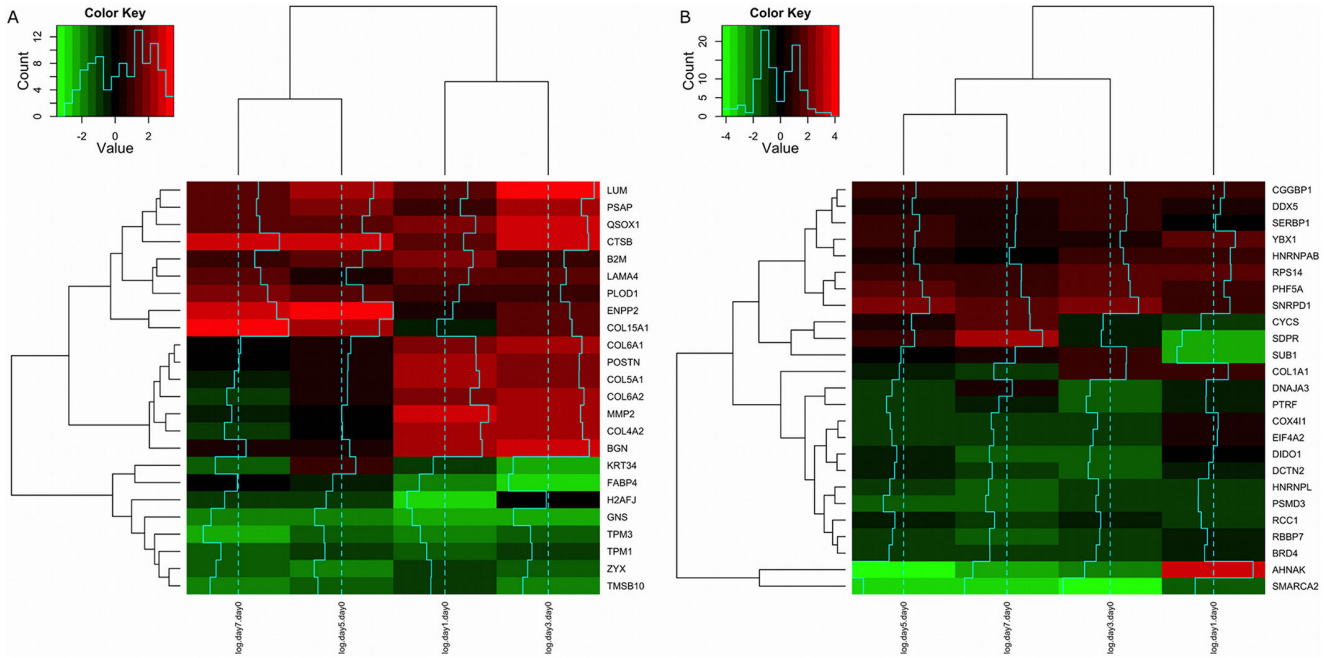

Figure 1. Heatmap of the residuals of the model $\text{logFC}_{protein} \sim \text{logFC}_{mRNA}$. Protein levels (logFC) of the set of secreted proteins are compared against the logFC of our data set and the residuals of the linear model analyzed; polysomal fraction (A) and total fraction (B). All time points are considered: day 1, 3, 5 and 7 (dendrogram on the top). Genes are on the rows (dendrogram on the left). Only data for genes with large absolute residuals are shown.
doi:10.1371/journal.pone.0075578.g001

Table 1. Linear model results for secreted and nuclear proteins at day 5.

SECRETOME				NUCLEAR			
Polysomal RNA				**Polysomal RNA**			
logFC	$microRNA_1$	$microRNA_2$	$adjusted R^2$	**logFC**	$microRNA_1$	$microRNA_2$	$adjusted R^2$
0.377†	-	-	0.155	0.0978	-	-	0.004
0.332†	miR-130a†	-	0.279	0.0952	miR-185*†	-	0.202
0.337†	miR-130b†	-	0.299	0.0831	miR-20b*†	-	0.175
0.322†	miR-130b†	miR-558†	0.321	0.105	miR-16-2*†	miR-185*†	0.300
Total RNA				**Total RNA**			
logFC	$microRNA_1$	$microRNA_2$	$adjusted R^2$	**logFC**	$microRNA_1$	$microRNA_2$	$adjusted R^2$
0.209†	-	-	0.100	0.0612	-	-	0.002
0.203†	miR-130a†	-	0.232	0.0410	miR-372†	-	0.191
0.205†	miR-19a†	-	0.228	0.0463	miR-106b†	-	0.181
0.205†	miR-150*†	-	0.239	0.0339	miR-523†	-	0.206

Results for applying linear models to the data at day 5 secreted and nuclear proteins. Both RNA fractions are considered. For each subtable (e.g. secretome-polysomal) the first row shows the results for a linear model without considering microRNA effect (the standard model: $logFC_{protein}$ vs. $logFC_{mRNA}$). The 2nd and 3rd row represent the values for univariate models, including the effect of only one miRNA. We selected the two most significant miRNAs. The last row shows the (multivariate) best model as determined by the BIC value. In several cases the best model is not multivariate, especially since BIC penalizes the number of parameters.
†means a significance level of $<1*10^{-3}$.
doi:10.1371/journal.pone.0075578.t001

bootstrap analysis as described in section 6. Our results ruled out this possibility, since for all significant miRNAs, much less than 5% of random models had explained variances comparable to BIC-selected "true" models. Fig. 2 shows an analysis indicating how many times each miRNA wins, comparing explained variances using "true" over random models (R^2 values are color coded). The miRNAs that win at least 95% of the times generate the best fitting models (more variance explained), and are shown in red. These miRNAs (red) could be distinguished from those winning in random models (>5%). Explained variances for both miRNA groups were compared and the differences were found to be statistically significant ($p<1*10^{-20}$, Kruskal-Wallis test); miRNAs winning in true models (>95% of the times) usually explain much more variance than miRNAs winning in random models (see Fig. S2).

4 Consequences of including miRNAs and alternative transcripts

While the effect of $\log FC_{mRNA}$ is significant for the secretome set (both fractions), it is not for the nuclear set (both RNA fractions), as shown in Table 1. Significant $\log FC_{mRNA}$ coefficients are higher for polysomal than for total RNA, which is expected since polysomal RNA reflects protein levels more accurately. Fig. 3 summarizes results for the best BIC models for the log-fold change in secreted proteins on day 5 with respect to day 0, for polysomal and total RNA. Fig. 3, (A) and (C) show the distribution of genes when comparing $\log FC_{mRNA}$ with $\log FC_{protein}$ not including the miRNA effect (base model). Fig. 3, (B) and (D) show the model including the effect of miR130b and miR-558 (polysomal) and miR-150* (total). While the base model performs poorly in predicting behavior of several genes (colored dots), in that they deviate from the predicted model line, our model shifts them towards a more expected position. In addition, among the shifted genes several established adipogenesis genes were found: FABP4, FABP5, LPL and ADIPOQ.

The coefficient for $\log FC_{mRNA}$ is low in the base model for both RNA fractions, ca. 0.377 (polysomal) and 0.209 (total). This coefficient decreases even more in our models. This indicates a range compression comparing protein fold-change with mRNA fold-changes (in log-log scale). This might be unexpected, however, translational efficiency (the number of protein molecules produced per mRNA molecule) may decay with the number of transcripts (see Appendix S1 (B) for more details). In fact, several studies have shown a decrease in translational efficiency [27,28,32], observed as a linear trend in the dot plot of absolute protein quantity vs mRNA quantity. Furthermore, as we show in Appendix S1 (A), the slope of this relation (1 indicating no decreasing translational efficiency with mRNA quantity, and 0 a complete decrease) is identical to the coefficient of $\log FC_{mRNA}$ in the linear models we have fit here.

