Chapter 5

Results of transfection experiments

The bidirectional perturbation experiments (i.e. the experiment in which cultures transfected with miRNA mimics were compared to cultures transfected with miRNA inhibitors) allowed to efficiently detect miR-124 mediated inhibition of gene expression (Chapter 4, section 4.2.1). Therefore, I decided to use this strategy to compile the list of putative direct targets of miR-124, and also to conduct bidirectional experiments to identify targets of other selected mouse miRNAs (the selection of miRNAs for these experiments is described in Chapter 3, section 3.3).

Before the start of these experiments, it was not known at what stage in development of cultures the bidirectional perturbation experiments would work most efficiently to identify miRNA targets. This uncertainty was due to differences between the selected miRNAs (Chapter 3). For example, miR-124 is highly expressed throughout the development of cultures, and miR-434-3p is upregulated (they are referred to as neuronal miRNAs), while miR-143, miR-145 and miR-25 are downregulated (they are referred to as non-neuronal miRNAs). It was not certain, whether the developmental timepoint optimal for identification of targets of neuronal miRNAs would be the same as for non-neuronal (and vice versa). Therefore, miRNA mediated effects of three non-neuronal miRNAs (miR-143, miR-145 and also cel-miR-67, which is a non-mouse miRNA and, hence, non-neuronal in the mouse) and of miR-124 were surveyed across the three developmental timepoints (**3DIV**, **4DIV** and **6DIV**[1]). This survey was conducted using the unidirectional overexpression strategy, where cultures transfected with miRNA mimics were compared to mock transfected cultures.

---

[1]The 4DIV timepoint was selected because of its importance as a switch point in gene expression of developing cultures: after 4DIV the ratio of abundances of neuritic transcripts and of somatic transcripts was above 1, i.e. similar to that in mature neurons (Chapter 3, section 3.1.3). The other two timepoints were picked around the 4DIV timepoint.

The first part of this chapter describes these unidirectional experiments, which resulted in the identification of the timepoints optimal for bidirectional perturbation experiments of neuronal and non-neuronal miRNAs. Additionally, results of these experiments suggested that the endogenous miRNA, miR-124, is a buffer of changes to the transcriptome of mature neurons, as discussed at the end of the first part of the chapter. The second part describes the bidirectional perturbation experiments, which resulted in lists of putative direct targets of several mouse miRNAs. Compilation of putative direct targets of cel-miR-67 from a unidirectional experiment will also be described. In the final part of the chapter, the methodology for identification of miRNA targets in this thesis is validated by comparing of miR-124 thesis targets to miR-124 targets previously reported in published literature.

5.1   Unidirectional overexpression experiments

5.1.1   The effect of cel-miR-67, miR-143 and miR-145 overexpression was maximal at 3DIV or 4DIV

Developmentally downregulated miRNAs were not expected to be highly expressed and be functional in mature neurons (Chapter 3, section 3.3). I selected two of these miRNAs, miR-143 and miR-145, for a series of unidirectional transfection experiments that are described in this section. In addition to these two mouse miRNAs, transfections of the mimic for a non-mouse miRNA, cel-miR-67, were also performed. This miRNA was identified in *Caenorhabditis elegans* and its seed region, *CACAACC*, was different from the seed region of any known mouse miRNA (as of miRBase release 14 (Griffiths-Jones et al., 2008, 2006; Griffiths-Jones, 2004)). Therefore, no specific category of mouse genes was expected to be under endogenous regulation of this miRNA.

Although these three miRNAs were unlikely to be involved in the regulation of normal neuronal function, their ectopic expression could still have an effect on neuronal gene expression. This is because the function of miRNAs is dependent on the generic RISC machinery (Introduction, section 1.1), and, in work by Lim and colleagues, the ectopic expression of a neuronal miR-124 and a muscular miR-1 in HeLa cells had a profound miRNA mediated effect on gene expression (Lim et al., 2005). Therefore it was not surprising that overexpression of cel-miR-67, miR-143 and miR-145 had an effect on gene expression in primary neuronal cultures, too (Figure 5.1).

Primary cultures were transfected with mimics of cel-miR-67 and miR-145 at 3DIV, 4DIV and 6DIV, and with mimics of miR-143 at 3DIV and 6DIV (Methods, section 2.5). Transfections of cel-miR-67 at 4DIV and 6DIV were independently repeated twice (specified by "A" and "B" indices). As these experiments were performed according to the unidirectional overexpression strategy, differential expression was estimated by comparison (contrast) of mRNA profiles of cultures transfected with miRNA mimics to those of mock transfected cultures. Analysis of the microarray data (Methods, section 2.7) revealed pronounced changes in mRNA profiles of cultures transfected with mimics of all three miRNAs (Figure 5.1). In each of the experiments downregulated genes ($P < 0.05$) were significantly enriched (hypergeometric test $P < 0.05$) in genes harbouring miRNA seed matching sites in their 3′UTRs (seed matching sites of 7(2) and 7(1A)-types, see Introduction, section 1.2.1 and Methods, section 2.7).

A trend toward decreasing impact of miRNA overexpression on gene expression was observed in these experiments. For all three miRNAs, the number of highly significantly downregulated genes decreased toward the 6DIV (e.g. numbers of genes downregulated beyond a strict cutoff $P < 1e − 06$ are shown in Table 5.1). Additionally, the most significant enrichment of seed matching sites containing genes was achieved in experiments at 3DIV or 4DIV for all three miRNAs (Figure 5.1). The trend toward decreasing miRNA impact at 6DIV was especially clear when the seed matching site distribution was assessed using the Sylamer program (van Dongen et al., 2008), which accounted for length and composition biases (Figure 5.2, description of the Sylamer method is in Methods, section 2.8).

miRNA	3DIV $P < 1e − 06$	4DIV $P < 1e − 06$	6DIV $P < 1e − 06$
cel-miR-67	7	61 *(exp A)*	1 *(exp A)*
		16 *(exp B)*	0 *(exp B)*
miR-143	40	*na*	0
miR-145	422	37	0

Table 5.1: Number of genes downregulated ($P < 1e − 06$) in miRNA overexpression experiments.

Another finding was the observation of significant biases in the distribution of miR-124 seed matching sites (7(2) and 7(1A)-types) in these experiments. Interestingly, the direction of these biases was reciprocal between 3DIV and 6DIV. At 3DIV transcripts that were upregulated in samples transfected with the mimics of miR-143, miR-145 and cel-miR-67 were depleted of seed matching sites for miR-124 (Figures 5.2a, 5.2b and 5.2c). In

transfections at 6DIV, on the other hand, the upregulated transcripts were enriched in the seed matching sites for miR-124 (Figures 5.2g, 5.2h, 5.2i and 5.2j). Interestingly, a similar trend was observed in gene expression profiles of the development of primary forebrain and hippocampal cultures (Chapter 3, section 3.2.3). There, transcripts that were upregulated early in development (in transition from 1DIV to 2DIV), were depleted of seed matching sites for miR-124 (Chapter 3, Figures 3.10a and 3.10b). On the other hand, transcripts that were upregulated later (in transition from 4DIV to 8DIV), were not depleted in the forebrain cultures (Chapter 3, Figure 3.10c), or enriched in the hippocampal cultures (Chapter 3, Figure 3.10d) for miR-124 seed matching sites.

In summary, the miRNA mediated effects of non-neuronal miRNAs, miR-143, miR-145 and cel-miR-67, were the strongest in transfection experiments at 3DIV or 4DIV. At 6DIV overexpression of these miRNAs had a lower direct effect on gene expression in the cultures. Unexpectedly, an inverse significant bias in the distribution of seed matching sites for miR-124 was observed.

Of the two timepoints (3DIV and 4DIV), the 4DIV timepoint was selected as optimal for the bidirectional experiments (section 5.2). This timepoint was preferred because the maximal seed matching site Sylamer enrichment was detected at 4DIV for miRNAs that were overexpressed at both 3DIV and 4DIV (Figures 5.2d and 5.2f).
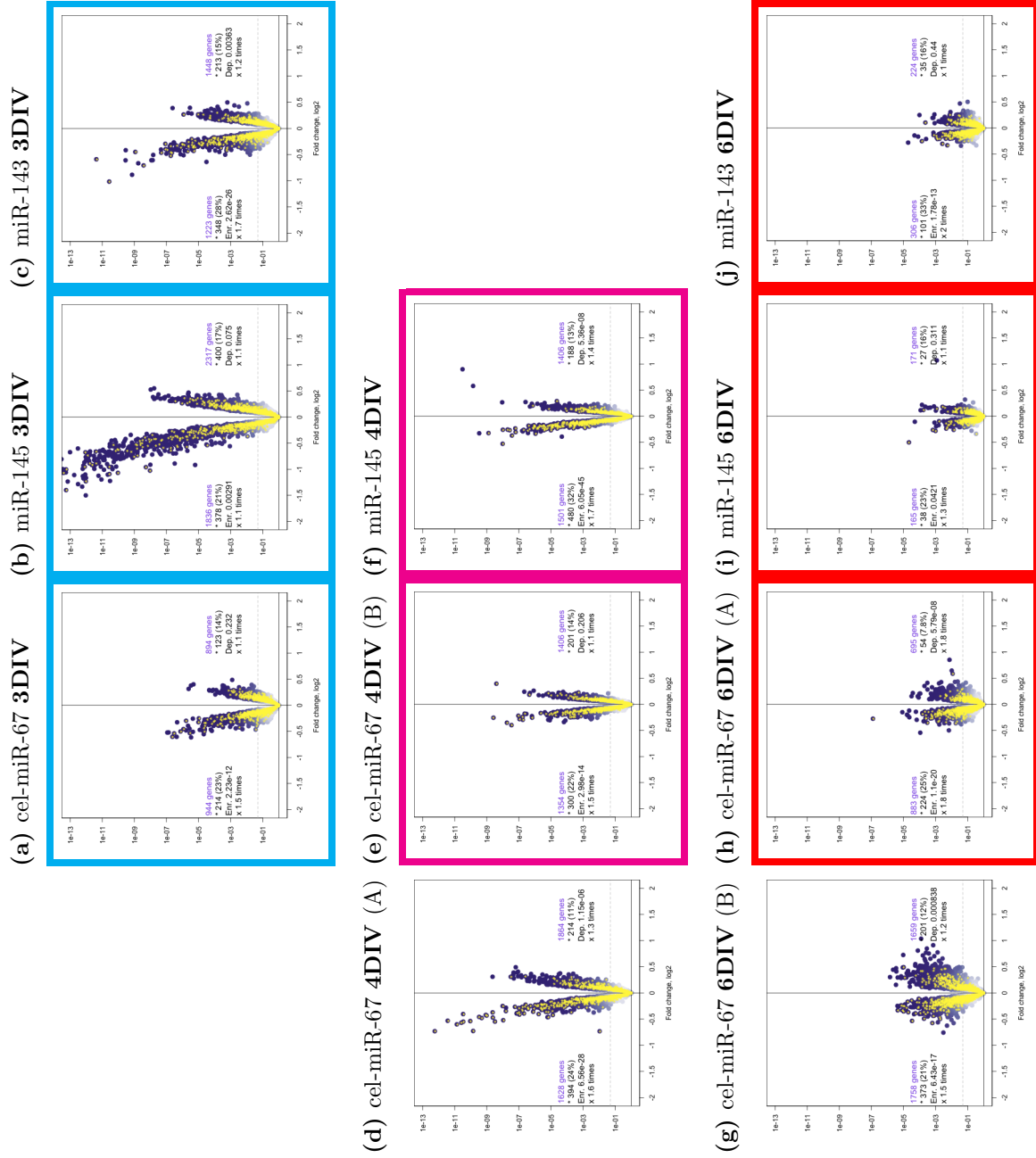
Figure 5.1: **Differential gene expression upon unidirectional over-expression of miR-143, miR-145 and cel-miR-67.**
[See the next page for caption]

Figure 5.1: Differential gene expression and seed enrichment upon unidirectional over-expression of miR-143, miR-145 and cel-miR-67. *[The figure is on the previous page]*
Genes detected by microarrays (using the standard Illumina detection call $P < 0.01$) are shown as the purple dots (the analysis of microarray data is described in Methods, section 2.7). The x-axes represent log_2 of gene expression fold change between samples transfected with miRNA mimics (for the miRNAs named in the titles to the subfigures) in comparison to the matched mock transfected samples. The y-axes represent P-value of differential expression (log_{10} scale), and the horizontal dashed grey lines show P-value cutoff of 0.05. The yellow asterisks mark genes with 3$'$UTRs harbouring one or more seed matching sites (7(2) or (7(1A)-types) for the miRNAs named in the titles to the subfigures. The text in the two halves of the plot area provides the following information: 1) The total number of genes with differential expression P-value more significant than the cutoff (0.05); 2) The total number (and percentage) of genes with seed matching sites for miR-124; 3) The hypergeometric P-value of enrichment (*Enr.*) or depletion (*Dep.*); 4) Fold enrichment or depletion of genes with the seed matching sites "$\times$ *times*" the number that is expected by chance alone. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed matching sites and the hypergeometric enrichment test is in Methods (section 2.8). The subfigures surrounded by the boxes of the same color describe experiments that were performed on the same batch of primary cultures (Methods, section 2.5).
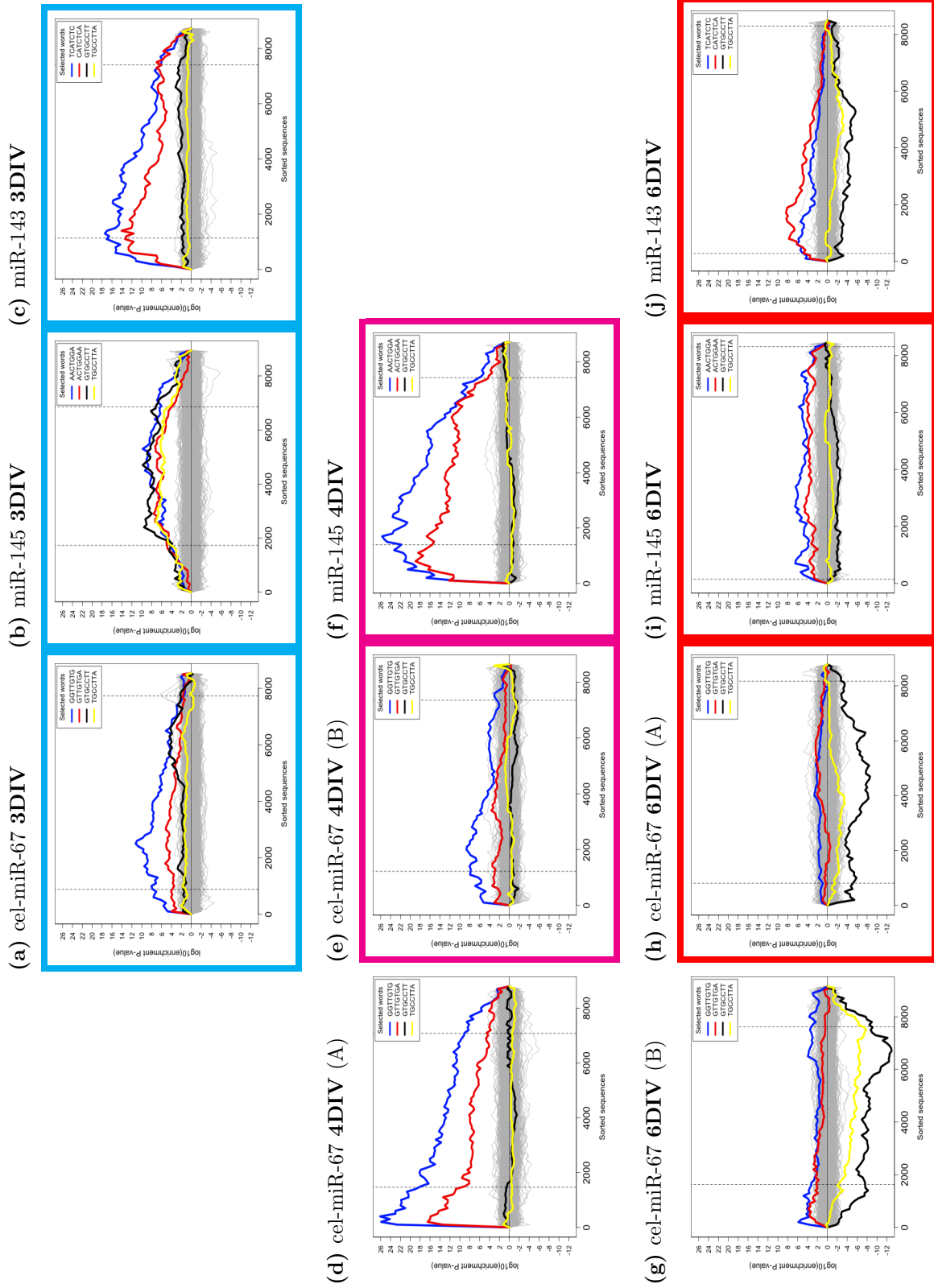
Figure 5.2: Sylamer analysis of biases in distribution of seed matching sites upon unidirectional over-expression of miR-143, miR-145 and cel-miR-67.
[See the next page for caption]

Figure 5.2: Sylamer analysis of biases in distribution of seed matching sites upon unidirectional over-expression of miR-143, miR-145 and cel-miR-67. *[The figure is on the previous page]* The x-axes represent sorted $3'$UTRs corresponding to detectably expressed genes (detected with the standard Illumina detection call $P < 0.01$, see Methods, section 2.7). These genes are ordered **from the most downregulated to the most upregulated** by fold change t-statistic for differential expression between samples transfected with miRNA mimics (for the miRNAs named in the titles to the subfigures) in comparison to the matched mock transfected samples. The y-axes represent the hypergeometric P-values for occurrence biases of 876 nucleotide words complementary to the seed regions (7(2) and 7(1A)-types) of the complete set of 581 distinct mouse miRNAs, according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Positive values on the y-axes correspond to an enrichment $(+|log_{10}(\text{P-value})|)$ and negative values to a depletion $(-|log_{10}(\text{P-value})|)$. The vertical dashed lines mark the P-value cutoff (0.05) on both sides of the ranked gene lists. The blue and the red lines show enrichment profiles of 7(2) and 7(1A)-type seed matching sites for the miRNAs named in the titles to the subfigures, the black and the yellow lines – for miR-124, the grey lines – for the rest of the distinct seed matching sites. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed regions and parameters of Sylamer (van Dongen et al., 2008) is in Methods (section 2.8). The full description of the Sylamer method is in the Introduction (section 2.8). The subfigures surrounded by the boxes of the same color describe experiments that were performed on the same batch of primary cultures (Methods, section 2.5).

5.1.2 The effect of miR-124 overexpression was maximal at 6DIV

Unlike the three miRNAs described in the previous section, miR-124 is known to have important functions in neuronal differentiation and development (Chapter 3, Table 3.3). Thus it was expected that its perturbation would have a significant impact on gene expression in primary cultures. It was not known, however, at which developmental timepoint perturbations of miR-124 would have the strongest impact.