Regulatory features we found help to explain protein level changes seen during adipogenesis, even though we used a limited data set. For this reason, in addition to analyzing significant miRNAs acting as predictor variables in protein-mRNA logFC relationships, we also analyzed the distribution of all miRNAs in all genes (with RNAseq data) having alternative transcripts.

5 Multiple miRNA functioning together in regulation

Evidence shows that multiple miRNAs may act together to co-regulate specific genes for normal function [33–36]. We investigated co-occurrence of miRNAs in our data set, and found established as well as novel regulatory correlations between them.

Table 2. Significant miRNAs at day 5 as obtained from the linear univariate model.

	Polysomal RNA	Total RNA
secreted	miR-103,miR-107,miR-130a,miR-130b	miR-103,miR-107,miR-130a
	miR-142-3p,miR-144,miR-148a	miR-130b,miR-142-3p,miR-144
	miR-148b,miR-150*,miR-152,miR-15a	miR-150*,miR-152,miR-15a
	miR-15b,miR-16,miR-190b,miR-195	miR-190b,miR-19a,miR-19b
	miR-19a,miR-220c,miR-28-3p,miR-29a	miR-210,miR-220c,miR-26a
	miR-29b,miR-29b-2*,miR-29c,miR-301a	miR-26b,miR-27a*,miR-28-3p
	miR-301b,miR-302a,miR-302d,miR-338-5p	miR-29a,miR-29b,miR-29b-2*
	miR-33a,miR-33a*,miR-33b,miR-340	miR-29c,miR-301a,miR-301b
	miR-486-5p,miR-509-5p,miR-510,miR-551b*	miR-338-5p,miR-33a,miR-33a*
	miR-553,miR-558,miR-569,miR-574-5p	miR-340,miR-361-5p
	miR-589*,miR-628-5p,miR-633,miR-672	miR-486-5p,miR-509-5p
	miR-768-3p,miR-768-5p,miR-891b	miR-510,miR-551b*,miR-553
		miR-558,miR-569,miR-574-5p
		miR-575,miR-582-3p,miR-587
		miR-589*,miR-604,miR-607
		miR-628-5p,miR-672
		miR-768-3p,miR-768-5p,miR-891b
nuclear	miR-143*,miR-16-2*,miR-185*,miR-20b*	miR-100,miR-106b,miR-10b*,miR-185*
	miR-346,miR-372,miR-378*,miR-587	miR-193a-5p,miR-222*,miR-28-5p
		miR-372,miR-433,miR-507
		miR-523,miR-548b-3p,miR-551b
		miR-576-5p,miR-621,miR-885-5p

Set of significant miRNAs in each data set. Underlined miRNAs correspond to those found in Zhang *et al.* (revision on miRNAs involved in adipogenesis) [31].
doi:10.1371/journal.pone.0075578.t002

In addition to the co-occurrence in the linear models described before, we now explored the correlation of miRNA occurrences in the different transcripts analyzing the presence/absence matrix of miRNAs by transcript, weighted by transcript usage differences between IN and CT inside genes (see subsection 7). Based on the total RNA fraction (reflects status of all transcripts, e.g. before degradation) we observed some miRNA pairs with significant correlations. We describe four of the cases found in our study. In all cases, we restricted our analysis to transcripts (rows in the matrix) in which at least one miRNA (of the two per comparison) is present and we compared the correlations obtained with the presence/absence matrix (1 and 0) with the values obtained with the matrix weighted by transcripts usage. First, the presence/absence matrix for miR-204 and miR-211 target sites was considered and a correlation was determined $r=0.044$ ($p>0.15$). When using the weighted matrix, we obtained a correlation value of $r=0.76$ ($p<1*10^{-15}$). Similarly, for transcripts targeted by miR-17 and miR-93, the correlation using the presence/absence matrix was $r=0.24$ ($p<1*10^{-15}$), whereas the correlation with the weighted matrix was$r=0.91$ ($p<1*10^{-15}$). For transcripts targeted by miR-17 and miR-20a a negative correlation is observed using the presence/absence matrix ($r=-0.79$, p-value$<1*10^{-15}$), however considering weighted data a significant positive correlation is observed ($r=0.57$, $p<1*10^{-15}$). Pair miR-34 and miR-449 presents a negative correlation in both cases ($r=-0.23$, $p<1*10^{-15}$ and $r=-0.79$ presence/absence matrix, $p<1*10^{-15}$ for our weighted data).

6 Alternative transcripts in relevant genes from other sources: PluriNet genes

The PluriNet is a protein-protein network with 299 members common to pluripotent stem cells based on gene expression profiles of 150 human cell samples. Such molecular network is believed to be involved in the differentiation and self-renewal of pluripotent stem cells [24].

We investigated 3′UTR length distribution of PluriNet transcripts for IN vs CT conditions. Similar trends were observed for the total and polysomal fraction. We found that positive differences correspond to longer 3′UTR under IN conditions, and negatives the converse situation (zero indicates no differences), when considering the weighted differences in length (as determined in 4). We first ranked all genes by 3′UTR length differences, and identified PluriNet genes within the ranking. As shown in Fig. 4, PluriNet genes accumulated near small negative differences but distributed evenly for all positive values. Of the 299 Plurinet genes, 216 were found in our dataset. 123 had positive differences in length (3′UTR longer in IN) and 88 negative (3′UTR longer in CT) with 5 having no differences. GO analysis of the 88 negative genes resulted in the following over-represented terms: metabolism of non-coding RNA ($p<1.1*10^{-2}$), snRNP assembly ($p<1.1*10^{-2}$), loading and methylation of Sm proteins onto SMN complexes ($p<1.1*10^{-2}$), RC complex during G2/M-phase of cell cycle ($p<1.55*10^{-2}$). In the set of positive correlated genes, one enriched term was found: nuclear part ($p<2.08*10^{-3}$).

Interestingly, according to the Cochran-Mantel-Haenszel statistic (with FDR$<$0.01) the following PluriNet genes showed

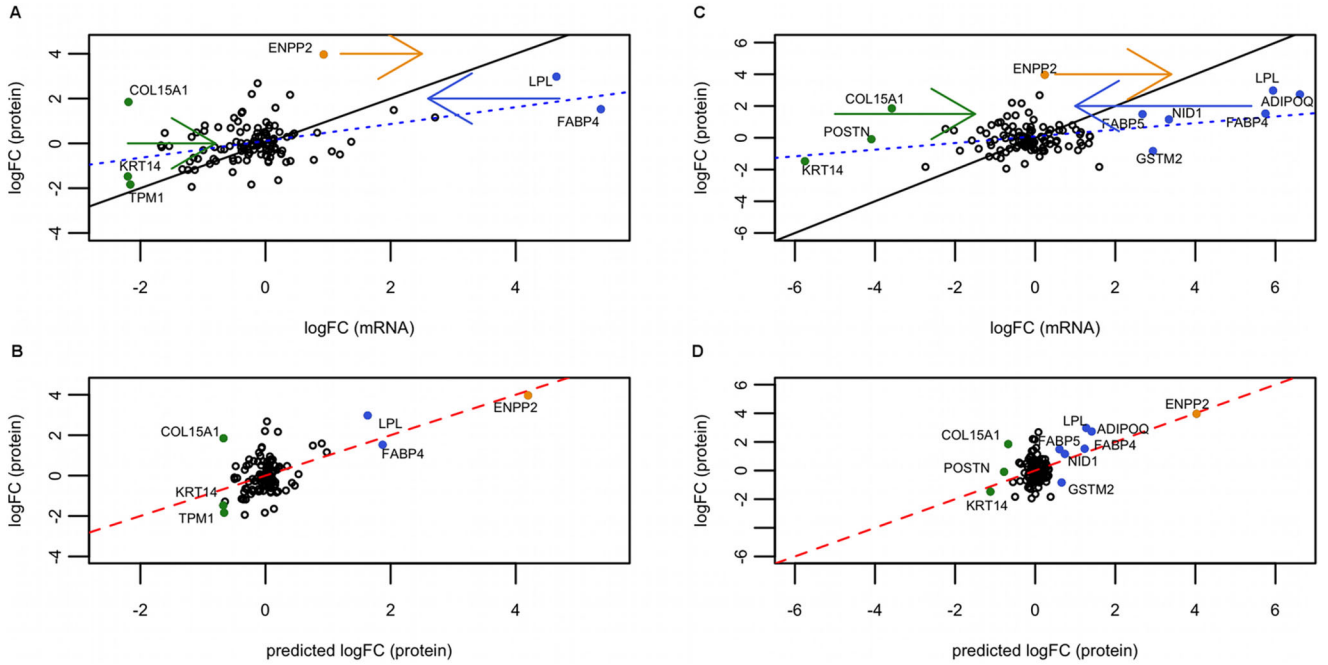

Figure 2. Bootstrap to asses our results for each RNA fraction and each protein set. Bootstrap results for total RNA fractions are shown in A (nuclear) and B (secretome). Polysomal fraction is shown in C (nuclear) and D (secretome). For each such pair of conditions, we performed a bootstrap analysis as explained in 0.6. For each miRNA we permute the values of the genes and calculate the explained variance from the resulting linear model. This procedure is repeated 1000 times. The y-axis represents how many times the "true" miRNA wins over the random model. The x-axis represents all miRNAs. The colors, from red to green, represent the explained variance from the current "true" model. It can be observed that the miRNAs win almost all times (the larger bars, almost reaching 1), explain the larger variance, and hence produce the best models (red).
doi:10.1371/journal.pone.0075578.g002

significant 3′UTR length differences between IN vs CT: PSMA3, PSMA4, PSME3, proteasome assembly (subunits and activator), HSPA8 (heat shock 70 kDa protein 8), SNRPF (small nuclear ribonucleoprotein polypeptide F), SUMO1 (small ubiquitin-like modifier which promotes SUMOylation), TMEM258 (transmembrane protein 258) and SNRPE (small nuclear ribonucleoprotein polypeptide E). Only SNRPE had a positive correlation, while the others had a negative correlation.

Discussion

We previously showed important differences in mRNAs changes comparing polysomal and total fractions during adipogenesis [12]. Furthermore, mRNA changes were poorly correlated with observed protein changes during differentiation [23]. Altogether, these results point to a very important role for post-transcriptional regulation in adipogenesis. To gain deeper insight into the mechanisms involved, we explored the differences observed in alternative transcript usage focusing on differences in the 3′UTR regions. These are relevant since they have well-known regulatory features, particularly involving small non-coding RNAs. An example showing how different miRNA binding sites can be generated in the 3′UTR of alternative transcripts is shown in Fig. 5. The gene illustrated is RER1, which is one of the significant genes in the polysome fraction in this study having alternative transcripts during adipogenesis. As indicated longer 3′UTRs may have additional miRNA binding sites.

Our results show that significant differences in transcript isoforms arise by APA during adipogenesis. A trend towards longer 3′UTR was observed in both RNA fractions, total (18/11 bases) and polysomal (20/12 bases). We proposed that this small differences in length were still sufficient for the generation of new miRNA binding sites. We tested this, by analyzing the pairwise differences between the 3′UTR length of transcripts and the corresponding differences in miRNA binding sites, for each gene. Our preliminary analysis showed that for the differences of interest (20, 18, 12 and 11 bases), out of the 16937 genes analyzed, 1235, 1204, 1132 and 1112 genes, respectively, differed in at least one miRNAs binding site.

The difference in the total RNA fraction is also consistent with the number of genes displaying a positive trend (3′UTR length IN>CT), which is significantly higher than those showing a negative trend. Regarding trend-length differences comparing IN and CT conditions, 182 genes showed statistically significant trends (FDR <0.01): 114 had a positive correlation value and 68 a negative value. Very similar trends were also observed for correlation values in the polysomal fraction. Two adipogenesis relevant genes, FABP4 and WNT2, appeared to exhibit APA and differential 3′UTR length during differentiation in our previous study [12] by visually inspection. Here we confirmed these results by analytical methods. In our earlier work, the FABP4 gene exhibited a much longer 3′UTR under IN compared with CT conditions. The WNT2 gene in contrast showed the opposite behavior having a longer 3′UTR under CT conditions. Results obtained in this study showed a (positive) difference of 103 bases and a significant correlation value of ~ 0.10 for the FABP4 gene, and for WNT2, a (negative) difference of -407 bases and a significance correlation value of ~ -0.48.

A protein-protein network was previously described for pluripotent stem cells (Plurinet) [24]. Construction of the network was based on gene expression profiles for 299 human proteins. We analyzed the distribution of differences in 3′UTR length for Plurinet genes having expression values in our dataset (216 in 299). As shown in Fig. 4, the distribution of length differences