In contrast to the outcome of the unidirectional overexpression of cel-miR-67, miR-143 and miR-145, the extent of gene expression changes increased dramatically between 3DIV and 6DIV in miR-124 overexpression experiments (Chapter 4, Figures 4.9a and 4.9c). At 3DIV, 847 genes were downregulated ($P < 0.05$) in contrast of cultures transfected with the mimic versus mock transfected cultures, while at 6DIV there were 2,070 downregulated ($P < 0.05$) genes. Additionally, the direct contribution of miR-124 to differential expression also appeared to have increased between 3DIV and 6DIV. Sylamer analysis of biases of miRNA seed matching site distributions in 3′UTRs showed that the peak of enrichment for miR-124 7(2)-type seed matching site increased between 3DIV and 6DIV from $P < 1e-14$ to $P < 1e-33$ (Chapter 4, Figures 4.8a and 4.8c). Therefore, the bidirectional perturbations at 6DIV were used to compile a list of putative miR-124 targets (see section 5.2.1). Additionally, based on the miR-124 effect at 6DIV as an example of an effect of a miRNA that is functional in neurons, the 6DIV timepoint was also selected for identification of targets of miRNAs upregulated in the development (section 5.2.3).

5.1.3 Endogenous miR-124 constrained gene expression in more mature primary neurons (6DIV)

As well as allowing identification of optimal timepoints for bidirectional experiments, the series of unidirectional perturbation experiments uncovered an interesting trend: the effect of overexpression of non-neuronal miRNAs decreased in more mature neurons (6DIV), while it increased toward 6DIV in the case of miR-124 overexpression experiments. Importantly, the increase in impact in miR-124 experiments demonstrated that the decrease in impact of non-neuronal miRNAs could not be explained by the decrease in transfection efficiency in more mature cultures.

Transfections of mimics of non-neuronal miRNAs (miR-143, miR-145 and cel-miR-67) had the biggest effect on gene expression in primary cultures at 3DIV and 4DIV, while it decreased at 6DIV. This effect is seen as shrinking of the plots of differential expression for all three non-neuronal miRNAs at 6DIV (Figure 5.1). Sylamer analysis

of seed matching site distributions in 3′UTRs of transcripts expressed in the cultures showed that enrichment of seed matching sites in 3′UTRs of downregulated transcripts also decreased at 6DIV (Figure 5.2). Therefore, both the overall effect of transfections and direct inhibition by the transfected non-neuronal miRNAs decreased in relatively more mature primary neuronal cultures (6DIV).

Interestingly, a reciprocal trend in the distribution of seed matching sites for miR-124 was observed in transfections of non-neuronal miRNAs at 3DIV and 6DIV. At 3DIV, transcripts that were upregulated upon the transfections of the mimics, were depleted of the seed matching sites for miR-124 (the black and yellow lines in Figures 5.2a, 5.2b and 5.2c). An inverse trend was observed at 6DIV, where upregulated transcripts were enriched in miR-124 seed matching sites (the black and yellow lines in Figures 5.2g, 5.2h, 5.2i and 5.2j). It is possible, that endogenous miR-124 imposes a limit to transcriptome changes in more mature neurons (6DIV) by moderating the extent of upregulation of the transcripts that harbour miR-124 seed matching sites. On the other hand, changes to the transcriptome in less mature neurons (3DIV and 4DIV), are relatively less restricted, because upregulated transcripts are neither enriched nor significantly depleted of the seed matching sites for miR-124.

Profiling of differential expression in developing primary neuronal cultures suggested an explanation for an increased capacity of miR-124 to moderate (or buffer) changes to the transcriptome in more mature neurons. Transcripts that were upregulated early in development of primary cultures (in transition from 1DIV to 2DIV) were seen to be depleted of miR-124 seed matching sites (Chapter 3, Figures 3.10a and 3.10b). Therefore, in very immature primary neurons transcripts that can in principle be inhibited by miR-124 (i.e. transcripts with seed matching sites for miR-124) may not yet be available. The relatively small scope for miR-124 activity in very immature neurons make sense biologically in the light of a rapid spurt of neurites and early synaptogenesis events that were observed in the early stages of development of primary neuronal cultures (Valor et al., 2007). Perhaps at these early stages it would be detrimental for neurons if limits to the changes in the transcriptome were imposed by highly expressed neuronal miRNAs, such as miR-124.

At later stages in development of cultures (the transition from 4DIV to 8DIV), 3′UTRs of developmentally upregulated transcripts were seen to be not depleted (primary forebrain cultures, Chapter 3, Figure 3.10c) or even enriched (primary hippocampal cultures, Chapter 3, Figure 3.10d) in seed matching sites for miR-124. Therefore, in more mature neurons there is scope for endogenous miR-124 to limit changes to the transcriptome,

which can explain the decrease in impact of transfections of non-neuronal miRNAs at 6DIV. Similarly, the increase in the impact of transfections of miR-124 itself at 6DIV can also be explained by the increased number of transcripts available for miR-124 mediated inhibition in more mature cultures.

It should be pointed out, that the hypothesis of miR-124 to buffer upregulated in mature neurons genes and by that restrict perturbations of the transcriptome, is speculative. However, this hypothesis is convenient as an effective theory at the moment, because it explains some of the observations that are described in Chapter 6. Experiments that can test this hypothesis are suggested in the Discussion (section 7.4).

Summary of section 5.1

Transfections of mimics of non-neuronal miRNAs (miR-143 and miR-145, cel-miR-67) in primary forebrain cultures at 3DIV, 4DIV and 6DIV elicited changes in gene expression, which were the most significant at 3DIV or 4DIV. Sylamer analysis suggested that the direct contribution of the transfected miRNAs to changes in differential gene expression was also greatest at 3DIV or 4DIV. In the end, the 4DIV timepoint was selected as the best timepoint to conduct bidirectional perturbation experiments aiming to identify putatively direct targets of non-neuronal miRNAs.

Transfections of mimics of the neuronal miR-124 at 3DIV and 6DIV elicited changes in gene expression at both of the timepoints. Sylamer analysis showed that the direct contribution of miR-124 mediated inhibition to differential gene expression increased from 3DIV to 6DIV. Therefore, the 6DIV timepoint was selected as the best timepoint for bidirectional experiments on neuronal miRNAs.

Additionally, significant biases in the distribution of miR-124 seed matching sites where observed in the experiments where non-neuronal miRNAs were exogenously added (transfected) into the primary neuronal cultures. The direction of these biases in the 6DIV experiments (an enrichment of miR-124 seed matching sites in the upregulated transcripts), in conjunction with the overall decrease in the effect of 6DIV transfections on gene expression, suggested that the endogenous miR-124 can act as a buffer for genes that are upregulated in more mature neurons (6DIV).

5.2 Bidirectional perturbation experiments

5.2.1 Identification of targets for a steady state expressed miRNA (miR-124)

Of miRNAs that were expressed at the steady state level in the development of primary neuronal cultures, I selected two for bidirectional transfection experiments[1]: miR-124 and miR-103 (Chapter 3, section 3.3). Bidirectional transfection experiments with miR-124 were conducted at all three experimental timepoints (3DIV, 4DIV and 6DIV), because of the special interest in this miRNA (Introduction, section 1.1.3). On the other hand, miR-103, although it was highly expressed, was not previously reported as a functionally important miRNA for neurons. Therefore, the miR-103 transfection experiment was performed at 4DIV, as this timepoint was conducive to detection of miRNA mediated effects in both miR-124 transfection (see below) and transfections of non-neuronal miRNAs (see section 5.2.2).

Both the bidirectional and unidirectional contrasts were available for analysis of miR-124 and miR-103 transfections (Chapter 4, section 4.2.1). In experiments on miR-124 at 3DIV and 6DIV, where matched mock transfected samples were available (Methods, section 2.5), the use of inhibition increased detection of enrichment of miR-124 seed matching sites (Chapter 4, Figure 4.8). In the miR-103 experiment both the mock transfection and the inhibition were conducted, and the bidirectional strategy increased enrichment of genes with miR-103 seed matching sites among downregulated genes (from approximately 1.1 times ($P < 0.039$) to 1.3 times ($P < 0.017$) more than expected by chance alone, Figure 5.3).

Despite the fact that the bidirectional strategy increased enrichment of genes with miRNA seed matching sites among downregulated genes for both miR-124 and miR-103, Sylamer analysis (Methods, section 2.8) showed that only in case of miR-124 the enrichment of seed matching sites was more significant than the background distribution (Figure 5.4). Therefore it was possible to compile putative direct targets for miR-124, but not for miR-103. Of the three miR-124 perturbation experiments, the biggest enrichment of miR-124 seed matching sites in 3′UTRs of downregulated genes was observed at 6DIV (Figure 5.4c). Therefore, putative direct targets of miR-124 were derived from the 6DIV experiments. The cutoff P-value of 0.01 in differential expression coincided well with the

---

[1]As introduced in Chapter 4 (section 4.2.1), the word **bidirectional** refers to comparison of transfections of miRNA mimics with transfections of miRNA inhibitors. On the other hand, the word **unidirectional** refers to comparison of transfections of miRNA mimics with mock transfections.
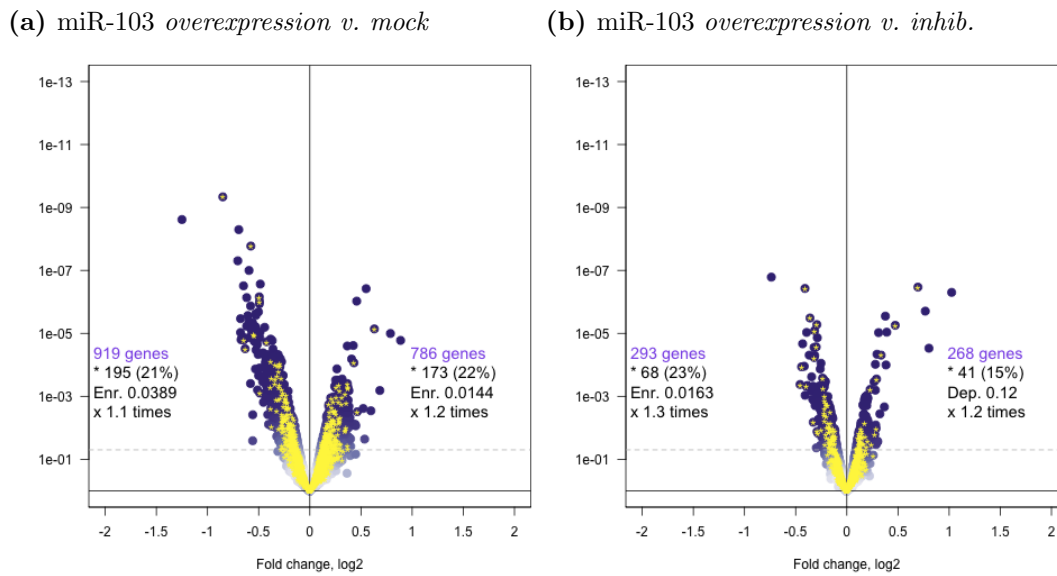
(a) miR-103 *overexpression v. mock*   **(b)** miR-103 *overexpression v. inhib.*