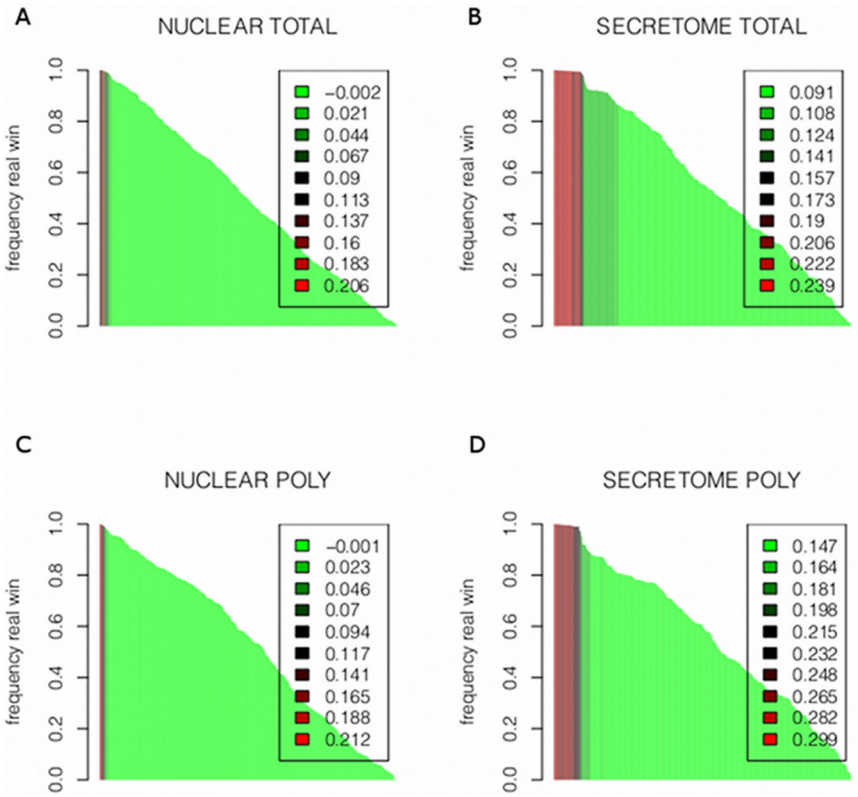

Figure 3. Linear models for day 5 secreted proteins represented graphically. (A, B) Polysomal fraction, (C, D) total RNA. (A) and (C): plot representing $\log FC_{mRNA}$ against $\log FC_{protein}$. The dashed blue line is the best fitting line of the base model, $logFC_{protein}$ against $logFC_{mRNA}$. The straight black line is the identity line (so you get an idea of the real coefficient of the model). The colored full dots are genes, which are moved after applying the model with miRNAs. Hence, they represent genes that are better explained by our model. The arrows indicate the direction of the movement. (B) and (D): plot representing our linear model including miRNA effect. In this case, the best (multivariate) model is shown: miR-130b and miR-558 (polysomal) and miR-150* (total). Full dots are the genes that were corrected by our model, being now closer to the protein prediction line of the model (red full line). Black identity line concurs with the red line. Note that the abscissas of (A) and (C) seem to have a compression of range with respect to the plots below, (B) and (D). This is not a compression, since they are different x-axis: (A) and (C) hold $\log FC_{mRNA}$ values, while (B) and (D) $\log FC_{protein}$.
doi:10.1371/journal.pone.0075578.g003

substantially deviates from the behavior of all genes. In particular, genes with much longer 3′UTRs in control cells compared with induced cells were under represented. Additionally, we found an enrichment of the term "metabolism of non-coding RNA" among genes with 3′UTR length CT>IN, which could be associated with post-transcriptional regulation.

The dataset of Molina et al., was analyzed to understand the potential role for APA in protein changes [23]. Even though the cell line used by these authors was murine, this dataset was the most suitable available to compare with our RNAseq experiment. Several studies indicated a reasonable conservation in regulatory networks between human and mouse [29,30]. Comparing differences between $\log FC_{protein}$ and the predicted protein quantity according to the $\log FC_{mRNA}$ ($\log FC_{predicted}$), some large residuals (gene differences) were observed using this dataset (Fig. 1). Adipogenic relevant genes FABP4, GNS, TPM1, TPM3, KRT34, TMSB10 and ZYX were among genes with larger negative differences, i.e, $\log FC_{protein} < logFC_{predicted}$. On the other hand, residuals with positive differences ($\log FC_{protein} > logFC_{predicted}$), include LUM, PSAP, QSOX1, COL15A1, POSTN, ENPP2 and LPL (total RNA fraction). In addition, we have found that the observed differences (residuals) do not correlate significantly with the absolute magnitude of change in mRNA. As such the differences can't be explained by the expected compression of range (see section Appendix S1 (A)).

Clear differences were observed in APA isoform usage comparing IN and CT conditions, as well as differences between predicted fold change (by mRNA) and observed protein fold change for some genes. To further investigate this discrepancy we compared explained variances of base models just including $\log FC_{mRNA}$ as predictive variable, against different models that incorporate miRNAs target site differences between transcripts as co-variables. The rationale behind including these miRNAs is to account for their potential effect on destabilizing or inhibiting translation resulting in discordance between the observed proteins and the mRNA levels. We have shown that hMSCs use their transcripts differentially during adipogenesis. We were able to test whether presence of miRNA binding sites is associated with change in the fate of specific transcripts by incorporating preferences for alternative transcripts (with alternative 3′UTR length) in our analyses. As summarized in Table 1, differences in explained variance were striking (even after adjusting for model complexity) when the effects of different miRNAs were introduced in the models. As expected, polysomal $\log FC_{mRNA}$ was higher correlated with $\log FC_{protein}$ than the corresponding correlation in total RNA. This can be seen in the explained variances of both datasets, i.e., secreted and nuclear proteins. More surprising,

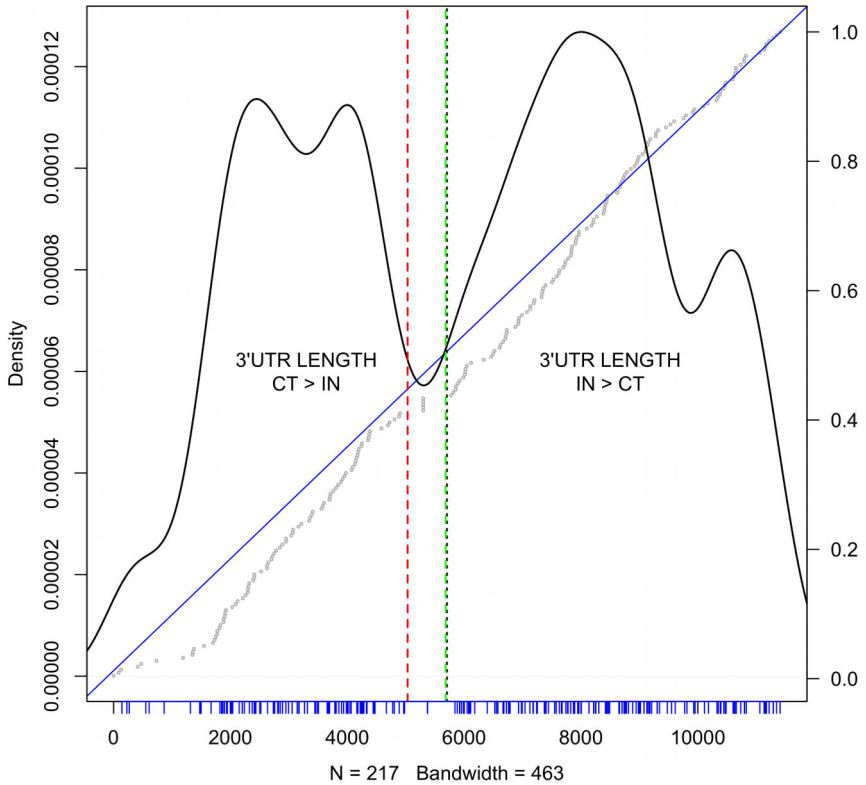

Figure 4. 3′UTR differences for PluriNet genes. On the x-axis one observes the ranking of 3′UTR lengths as determined in section 1 of all genes used for logFC calculations in the total RNA fraction. The ranking of genes belonging to the PluriNet are shown as densities (y-axis on the left). Negative lengths (CT>IN) lie to the left of the red dashed line. Positive values are to the right of the green dashed line. The wide space between those lines correspond to genes with no differences in 3′UTR length. The median of the rankings is represented as a doted black line. Tick marks in blue represent the ranking positions of the PluriNet genes. On top of the density plot the cumulative distribution of rankings is shown. The straight blue line has slope 1 and intersect 0. Gray dots represent the cumulative ranking of the PluriNet genes. The y-axis to the right indicates the meassure of this cumulative ranking. An under-representation of PluriNet genes with high negative values and a slight over-representation of positive values is observed. Moreover, only marginal PluriNet genes are presenting values of 0.
doi:10.1371/journal.pone.0075578.g004

however, is that changes in nuclear proteins were very poorly correlated with changes in mRNAs (the coefficient for $\log FC_{mRNA}$ was never significant, even in absence of other co-variables). While several reasons might account for this, mechanisms involving protein translocation could be collaborating to this lack of correlation.

A range compression of $\log FC_{protein}$ compared with $\log FC_{mRNA}$ can be seen in the slope of Fig. 3 (A and C) and in coefficients for $\log FC_{mRNA}$ in Table 1. If translational efficiency decreases with increased mRNA levels (competition for scarce resources, e.g., ribosomes) in such a way that a linear trend is observed in log-log scale when plotting amounts of protein vs mRNA, the observed range compression would be expected (see section Appendix S1 (B)). In fact, this trend was observed in several studies [27,27,32] and a coefficient of ~ 0.50 for *Saccharomyces cerevisiae* was determined [32]. We calculated a coefficient of ~ 0.35 for comparisons with the secretome dataset, a reasonable estimate. We may be underestimating this coefficient since our comparisons and analyses are between species (mouse and human). Moreover, as we are only considering up to 214 genes, our coefficient may not correspond to a global scenario in the cell. Finally, even though a significant improvement in explained variances is found by incorporating miRNAs in models, the small changes in $\log FC_{mRNA}$ coefficients indicate that the improvement in performance is basically obtained by adjusting the prediction of "poorly-behaved" genes. In addition, the linear models presented here also reveal several genes whose regulation might be explained by specific miRNAs included in the models. In particular, we observed that the following genes were better fit by miRNA-models than the base model: ENPP2, LPL, FABP4, KRT14, TPM1, COL15A1 (polysomal RNA) and ENPP2, LPL, ADIPOQ, FABP5, FABP4, NID1, GSTM2, COL15A1, POSTN, KRT14 (total RNA). In the case of polysomal RNA, miR-130b and miR-558 were the miRNAs included in the model, whereas miR-150* was the co-variable in the model considering total RNA. It is worth mentioning, that we are only considering presence of miRNA binding sites, the expression levels of the miRNAs themselves is not included in our work.

Table 2 lists all significant miRNAs for which one-miRNA models were constructed, and also indicates which are previously mentioned as relevant for adipogenesis according to the revision of Zhang et al. [31]. In particular, we found 8 significant miRNAs of the 23 previously identified. Additionally, we found several miRNAs involved in other differentiation processes not described by Zhang et al. These include miR-142-3p, miR-16 and miR-15a which are associated with (TPA)-induced differentiation of human leukemia cells (HL-60) to monocyte/macrophage-like cells [37]. Also, miR-144 was implicated in erythroid differentiation [38] and miR-148a, miR-26, miR-378, miR-486 and miR-29 were identified in skeletal myogenic differentiation [39], and miR-10

	RER1-			
miRNA	004	001	201	002
218			1	1
96*	1			1
197	1			
92b	1			
212	1			
220	1	1		
323-5p	1			
126	1		1	1
887	1			1
502-5p		1	1	1
616*		1	1	1
30d*		1		
30a*		1	1	1
623		1	1	
22*		1		
220b		1		
188-5p		1		
623		1		1
424*		1		
582-5p		1		
651		1	1	1
548b-3p		1		
548d-5p		1		
548b-5p		1		
548c-5p		1		
616*		1		
938		1		1
424*				1

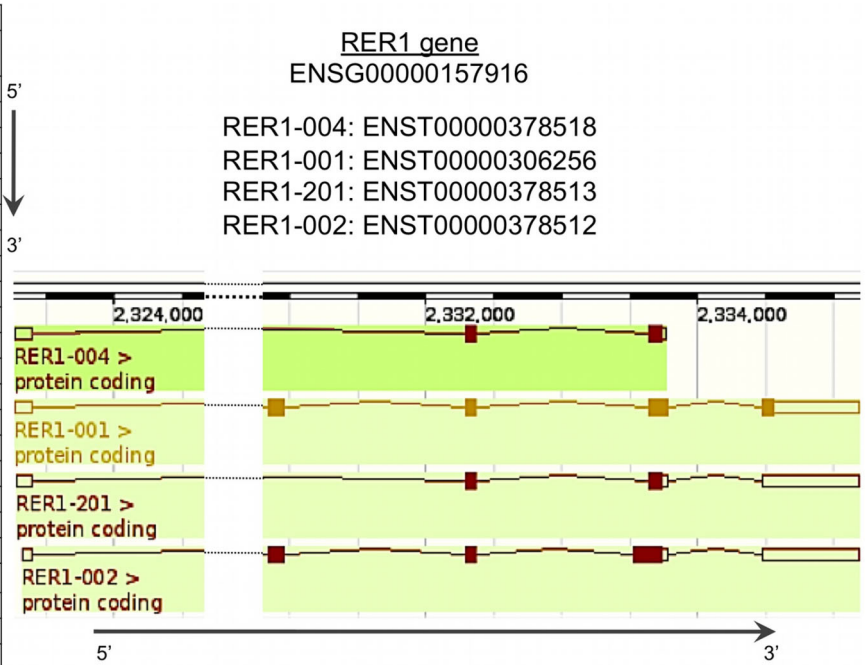

Figure 5. An example of how different microRNAs binding sites arise from alternative transcripts. The table shows the presence of the miRNAs in the transcripts. The longer the 3′UTR the more binding sites are seen.
doi:10.1371/journal.pone.0075578.g005

was involved in endodermal differentiation [40]. Hence, miRNAs identified using our *in silico* analysis were previously found to be involved in several differentiation processes (including adipogenesis) by experimental methods.

Co-occurrence of miRNAs is not unusual; several miRNAs have been found to work together in gene regulation. Based on differences observed in alternative transcript usage, we explored miRNA co-occurrence in adipogenesis. We have found several strong associations in our presence/absence matrix weighted by differences in transcripts usage. Here we discuss some examples. Our primary analysis shows a statistically significant, but relatively trivial (since they are homologous) co-ocurrence of miR-204 and miR-211, whose common target is the Runx2 gene. miR-204/211 inhibits expression of Runx2, which inhibits osteogenesis and promotes adipogenesis of mesenchymal progenitor cells and bone marrow stromal cells [33]. We also observed a highly significant association of miRNA pair miR-17 and miR-93. They belong to the family including miR-17-5p, miR-20a, miR-93, and miR-106a, are differentially expressed in developing mouse embryos and have a controlling function in stem cell differentiation [41]. They are also key regulators of induced pluripotent stem cells and play a role in reprogramming efficiency of such cells [34]. On the other hand, miR-34 and miR-449 are negatively correlated in our data set implying that the presence of one results in the absence of the other. Both miRNAs belong to the same family; miR-449a, b and c are strong inducers of cell death, cell cycle arrest and cell differentiation; miR-34 is activated with expression of p53 protein and miR-449 is induced by E2F1, a cell cycle regulatory transcription factor. They are responsible for an asymmetric feedback loop that keeps the balance between E2F and p53 functions. miR-449 helps to ensure normal cell function but is also involved in maintaining a close interaction between cell differentiation and tumor suppression [35].

In summary, in the present work we found interesting and consistent differences in transcript isoforms used during adipogenesis. We found that, in general, induced cells had longer 3′UTRs compared with undifferentiated hMSCs. Furthermore, we characterized these differences by identifying genes whose transcripts had important differences in miRNAs target sites. Additionally, we demonstrated that by incorporating the effect of several miRNAs and alternative transcript usage in linear models, we were able to substantially improve prediction of $\log FC_{protein}$ over the base model that only includes $\log FC_{mRNA}$. We need to expand our dataset by obtaining more accurate proteomic data to further corroborate our findings. Our results indicate that post-transcriptional regulation plays a key role in differentiation.