Figure 5.3: Differential gene expression and seed matching site enrichment in miR-103 transfection experiments at 4DIV.
Genes detected by microarrays (using the standard Illumina detection call $P < 0.01$) are shown as the purple dots (analysis of microarray data is described in Methods, section 2.7). The x-axes represent log_2 of gene expression fold change between samples transfected: 5.3a – with the mimic of miR-103 in comparison to the matched mock transfection at 4DIV; 5.3b – with the mimic of miR-103 in comparison to transfection with the inhibitor of miR-103 at 4DIV; The y-axes represent P-value of differential expression (log_{10} scale), and the horizontal dashed grey lines show P-value cutoff of 0.05. The yellow asterisks mark genes [encoding transcripts] with 3′UTRs harbouring one or more seed matching sites (7(2) or 7(1A)-types) for miR-103. The text in the two halves of the plot area provides the following information: 1) The total number of genes with differential expression P-value more significant than the cutoff (0.05); 2) The total number (and percentage) of genes with seed matching sites for miR-103; 3) The hypergeometric P-value of enrichment (*Enr.*) or depletion (*Dep.*); 4) Fold enrichment or depletion of genes with the seed matching sites "× *times*" the number that is expected by chance alone. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed matching sites and the hypergeometric enrichment test is in Methods (section 2.8).

slope and the peak of the Sylamer distribution of miR-124 seed matching sites (both 7(2) and 7(1A)-types, see blue and red line and the first vertical dashed line in Figure 5.4c). Overall, there were 399 genes that contained one or more miR-124 seed matching sites in their 3′UTRs and were downregulated beyond the $P < 0.01$ cutoff. These putative direct miR-124 targets are listed in Supplementary Data, Table A.9.
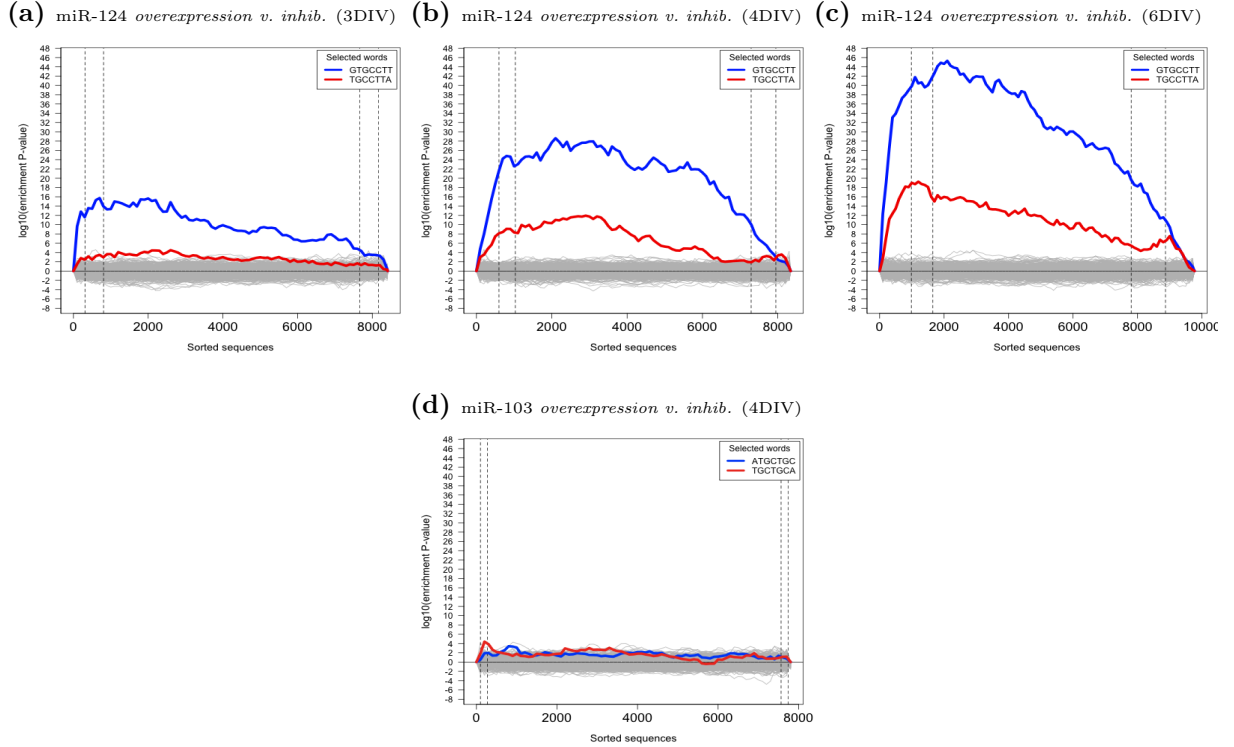
Figure 5.4: Sylamer analysis of biases in distributions of seed enrichment in miR-124 and miR-103 transfection experiments.

The x-axes represent sorted 3'UTRs corresponding to detectably expressed genes (detected with the standard Illumina detection call $P < 0.01$, see Methods, section 2.7). These genes are ordered **from the most downregulated to the most upregulated** by fold change t-statistic for differential expression in the following transfections: 5.4a – with the mimic of miR-124 in comparison to the transfection with the inhibitor of miR-124 at 3DIV; 5.4b – with the mimic of miR-124 in comparison to the transfection with the inhibitor of miR-124 at 4DIV; 5.4c – with the mimic of miR-124 in comparison to the transfection with the inhibitor of miR-124 at 6DIV; 5.4d – with the mimic of miR-103 in comparison to the transfection with the inhibitor of miR-103 at 4DIV. The y-axes represent the hypergeometric P-values for occurrence biases of 876 nucleotide words complementary to the seed regions (7(2) and 7(1A)-types) of the complete set of 581 distinct mouse miRNAs, according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Positive values on the y-axes correspond to an enrichment ($+|log_{10}(\text{P-value})|$) and negative values to a depletion ($-|log_{10}(\text{P-value})|$). The vertical dashed lines mark the P-value cutoffs (0.01 and 0.05) on both sides of the ranked gene lists. The blue and the red lines show enrichment profiles of 7(2) and 7(1A)-type seed matching sites for miR-124 or miR-103 (see the titles for the subfigures), the grey lines – for the rest of the distinct seed matching sites. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed regions and parameters of Sylamer (van Dongen et al., 2008) is in Methods (section 2.8). The full description of the Sylamer method is in the Introduction (section 2.8).

5.2.2 Identification of targets for downregulated miRNAs in development (miR-143, miR-145 and miR-25) and of a non-mouse miRNA (cel-miR-67)

Of miRNAs that were downregulated in the development of primary neuronal cultures, I selected three for bidirectional transfection experiments: miR-143, miR-145 and miR-25 (the selection is described in Chapter 3, section 3.3). These experiments were performed on a single batch of cultures (Methods, section 2.5) at 4DIV. The 4DIV timepoint was selected as results of unidirectional overexpression experiments (see section 5.1.1) suggested that detection of targets of putatively non-neuronal miRNAs was the most efficient at 4DIV (Figure 5.2). In addition to the compilation of lists of putative direct targets of miR-143, miR-145 and miR-25, identification of putative direct targets of cel-miR-67 (a non-mouse miRNA) is also described at the end of this section.

By conducting a mock transfection, in addition to overexpression and inhibition, it was possible to demonstrate that the use of bidirectional contrasts improved detection of miRNA-mediated effects for all three of the downregulated miRNAs (miR-143 (Figures 5.5a and 5.5d), miR-145 (Figures 5.5b and 5.5e) and miR-25 (Figures 5.5c and 5.5f)). Upon bidirectional perturbation of miR-143, the enrichment of the genes with the seed matching sites complementary to miR-143 increased among downregulated genes (differential expression $P < 0.05$) from approximately equal to that expected by chance alone to 1.1 times more than expected ($P < 0.05$). This increase was also true for miR-145 (1.5 fold enrichment in contrast with mock transfection and 1.6 times in contrast with inhibition), and for miR-25 (increasing from 2.4 times to 2.7 times).

Sylamer analysis (Methods, section 2.8) showed that bidirectional perturbation of miR-143, miR-145 and miR-25, leads to identification of significant miRNA-mediated effects in each of the experiments (Figure 5.6). In all experiments, the enrichment of seed matching sites (7(2) and 7(1A)-types) for the transfected miRNAs was more significant than that of all other tested nucleotide words (shown by the grey lines, Figure 5.6). However, in miR-143 and miR-145 experiments the peak of enrichment of the seed matching sites for several other nucleotide words approached that of the seed matching sites for miR-145 and miR-143. These nucleotide words were either of low complexity (e.g. *GCCCCGG* in the case of miR-143 experiment), were related to the polyadenylation site, or corresponded to miRNAs with low abundance and unknown function. Perhaps these distributions were indicative of interesting biological phenomena, however they were unlikely to be

directly related to activity of the transfected miRNAs[1], and therefore they were not studied further.