Materials and Methods

1 Ethics statement

Samples were isolated and collected after obtention of written informed consent, agreeing with guidelines for research involving human subjects, and with the approval of the Ethics Committee of Fundação Oswaldo Cruz, Brazil (approval number 419/07), as previously mentioned in [12].

2 Sample description

We used samples described by Spangenberg *et al.* [12]. Raw data is available under the accession number E-MTAB-1366 in the ArrayExpress repository. Stem cells were obtained from

Table 3. Mapping statistics of RNA-seq.

donor	condition	raw data	reads for mapping	mapped	unmapped	junctions	%
61	CT_poly	15105571	15041140	8026275	8462131	38127	53.6
61	IN_poly	18367050	18280311	10057455	10359395	32762	55.2
67	CT_poly	40032577	39862820	19398037	23642317	40995	48.8
67	IN_poly	148700586	147993700	55932659	103696209	37807	37.8
67	CT_total	8883206	8845973	4436690	5185133	39802	50.6
70	CT_poly	17473812	17403946	9415117	9862766	39336	54.3
70	IN_poly	32280923	32151368	16831536	19327229	32614	52.5
70	CT_total	121079661	120741759	58016204	71585612	39275	48.1
61	IN_total	31667090	31573343	16498438	18437894	57296	52.4
67	IN_total	27685080	27615016	13794780	16630549	53312	50.1
70	IN_total	60059063	59886179	31756854	34628161	47819	53.1
61	CT_total	22644356	22584805	11745550	12531817	54097	52.2
67	CT_total	34358013	34267292	19408351	19849852	32832	56.7

Mapping data of SOLiD runs. Following data is shown: donor number, condition considered (CT or IN, and polysomal or total RNA), number of raw reads obtained from the sequencing process, number of reads considered for mapping, number of mapped reads, unmapped reads, and the percentage of mapped reads.
doi:10.1371/journal.pone.0075578.t003

adipose tissue of three obese human donors. hASCs were isolated, cultured and characterized as previously described [42]. Briefly, adipogenesis was induced with 6 day-cycles of induction/maintenance over 21 days. Induction medium contained the adipogenic inducers insulin, dexamethasone, indomethacin and IBMX; maintenance medium contained insulin. Medium was changed every 3 days. The degree of adipogenic differentiation was determined by assessing cytoplasmic accumulation of triglycerides by staining with Oil Red O or Nile Red (Sigma-Aldrich). Samples were taken at time point 0 (control samples, CT) and then after three days (induced samples, IN).

A total of 13 samples were sequenced with SOLiD4 System (Applied Biosystems), 7 CT (2 polysomal-associated RNA and 6 total RNA samples) and 6 IN (3 polysomal-associated RNA and 3 total RNA). Table 3 shows an overview of samples. The proteomic data used in this study is from Molina *et al.* [23]. They quantified two sets of 3T3-L1 murine proteins with SILAC: 280 nuclear and 147 secreted proteins, with a total of 427 proteins. These were analyzed during adipogenesis (at day 0, 1, 3, 5 and 7).

While our RNA-seq data is from human donors, nevertheless we decided to compare it against murine proteomic data. Of course, this assumes a high conservation at protein level between this two organisms in the involved networks, a fact relatively supported by recent studies [29,30]. Furthermore, at transcriptional level, some studies have shown that a conservation is also seen for several genes [43].

3 Primary analysis of SOLiD RNA-seq samples

Table 3 summarizes results of the mapping procedure with *tophat2* and *cuffdiff*. We obtained a median of 52% mapped reads in the 13 samples. Information on transcript usage for 62134 ensembl gene ids was obtained from *cuffdiff* for total and polysomal RNA samples. These were filtered according to the quality status of transcripts, because the low number of reads might compromise determination of FPKM. After filtering we obtained 61381 for both sets, polysomal and total RNA. From those genes, 21647 have annotated 3′UTRs according to ensembl annotation, corresponding to 74803 transcripts.

4 Summarizing transcript differences

We calculated the relative frequency of each transcript for each condition (IN and CT), and weighted the transcript 3′UTR length by the differences in frequency (we did this for each gene). To assess the significance of the differences observed above, we tested our data using the Cochran-Mantel-Haenszel statistic, a test of linear trend alternative to independence [26], which is more sensitive than a standard χ^2 test if a linear trend holds. Additionally, for each gene we calculated and analyzed the Pearson-r distribution between 3′UTR length and condition (CT = 0, IN = 1) [26].

5 Mapping and annotation

13 samples were mapped onto the reference genome (hg19 GR37p2) using *tophat2* [44]. *cufflinks* [45] v2.1.1 was then used for transcript assembly. Determination of isoform abundance was done with *cuffdiff* v2.1.1. The annotation file used for counting was based on the genome version Hg19 Gr37p10 (August 2012), downloaded from the ensembl. The 3′UTR annotation file was also created from the ensembl (version Hg19Gr37p10, 15 August 2012) human gff annotation file. The miRNA target information considered is the one included in the R package microRNA, from Gentleman and Falcon [46], which is also based on ensembl. Currently, it contains a total of 694 miRNAs targeting a total of 34507 transcripts.

Mapping, gene expression assessment and differential expression determination in our earlier work was performed using the *Rsubread* and *edgeR* R packages.

6 Linear model for correlation of microRNAs with protein levels

We developed a linear model approach to show the influence of miRNAs targeting 3′UTR regions of transcripts on respective protein expression levels.

Our starting point is data generated from *cuffdiff* software. An abundance normalized measure, FPKM, is first obtained for each transcript isoform which represents the number of fragments per kilobase per million fragments falling on each feature (e.g.,

Transcripts	IN_{prop}	CT_{prop}	$Prop_{IN-CT}$	$miRNA_1$	$miRNA_2$	...	$miRNA_{694}$
$isoform_A$	0.3	0.6	−0.3	1	0	0	0
$isoform_B$	0.4	0.1	0.3	1	1	0	1
$isoform_C$	0.1	0.2	−0.1	0	1	1	0
$isoform_D$	0.2	0.1	0.1	1	0	0	0
$gene_X$	1	1	0	0.1	0.2	−0.1	0.3
$isoform_A$	0.6	0.2	0.4	0	0	1	1
$isoform_B$	0.05	0.1	−0.05	1	0	1	0
$isoform_C$	0.35	0.7	−0.35	1	0	0	1
$gene_Y$	1	1	0	−0.4	0	0.35	0.05
⋮	...	...	...	...	...	...	...

↓

Gene	$logFC_{pro}(3)$	$logFC_{pro}(5)$	$logFC_{pro}(7)$	$logFC_{mRNA}$	$miRNA_1$	$miRNA_2$	...
$gene_X$	2.1	3.2	2.1	1.9	0.1	0.2	...
$gene_Y$	−1.2	−1.4	2.9	2.4	−0.4	0	...
$gene_Z$	−0.9	2.4	3.9	5.4	0.6	−0.1	...
⋮	...	...	...	...	...	...	...