Interestingly, Sylamer enrichment of miR-25 seed matching sites was very significant (Figure 5.6c). The Sylamer enrichment peak for 7(2)-type seed matching site for miR-25 ($GTGCAAT$) corresponded to a hypergeometric enrichment P-value below $1e - 52$, which was the most significant Sylamer enrichment P-value detected in this thesis project. Surprisingly, the effect of miR-25 on expression of transcripts harbouring seed matching sites for miR-25 was even more dramatic than that of miR-124 ($P < 1e-44$ in transfection at 6DIV, see Figure 5.4c). The extremely high enrichment of seed matching sites for miR-25 was likely a combination of high efficiency of miR-25 as a guide of the RISC to destabilise its targets (Introduction, section 1.1.2), and a relatively small repertoire of its potential targets[2]. It should be noted that in transfection experiments performed in this thesis, there were substantial fluctuations in Sylamer enrichment results[3]. Therefore the result of one miR-25 transfection experiment cannot serve as definitive evidence of its exceptional properties. However, transfection of miR-25 into mutant mouse embryonic stem cells was found to lead to a similarly significant destabilisation of seed matching site containing mRNAs (Matthew Davis, personal communication). Interestingly, expression of miR-25 was shown to be induced in tumors (Poliseno et al., 2010), which is a relatively unusual for miRNAs property (as miRNAs are generally reduced in tumors (Thomson et al., 2006; Lotterman et al., 2008)). Therefore, a strong inhibition of miR-25 targets in differentiated cell types, such as cells in primary neuronal cultures, may be related to the role of miRNAs in carcinogenesis. This makes the role of miR-25 in primary neuronal cultures a relevant subject for research of miRNAs in cancer (Discussion, section 7.4).

Identification of significant enrichment for miR-143, miR-145 and miR-25 seed matching sites in downregulated genes enabled the compilation of lists of their putative direct targets. As in the case of miR-124 (see section 5.2.1), the 0.01 P-value cutoff of differential expression corresponded well to the peaks in the enrichment of seed matching sites for

---

[1]It should be noted that these enrichments peaked in the right half of the plot (i.e. the dominant trend was depletion of these seed matching sites in the 3′UTRs of upregulated genes), whilst in case of direct miRNA mediated effects on gene expression the peak is expected to be in the left half (i.e. consistent with a enrichment of seed matching sites in the 3′UTRs of downregulated genes being a dominant trend). For interpretation of Sylamer plots see Introduction, section 1.2.3.

[2]Of genes represented on the microarray platform there were 2,083 that uniquely corresponded to transcripts harbouring one or more seed matching sites for miR-25, while this number is over 3,000 for miR-143 and miR-145.

[3]For example, the Sylamer enrichment P-values of enrichment of the seed matching sites for cel-miR-67 differed by more than 10 orders of magnitude in replicate transfections at 4DIV, see Figures 5.2d and 5.2e.

transfected miRNAs (the first dashed vertical lines in Figure 5.6). Selection of downregulated genes (differential expression $P < 0.01$), harbouring seed matching sites for the transfected miRNAs in their 3′UTRs, identified 272 putative direct targets of miR-143, 358 targets of miR-145 and 196 targets of miR-25. These putative direct miRNA targets are listed in Supplementary Data, Table A.10 (miR-143 targets), Table A.11 (miR-145 targets) and Table A.12 (miR-25 targets).

Transfections of cel-miR-67 (naturally expressed in *Caenorhabditis elegans*) were only possible using the unidirectional overexpression design as its inhibition was not available in mouse cells. The maximal enrichment of cel-miR-67 seed matching sites in 3′UTRs of downregulated in the transfections genes was detected in one of the experiments at 4DIV (the experiment marked by the index "A", see Figures 5.1d and 5.2d). Therefore this experiment was used to compile putative direct targets of cel-miR-67 in mouse primary neuronal cultures. A P-value cutoff of 0.05 was used[4] and it led to compilation of a list of 394 putative direct cel-miR-67 targets. These putative direct targets of cel-miR-67 in mouse primary neurons are listed in Supplementary Data, Table A.13.

---

[4]The 0.05 cutoff, rather than 0.01, was used for compilation of target of cel-miR-67, because the design of cel-miR-67 transfection experiments on cel-miR-67 was not optimised for detection of direct miRNA-mediated effects and it was performed according to the unidirectional strategy. For two out of three unidirectional transfections of the mimics of non-neuronal miRNAs at 4DIV the 0.05 cutoff approximately coincided with the Sylamer enrichment peak (Figure 5.2).

(a) miR-143 *overexpression v. mock* **(b)** miR-145 *overexpression v. mock* **(c)** miR-25 *overexpression v. mock*
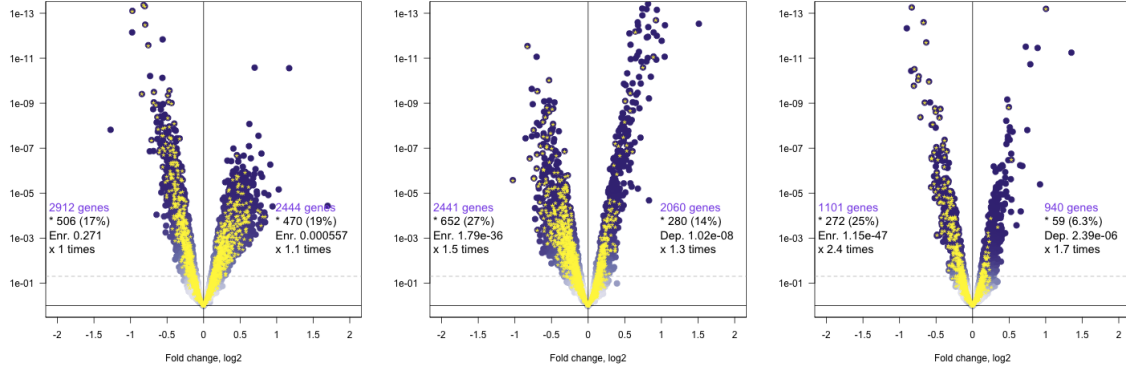
(d) miR-143 *overexpression v. inhib.* **(e)** miR-145 *overexpression v. inhib.* **(f)** miR-25 *overexpression v. inhib.*
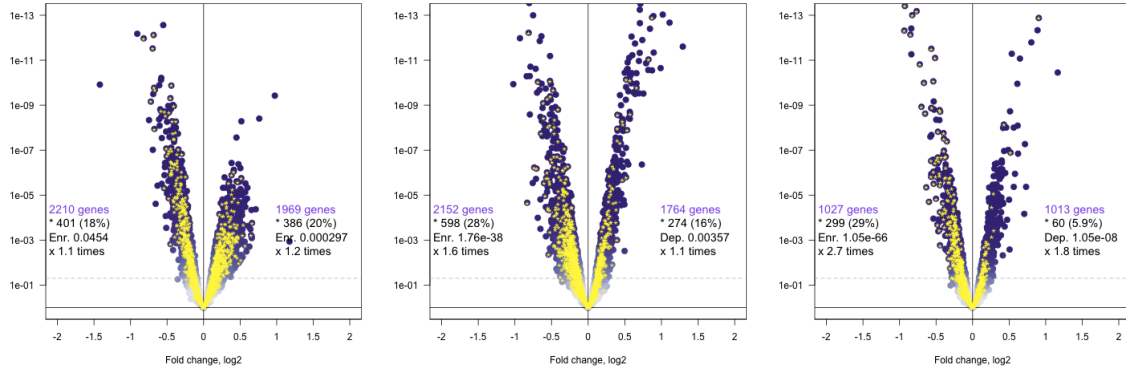
Figure 5.5: Differential gene expression and seed matching site enrichment in miR-143, miR-145 and miR-25 transfections experiments at 4DIV.
Genes detected by microarrays (using the standard Illumina detection call $P < 0.01$) are shown as the purple dots (analysis of microarray data is described in Methods, section 2.7). The x-axes represent log_2 of gene expression fold change between samples transfected: 5.5a – with the mimic of miR-143 in comparison to the matched mock transfection at 4DIV; 5.5b – with the mimic of miR-145 in comparison to the matched mock transfection at 4DIV; 5.5c – with the mimic of miR-25 in comparison to the matched mock transfection at 4DIV; 5.5d – with the mimic of miR-143 in comparison to the transfection with the inhibitor of miR-143 at 4DIV; 5.5e – with the mimic of miR-145 in comparison to the transfection with the inhibitor of miR-145 at 4DIV; 5.5f – with the mimic of miR-25 in comparison to the transfection with the inhibitor of miR-25 at 4DIV. The y-axes represent P-value of differential expression (log_{10} scale), and the horizontal dashed grey lines show P-value cutoff of 0.05. The yellow asterisks mark genes [encoding transcripts] with 3'UTRs harbouring one or more seed matching sites (7(2) or 7(1A)-types) for the miRNAs in the titles of the subfigures. The text in the two halves of the plot area provides the following information: 1) The total number of genes with differential expression P-value more significant than the cutoff (0.05); 2) The total number (and percentage) of genes with seed matching sites for the miRNAs in the titles to the subfigures; 3) The hypergeometric P-value of enrichment (*Enr.*) or depletion (*Dep.*); 4) Fold enrichment or depletion of genes with the seed matching sites "$\times$ *times*" the number that is expected by chance alone. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7); the identification of seed matching sites and the hypergeometric enrichment test is in Methods (section 2.8).

(a) miR-143 *overexpression v. inhib.* **(b)** miR-145 *overexpression v. inhib.* **(c)** miR-25 *overexpression v. inhib.*
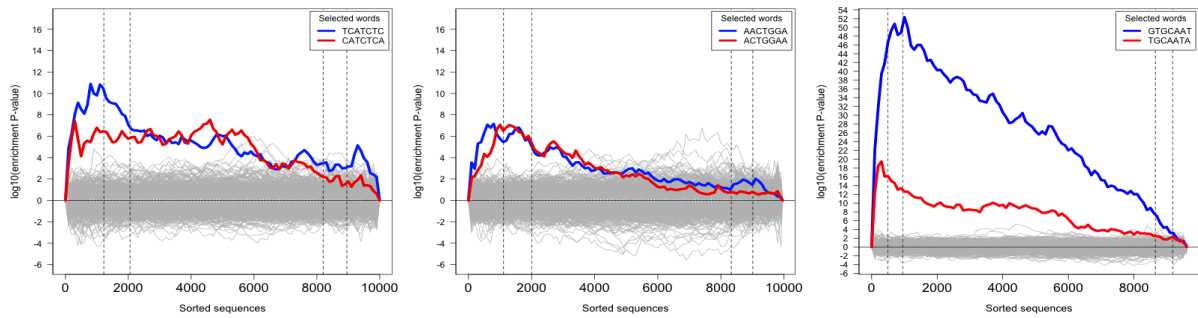

Figure 5.6: Sylamer analysis of biases in distributions of seed matching sites in miR-143, miR-145 and miR-25 transfection experiments at 4DIV.
The x-axes represent sorted 3′UTRs corresponding to detectably expressed genes (detected with the standard Illumina detection call $P < 0.01$, see Methods, section 2.7). These genes are ordered **from the most downregulated to the most upregulated** by fold change t-statistic for differential expression in the following transfections: 5.6a – with the mimic of miR-143 in comparison to the transfection with the inhibitor of miR-143 at 4DIV; 5.6b – with the mimic of miR-145 in comparison to the transfection with the inhibitor of miR-145 at 4DIV; 5.6c – with the mimic of miR-25 in comparison to the transfection with the inhibitor of miR-25 at 4DIV. The y-axes represent the hypergeometric P-values for occurrence biases of 876 nucleotide words complementary to the seed regions (7(2) and 7(1A)-types) of the complete set of 581 distinct mouse miRNAs, according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Positive values on the y-axes correspond to an enrichment $(+|log_{10}(\text{P-value})|)$ and negative values to a depletion $(-|log_{10}(\text{P-value})|)$. The vertical dashed lines mark the P-value cutoffs (0.01 and 0.05) on both sides of the ranked gene lists. The blue and the red lines show enrichment profiles of 7(2) and 7(1A)-type seed matching sites for either of miR-143, miR-145 or miR-25 (see the titles of the subfigures), the grey lines – for the rest of the distinct seed matching sites. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed regions and parameters of Sylamer (van Dongen et al., 2008) is in Methods (section 2.8). The full description of the Sylamer method is in the Introduction (section 2.8).

5.2.3   Identification of targets for an upregulated miRNA in development (miR-434-3p)

Of miRNAs that were upregulated in the development of primary cultures, I selected four for bidirectional transfection experiments. These miRNAs were miR-370, miR-410, miR-551b and miR-434-3p (the selection is described in Chapter 3, section 3.3). Unlike miRNAs of the downregulated category and the non-mouse cel-miR-67, the upregulated miRNAs were assumed to be functional in neurons (Chapter 3, section 3.3). Bidirectional perturbations at 6DIV worked best for induction of miRNA mediated effects in miR-124 transfection experiments (Figure 5.1.2), which has several known functions in neurons (Introduction, section 1.1.3). Therefore transfection experiments to identify targets of upregulated miRNAs were performed according to the bidirectional strategy at 6DIV. Additionally, levels of expression of the upregulated miRNAs were increasing during development (Chapter 3, section 3.2.1), therefore it was reasoned that inhibition of these miRNAs may have a greater effect at the later timepoint of 6DIV.

Bidirectional perturbation of three miRNAs (miR-370, miR-410 and miR-551b) was conducted in one batch of cultures, while the miR-434-3p experiment was performed separately (Methods, section 2.5). By carrying out the mock transfection as a part of the miR-434-3p experiment, it was possible to conclude that the bidirectional perturbation strategy had worked successfully to improve detection of direct miRNA mediated effects. Enrichment of transcripts with miR-434-3p seed matching sites among all downregulated transcripts ($P < 0.05$) was approximately 1.2 fold more than expected ($P < 1.48e - 06$) in case of the unidirectional contrast, and 1.7 fold more than expected ($P < 3.1e - 15$) in case of the bidirectional contrast (Figure 5.7).

Sylamer analysis (Methods, section 2.8), showed that bidirectional perturbation of only miR-434-3p produced a significant miRNA-mediated effect on gene expression (Figure 5.8). Significant enrichment for miR-434-3p seed matching sites in 3′UTRs of downregulated genes allowed the compilation of a list of its putative direct targets. The differential expression cutoff was again chosen to be $P < 0.01$, as it corresponded well to the Sylamer peak of the enrichment of the 7(2)-type seed matching site complementary to the seed region of miR-434-3p (the first dashed vertical line in Figure 5.8a). There were 112 genes downregulated beyond the cutoff ($P < 0.01$) and containing one or more seed matching sites (7(2) or 7(1A)-types) for miR-434-3p, which comprised the list of its putative direct targets. The putative direct targets of miR-434-3p are listed in Supplementary Data, Table A.14.
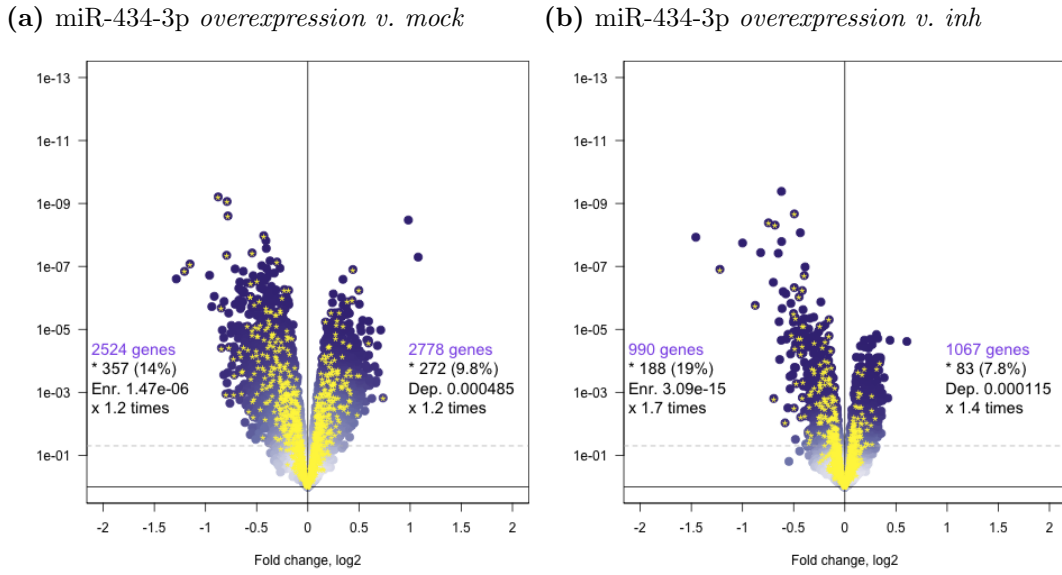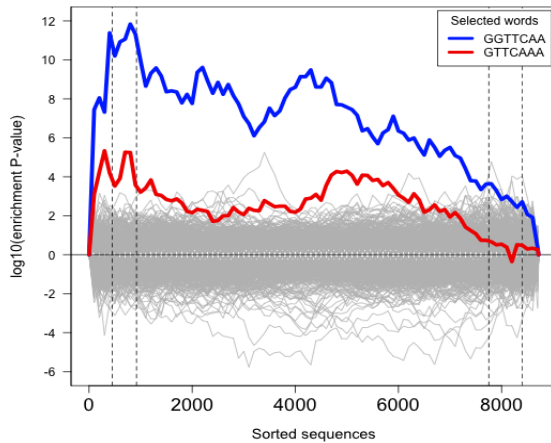
(a) miR-434-3p *overexpression v. mock*   **(b)** miR-434-3p *overexpression v. inh*

Figure 5.7: Differential gene expression and seed matching site enrichment in miR-434-3p transfection experiments at 6DIV.
Genes detected by microarrays (using the standard Illumina detection call $P < 0.01$) are shown as the purple dots (the analysis of microarray data is described in Methods, section 2.7). The x-axes represent log_2 of gene expression fold change between samples transfected: 5.7a – with the mimic of miR-434-3p in comparison to the matched mock transfection at 6DIV; 5.7b – with the mimic of miR-434-3p in comparison to the transfection with the inhibitor of miR-434-3p at 6DIV. The y-axes represent P-value of differential expression (log_{10} scale), and the horizontal dashed grey lines show P-value cutoff of 0.05. The yellow asterisks mark genes [encoding transcripts] with 3′UTRs harbouring one or more seed matching sites (7(2) or 7(1A)-types) for miR-434-3p. The text in the two halves of the plot area provides the following information: 1) The total number of genes with differential expression P-value more significant than the cutoff (0.05); 2) The total number (and percentage) of genes with seed matching sites for miR-434-3p; 3) The hypergeometric P-value of enrichment (*Enr.*) or depletion (*Dep.*); 4) Fold enrichment or depletion of genes with the seed matching sites "× *times*" the number that is expected by chance alone. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed matching sites and the hypergeometric enrichment test is in Methods (section 2.8).