Figure 6. Representative table for constructing the model. For each gene we determined the proportion of FPKM in each sample and calculated the differences ($Prop_{IN}-Prop_CT$). Furthermore, we determined the miRNAs targeting transcripts (inside 3′UTRs). A total of 694 were considered. The isoform has a 1 in $miRNA_1$ if that miRNA is present in that transcript, a 0 otherwise. For each $miRNA$ (eg. $miRNA_1$) corresponding to one gene (e.g. $gene_X$), the $Prop_{IN-CT}$ vector is multiplied by the presence/absence vector of $miRNA_1$ (with assigned 1 s and 0 s). The intermediate result is, thus, a vector having the respective $Prop_{IN-CT}$ value if $miRNA_i$ was present in the isoform and 0 otherwise ($\vec{v}=\{-0.3,0.3,0,0.1\}$). The resulting vector $\vec{v}$ is summed giving a total value for $miRNA_1$ for $gene_X$ ($sum(\vec{v})=0.1$). This represents the mean weighted usage of the miRNA in that specific gene. Larger positive values indicate that the miRNA is used more (appears more often) in IN than in CT. Larger negative values represent a higher usage in CT (values around 0 indicate same usage in both). The same procedure is done for each miRNA (so a vector of 694 values is assigned to $gene_X$) and for each gene. The gene wise table below in addition to showing the resulting values calculated above, also shows the other data needed for the model; the $logFC_{protein}$ values (at day 3, 5 and 7, from Molina *et al.*) and the respective $logFC_{mRNA}$ values (our data).
doi:10.1371/journal.pone.0075578.g006

transcript). A FPKM value is calculated for each condition and each transcript, which allows determination of differential isoform usage. The proportion of each transcript isoform for each gene was determined under all conditions based on the FPKM values. Proportions in control samples are subtracted from the proportions in induced samples (IN) to determine the differences in isoform usage. Differences in proportions of each isoform for each gene ($Prop_{IN-CT}$) and the presence of miRNA binding sites in transcript 3′UTRs (represented as 1 s in Fig. 6) were determined. The $Prop_{IN-CT}$ value is multiplied by the corresponding miRNA binding site present and the resulting vector is summed for a given gene (Fig. 6). This results in one value for each miRNA binding site for each gene, which represents a weighted mean for usage of that miRNA for that gene. Large positive values (closer to 1) are miRNAs highly used in IN samples, large negative values (closer to −1) are those most used in CT. In other words, values closer to 1 correspond to miRNAs targeting transcripts preferentially used in IN samples, and those with values closer to −1 are preferentially used in CT. Note that a given miRNA might have several binding sites in a given 3′UTR, nevertheless we considered one or more sites as either present or absent with no multiplicity value assigned. This is still a matter open for discussion, since several studies have shown cooperative effects in the past [47–50], while others suggested the opposite behavior in large and comprehensive human and mouse datasets [18,51]. We have also run our analysis considering the cooperative effect, obtaining conceptually similar results (data not shown). However, for simplicity reasons, we decided to consider the simplest model accepted and used the present/absent values. Since such values are determined for each gene and for each miRNA, results can be presented in a table with #of genes × #of microRNAs. For each day d (1, 3, 5 and 7), miRNA i and assuming $e \sim N(0,1)$, we applied following model:

$$logFC_{prot_d} = logFC_{mRNA} + microRNA_i + e_{d,i},$$

so we can determine the effect of each microRNA on protein level.

The possibility that significant miRNAs coefficients arise by chance was assessed by bootstrap analysis. We randomly assigned the existing values to genes for each miRNA, and calculated the explained variance from the linear model. We repeated this procedure 1000 times. The proportion of times the variance explained by the random model was larger than the "true" model was determined for each miRNA for the four datasets (nuclear, secreted vs total, polysomal). We arbitrarily set a threshold of 5% (times the random wins over the "true") for each dataset and compared the explained variances of the two groups (random vs. "true") using the Kruskal-Wallis test.

7 Determining significative correlation for co-occurring microRNAs

Co-occurrence of miRNAs was investigated to demonstrate regulatory effects. We analyzed the complete presence/absence table of miRNAs in human (downloaded from the *microRNA* R package). This table contains all transcripts analyzed (34507) in which 1 is assigned if $microRNA_i$ is present in that transcript, and a 0 if not, for all miRNAs considered (694). We compared pairwise correlations for all miRNAs based on that information and the same in our weighted data set. This means, we also determined the correlation of miRNAs, but weighted by proportion of the transcripts used. If a transcript with a given miRNA is used

only 40% of the time by the gene, the miRNA value assigned would be 0.4, and not a simple 1.

Not all entries were used for each pairwise correlation; we eliminate all entries in which both miRNAs had values of 0, i.e., pairwise-zero entries. Several of such entries exists, since not every transcript has either one of the miRNAs considered (in most cases, they have neither). With such strategy we have compared the correlations found by the presence/absence table, and the ones obtained by our weighted filtered data.

Supporting Information

Figure S1 Heatmap of the residuals of the model $\mathbf{logFC}_{protein} \sim \mathbf{logFC}_{mRNA}$ of nuclear proteins. Protein levels (logFC) of the set of nuclear proteins are compared against the logFC of our data set and the residuals of the linear model analyzed; polysomal fraction (A) and total fraction (B). All time points are considered: day 1, 3, 5 and 7 (dendrogram on the top). Genes are on the rows (dendrogram on the left). Only data for genes with large absolute residuals are shown.
(TIFF)

Figure S2 Box plot to show the distribution of random and ''true'' models in the bootstrap. All comparisons are shown (polysomal-secreted, polysomal-nuclear, total-secreted, total-nuclear). For each such dataset, bootstrap was performed, and two groups were determined. Low-Random group holds models in which ''true'' miRNAs data won over random sampling of the miRNA values at least 95% of the time. The High-Random group corresponds to miRNAs in which random sampling of miRNA values produce models that are better than the ''true'' more than 5% of the time.
(TIFF)

Appendix S1 (A) Range compression is observed in protein log fold-change (in our data), when $\log FC_{mRNA}$ is considered as predictor. The size of this effect is the translational efficiency (in log-log scale) as a function of the quantity of mRNA. (B) Messenger exponential decay with alternative target miRNA sites. We show that the basic assumption underlying the way in which we modeled the effect of miRNAs is an exponential decay of mRNA as a function of differential target sites.
(PDF)

Acknowledgments

We are indebted to Tamara Fernandez for helpful discussions on the manuscript. We are also grateful to Paul Gill for comments on the manuscript and correcting the language.

Author Contributions

Conceived and designed the experiments: LS AC BD HN. Analyzed the data: LS HN. Contributed reagents/materials/analysis tools: LS AC BD HN. Wrote the paper: LS HN.

References

1. Pittenger MF (1999) Multilineage potential of adult human mesenchymal stem cells. Science 284: 143–147.
2. Rosenbaum AJ, Grande DA, Dines JS (2008) The use of mesenchymal stem cells in tissue engineering: A global assessment. Organogenesis 4: 23–27.
3. Tae SK, Lee SH, Park JS, Im GI (2006) Mesenchymal stem cells for tissue engineering and regenerative medicine. Journal of Cellular Physiology 1: 341–347.
4. Uccelli A, Mancardi G, Chiesa S (2008) Is there a role for mesenchymal stem cells in autoimmune diseases? Autoimmunity 41: 592–595.
5. Boyle AJ, McNiece IK, Hare JM (2010) Mesenchymal stem cell therapy for cardiac repair. Methods In Molecular Biology 660: 65–84.
6. Jain M, Pfister O, Hajjar RJ, Liao R (2005) Mesenchymal stem cells in the infarcted heart. Coronary Artery Disease 16: 93–97.
7. Baer PC, Geiger H (2012) Adipose-derived mesenchymal stromal/stem cells: tissue localization, characterization, and heterogeneity. Stem cells international 2012: 812693.
8. Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M (2005) Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. Science 308: 1472–1477.
9. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, et al. (2002) A stem cell molecular signature. Science 298: 601–604.
10. Song L, Webb NE, Song Y, Tuan RS (2006) Identification and functional analysis of candidate genes regulating mesenchymal stem cell self-renewal and multipotency. Stem Cells 24: 1707–1718.
11. Jääger K, Islam S, Zajac P, Linnarsson S, Neuman T (2012) RNAseq analysis reveals different dynamics of differentiation of human dermis- and adipose-derived stromal stem cells. PLoS ONE 7: e38833.
12. Spangenberg L, Shigunov P, Abuda APR, Cofré AR, Stimamiglio MA, et al. (2013) Polysome profiling shows extensive posttranscriptional regulation during human adipocyte stem cells differentiation into adipocytes. Stem Cell Research 1: 341–347.
13. Kolle G, Shepherd JL, Gardiner B, Kassahn KS, Cloonan N, et al. (2011) Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. Genome Research 21: 2014–25.
14. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Research 18: 610–621.
15. Koh W, Sheng C, Tan B, Lee Q, Kuznetsov V, et al. (2010) Analysis of deep sequencing microRNA expression profile from human embryonic stem cells derived mesenchymal stem cells reveals possible role of let-7 microRNA family in downstream targeting of hepatic nuclear factor 4 alpha. BMC Genomics 11: S6.
16. Fromm-Dornieden C, Von Der Heyde S, Lytovchenko O, Salinas-Riester G, Brenig B, et al. (2012) Novel polysome messages and changes in translational activity appear after induction of adipogenesis in 3T3-L1 cells. BMC Molecular Biology 13: 9.
17. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563.
18. Boutet SC, Cheung TH, Quach NL, Liu L, Prescott SL, et al. (2012) Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. Cell Stem Cell 10: 327–336.
19. Liaw HH, Lin CC, Juan HF, Huang HC (2013) Differential microRNA regulation correlates with alternative polyadenylation pattern between breast cancer and normal cells. PLoS One 8: e56958.
20. Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Molecular Cell 43: 853–866.
21. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320: 1643–1647.
22. Ji Z, Lee JY, Pan Z, Jiang B, Tian B (2009) Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proceedings of the National Academy of Sciences of the United States of America 106: 7028–7033.
23. Molina H, Yang Y, Ruch T, Kim JW, Mortensen P, et al. (2009) Temporal profiling of the adipocyte proteome during differentiation using a 5-plex silac based strategy. J Proteome Res 8: 48–58.
24. Müller FJ, Laurent LC, Kostka D, Ulitsky I, Williams R, et al. (2008) Regulatory networks define phenotypic classes of human stem cell lines. Nature 455: 401–405.
25. Fu Y, Sun Y, Li Y, Li J, Rao X, et al. (2011) Differential genome-wide profiling of tandem 3′UTRs among human breast cancer and normal cells by high-throughput sequencing. Genome Research 21: 741–747.
26. Agresti A (2007) An Introduction to Categorical Data Analysis. John Wiley and Sons, 400 pp.
27. Stevens SG, Brown CM (2013) *In silico* estimation of translation efficiency in human cell lines: potential evidence for widespread translational control. PLoS One 8: e57625.
28. Tuller T, Kupiec M, Ruppin E (2007) Determinants of protein abundance and translation efficiency in S. *cerevisiae*. PLoS Computational Biology 3: 10.
29. Fernandez-Tresguerres B, Caon S, Rayon T, Pernaute B, Crespo M, et al. (2010) Evolution of the mammalian embryonic pluripotency gene regulatory network. Proceedings of the National Academy of Sciences of the United States of America 107: 19955–19960.
30. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, et al. (2010) Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast cancer research BCR 12: R21.
31. Zhang R, Wang D, Xia Z, Chen C, Cheng P, et al. (2013) The role of microRNAs in adipocyte differentiation. Frontiers of medicine 7: 223–230.
32. Futch B, Latter GI, Monardo P, McLaughlin CS, Garrels JI (1999) A sampling of the yeast proteome. Molecular and celular biology 19: 7357–7368.

33. Huang J, Zhao L, Xing L, Chen D (2010) MicroRNA-204 regulates Runx2 protein expression and mesenchymal progenitor cell differentiation. Stem cells Dayton Ohio 28: 357–64.
34. Li Z, Yang CS, Nakashima K, Rana TM (2011) Small RNA-mediated regulation of iPS cell generation. the The European Molecular Biology Organization Journal 30: 823–834.
35. Liz M, Klimke A, Dobbelstein M (2011) MicroRNA-449 in cell fate determination. Cell Cycle 10: 2874–2882.
36. Meenhuis A, Van Veelen PA, De Looper H, Van Boxtel N, Van Den Berge IJ, et al. (2011) MiR- 17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. Blood 118: 916–925.
37. Kasashima K, Nakamura Y, Kozu T (2004) Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells. Biochemical and Biophysical Research Communications 322: 403–410.
38. Fu YF, Du TT, Dong M, Zhu KY, Jing CB, et al. (2009) MiR-144 selectively regulates embryonic alpha-hemoglobin synthesis during primitive erythropoiesis. Blood 113: 1340–1349.
39. Zhang J, Ying ZZ, Tang ZL, Long LQ, Li K (2012) MicroRNA-148a promotes myogenic differentiation by targeting the ROCK1 gene. The Journal of Biological Chemistry.
40. Tzur G, Levy A, Meiri E, Barad O, Spector Y, et al. (2008) MicroRNA expression patterns and function in endodermal differentiation of human embryonic stem cells. PLoS ONE 3: 14.
41. Foshay KM, Gallicano GI (2009) miR-17 family miRNAs are expressed during early mammalian development and regulate stem cell differentiation. Dev Biol 326: 431–433.
42. Rebelatto CK, Aguiar AM, Moreto MP, Senegaglia AC, Hansen P, et al. (2008) Dissimilar differentiation of mesenchymal stem cells from bone marrow, umbilical cord blood, and adipose tissue. Experimental biology and medicine Maywood NJ 233: 901–913.
43. Zambelli F, Pavesi G, Gissi C, Horner DS, Pesole G (2010) Assessment of orthologous splicing isoforms in human and mouse orthologous genes. BMC Genomics 11: 534.
44. Trapnell C, Pachter L, Salzberg SL (2009) Tophat: discovering splice junctions with RNAseq. Bioinformatics 25: 1105–1111.
45. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNAseq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28: 511–515.
46. Gentleman R, Falcon S (2012) microRNA: Data and functions for dealing with microRNAs. R package version 1.16.0.
47. Doench JG, Petersen CP, Sharp PA (2003) siRNAs can function as miRNAs. Genes & Development 17: 438–442.
48. Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular Cell 27: 91–105.
49. Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, et al. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA 13: 1894–1910.
50. Bartel DP (2009) Review microRNAs:target recognition and regulatory functions. Cell 136: 215–233.
51. Hu Z (2009) Insight into microRNA regulation by analyzing the characteristics of their targets in humans. BMC Genomics 10: 594.