Summary of section 5.2

Using timepoints that were identified in unidirectional overexpression experiments (section 5.1) as best allowing identification of putative direct targets of miRNAs, the bidirectional transfection experiments were performed on the nine selected mouse miRNAs. As a result of these experiments, lists of putatively direct targets were compiled for miR-124, miR-434-3p, miR-143, miR-145 and miR-25. Additionally, using results of a unidirectional overexpression experiment, the list of putative direct targets was compiled for a non-mouse miRNA, cel-miR-67.
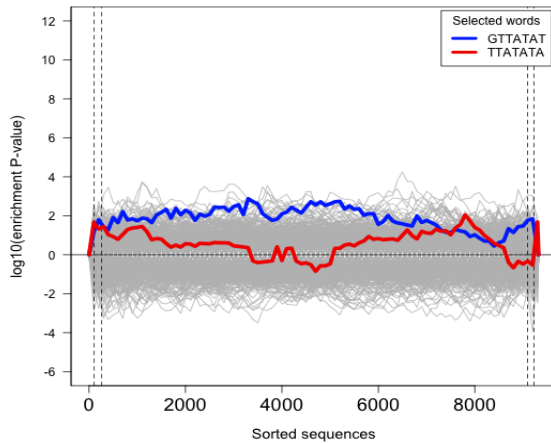
(a) miR-434-3p *overexpression v. inhib.*

(b) miR-370 *overexpression v. inhib.*
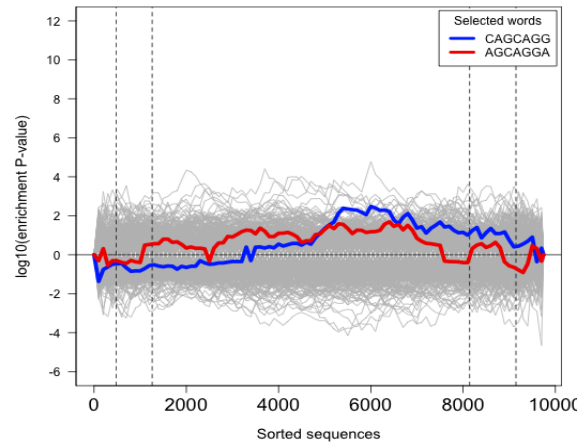
(c) miR-410 *overexpression v. inhib.*

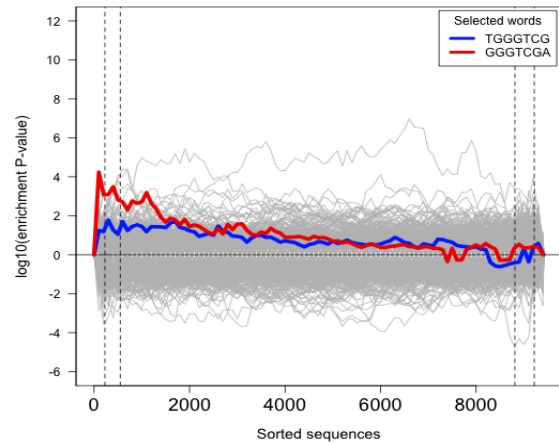**(d)** miR-551b *overexpression v. inhib.*

Figure 5.8: Sylamer analysis of distribution biases in distributions of seed matching sites in miR-434-3p, miR-370, miR-410 and miR-551b transfection experiments at 6DIV.
The x-axes represent sorted 3′UTRs corresponding to detectably expressed genes (detected with the standard Illumina detection call $P < 0.01$, see Methods, section 2.7). These genes are ordered **from the most downregulated to the most upregulated** by fold change t-statistic for differential expression in the following transfections: 5.8a – with the mimic of miR-434-3p in comparison to the transfection with the inhibitor of miR-434-3p at 6DIV; 5.8b – with the mimic of miR-370 in comparison to the transfection with the inhibitor of miR-370 at 6DIV; 5.8c – with the mimic of miR-410 in comparison to the transfection with the inhibitor of miR-410 at 6DIV; 5.8d – with the mimic of miR-551b in comparison to the transfection with the inhibitor of miR-551b at 6DIV. The y-axes represent the hypergeometric P-values for occurrence biases of 876 nucleotide words complementary to the seed regions (7(2) and 7(1A)-types) of the complete set of 581 distinct mouse miRNAs, according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Positive values on the y-axes correspond to an enrichment ($+|log_{10}$(P-value)$|$) and negative values to a depletion ($-|log_{10}$(P-value)$|$). The vertical dashed lines mark the P-value cutoffs (0.01 and 0.05) on both sides of the ranked gene lists. The blue and the red lines show enrichment profiles of 7(2) and 7(1A)-type seed matching sites for either of miR-434-3p, miR-370, miR-410 or miR-551b (see the titles of the subfigures), the grey lines – for the rest of the distinct seed matching sites. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of the seed regions and parameters of Sylamer (van Dongen et al., 2008) is in Methods (section 2.8). The full description of the Sylamer method is in the Introduction (section 2.8).

5.3 Validation of the methodology

The previous section described identification of putative direct targets for six miRNAs: miR-124, miR-434-3p, miR-25, miR-143, miR-145 and cel-miR-67. Targeting of miR-124 was previously studied and published before (Table 5.2). Therefore it was possible to validate methods to identify miRNA targets used in this thesis by comparison of miR-124 targets identified in the thesis to those previously published (or inferred from previously published data).

System	Detection	Reference
HeLa cell line	microarrays	(Lim et al., 2005)
HepG2 cell line	microarrays	(Wang and Wang, 2006)
293 cell line	microarrays	(Karginov et al., 2007)
CAD cell line	microarrays	(Makeyev et al., 2007)
HeLa cell line	protein mass-spectrometry	(Baek et al., 2008)
P13 mouse neocortex	RNA sequencing	(Chi et al., 2009)
HEK293T cell line	microarrays	(Hendrickson et al., 2009)

Table 5.2: Previously published genome-wide studies of miR-124 targeting

To date there are at least seven published experiments aiming at a genome wide description of miR-124 targeting at RNA and protein levels (Table 5.2). This section describes comparison of putative direct targets identified for miR-124 in this thesis (see Chapter 5, section 5.2.1) and targets identified in three of the previously published studies. These three studies were:

- Overexpression of miR-124 in HeLa cells published by **Lim *et al.*** in 2005 (Lim et al., 2005). This was the first study to characterize a global effect of miRNAs on the transcriptome, and multiple miR-124 targets from this study were subsequently experimentally validated elsewhere (Conaco et al., 2006).

- Overexpression of miR-124 in the mouse CAD cell line published by **Makeyev *et al.*** in 2007 (Makeyev et al., 2007). Comparison to the targets that could be derived from this work was more relevant to neurons than the HeLa targets, because CAD cells were originally obtained from a mouse brain tumor (Qi et al., 1997), while HeLa was derived from a human cervical carcinoma (Scherer et al., 1953).

- Identification of miR-124 targets with Argonaute HITS-CLIP method, published by **Chi *et al.*** in 2009 (Chi et al., 2009). Unlike the other two studies, in this study the cell cultures and transfections were not used. Instead, RNA was precipitated from the P13 neocortex and analyzed with new generation sequencing technology. Therefore miR-124 targets identified in this work provided an insight to what the

miR-124 targeting repertoire may be like in a relatively unperturbed and naïve brain (referred to as the *in vivo* targets).

5.3.1   Comparison to targets identified by Lim *et al.*

Lim et al. (Lim et al., 2005) described microarray analysis of gene expression changes upon delivery of miRNA mimics with cationic lipid transfection into HeLa cultures. The authors observed that two tissue specific miRNAs, miR-124 and miR-1, caused significant differential gene expression in HeLa cells. Downregulated genes were enriched for relevant miRNA seed matching sites in their 3′UTRs, which suggested direct miRNA-mediated inhibition of a fraction of these genes. Interestingly, genes that were downregulated by miR-124 in HeLa were found to be relatively lowly expressed in the brain, while genes downregulated by miR-1 were lowly expressed in the muscle. The authors made the suggestion that this observation indicated that the two miRNAs were involved in maintenance of cognate tissue identity, and that it was possible to identify functional targets of miR-124 and miR-1 in HeLa system. The list of miR-124 targets produced by Lim *et al.* was subsequently used as a bench-mark list of experimentally derived miR-124 direct targets (Conaco et al., 2006).

A significant overlap was detected between putative direct miR-124 targets identified in this thesis and the HeLa targets (Figure 5.9). Lim and colleagues identified 129 genes as significantly down-regulated upon miR-124 transfection and containing one or more putative miR-124 target sites. Of these putative direct targets in HeLa, 102 were homologous to mouse genes (according to HomoloGene, version 64 (Sayers et al., 2010)) and contained miR-124 heptamer seed matching sites. Exactly 50.0% of these genes ($P < 1.3e - 22$) were identified as putative miR-124 targets in this work (Chapter 5, section 5.2.1).

5.3.2   Comparison to targets derived from Makeyev *et al.*

In a study of the role of miR-124 by Makeyev *et al.*, CAD cells were transfected with a plasmid expressing pre-miR-124-2, a precursor of miR-124 (Makeyev et al., 2007). The CAD cell line was derived from a mouse brain tumor, and it expressed a range of neuron-specific proteins and was capable of neuronal-like differentiation upon serum deprivation (Qi et al., 1997).

Makeyev and colleagues were able to purify a population of cells of which approximately 100% was transfected with a construct expressing miR-124. To achieve this, the
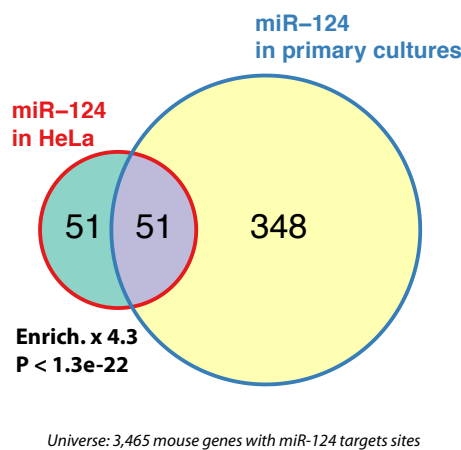
Figure 5.9: A significant intersection of miR-124 targets with those identified by Lim et al.
The Venn diagram shows counts of putative direct targets of miR-124 that were inferred from the transfection experiment of this thesis (*miR-124 in primary cultures*) and from the HeLa transfection experiment (Lim et al., 2005) (*miR-124 in HeLa*). The test universe was 3,465 mouse genes, with 3′UTRs containing one or more 7(2) or 7(1A)-type seed matching site for miR-124. The text shows fold enrichment above what is expected by chance alone and the hypergeometric P-value for the intersection.

pre-miR-124-2 sequence was inserted into an intron of a fluorescent reporter gene, which enabled the authors to FACS-sort transfected CAD cells. Subsequently, gene expression changes in cells transfected with pre-miR-124-2 were assessed using microarrays. Microarray results were deposited in the Gene Expression Omnibus (GEO) database (Sayers et al., 2010), GEO ID GSE8498.

Analysis of the microarray data obtained by Makeyev and colleagues (Methods, section 2.7) revealed that genes downregulated in CAD cells upon transfection with miR-124 expressing plasmid were significantly enriched in genes with miR-124 seed matching sites in their 3′UTRs (Figure 5.10a). Sylamer analysis of seed matching sites distribution showed that this enrichment was independent of length and composition biases (Figure 5.10b).

Analysis of the miR-124 seed matching site distribution enabled identification of putative direct miR-124 targets. The Sylamer enrichment peak approximately coincided with the 0.05 P-value cutoff for differential expression (Figure 5.10b). Of the genes that were downregulated with P-value < 0.05, 641 contained the 7(2) or 7(1A) seed matching sites for miR-124 and comprised a list of candidate direct targets of miR-124 in CAD cells.

A significant intersection (198 genes in common, which is approximately 2.7 more than expected by chance alone, $P < 1.6e - 52$) was observed between putative miR-124 targets in CAD cell line (Makeyev et al., 2007) and miR-124 targets identified in this thesis (Figure 5.11).
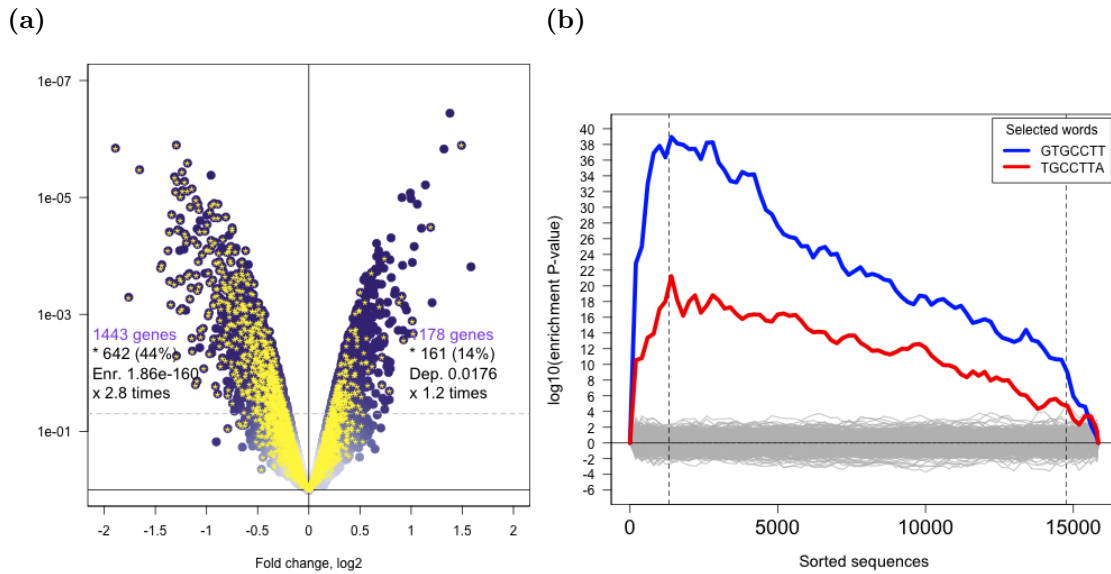
Figure 5.10: Differential expression and shifts in seed matching site distributions induced by miR-124 in CAD cells.
(Figure 5.10a) The x-axis represents log_2 of gene expression fold change of all genes represented on the Affymetrix platform (see Methods, section 2.7) between samples of the CAD cell line transfected with a miR-124 expressing plasmid in comparison the matched mock transfected samples. The y-axis represents P-value of differential expression (log_{10} scale), and the horizontal dashed grey lines show P-value cutoff of 0.05. The yellow asterisks mark genes [encoding transcripts] with 3′UTRs harbouring one or more seed matching sites (7(2) or 7(1A)-types) for miR-124. The text in the two halves of the plot area provides the following information: 1) The total number of genes with differential expression P-value more significant than the cutoff (0.05); 2) The total number (and percentage) of genes with seed matching sites for miR-124; 3) The hypergeometric P-value of enrichment (*Enr.*) or depletion (*Dep.*); 4) Fold enrichment or depletion of genes with the seed matching sites "× *times*" the number that is expected by chance alone. (Figure 5.10b) The x-axis represents sorted 3′UTRs corresponding to all genes represented on the Affymetrix platform (see Methods, section 2.7) ordered by fold change t-statistic for differential expression between samples of the CAD cell line transfected with a miR-124 expressing plasmid in comparison to the matched mock transfected samples. The sequences are **sorted from the most downregulated on the left to the most upregulated on the right**. The y-axis represents the hypergeometric P-values for occurrence biases of 876 nucleotide words complementary to the seed regions (7(2) and 7(1A)-types) of the complete set of 581 distinct mouse miRNAs, according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Positive values on the y-axis corresponds to an enrichment ($+|log_{10}(\text{P-value})|$) and negative values to a depletion ($-|log_{10}(\text{P-value})|$). The vertical dashed lines mark the P-value cutoff (0.05) on both sides of the ranked gene lists. The blue and the red lines show the enrichment profiles of 7(2) and 7(1A)-type seed matching sites for miR-124. The grey lines show the enrichment for the rest of the distinct seed matching sites. The mapping of microarray probes to mRNA transcripts, and transcripts to genes, is described in Methods (section 2.7). The identification of seed regions and parameters of Sylamer (van Dongen et al., 2008) is in Methods (section 2.8). The full description of the Sylamer method is in the Introduction (section 2.8).
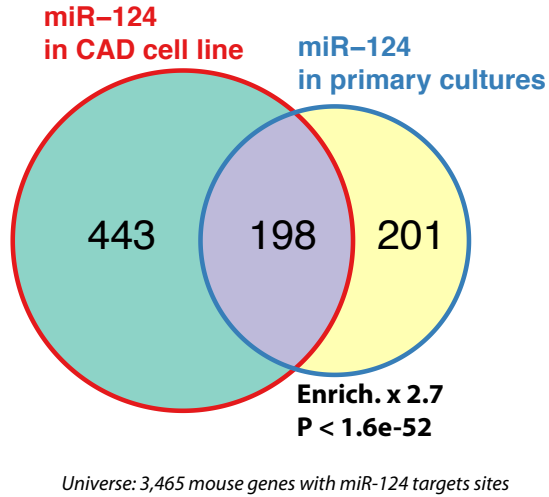
Figure 5.11: A significant intersection of miR-124 targets with those derived from Makeyev et al.
The Venn diagram shows counts of putative direct targets of miR-124 that were inferred from the transfection experiment of this thesis (*miR-124 in primary cultures*) and from the HeLa transfection experiment (Makeyev et al., 2007) (*miR-124 in CAD cell line*). The test universe was 3,465 mouse genes, with 3′UTRs containing one or more 7(2) or 7(1A)-type seed matching site for miR-124. The text shows fold enrichment above what is expected by chance alone and the hypergeometric P-value for the intersection.
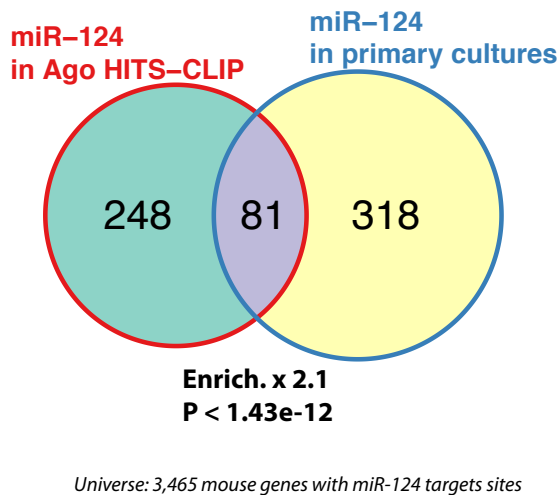

Figure 5.12: A significant intersection of miR-124 targets with those identified by Chi et al.
The Venn diagram shows counts of putative direct targets of miR-124 that were inferred from the transfection experiment of this thesis (*miR-124 in primary cultures*) and from the HITS-CLIP experiment (Chi et al., 2009) (*miR-123 in Ago HITS-CLIP*). The test universe was 3,465 mouse genes, with 3′UTRs containing one or more 7(2) or 7(1A)-type seed matching site for miR-124. The text shows fold enrichment above what is expected by chance alone and the hypergeometric P-value for the intersection.

5.3.3   Comparison to targets identified by Chi *et al.*

The basis of the two described above studies (Lim et al., 2005; Makeyev et al., 2007) was the overexpression of miR-124 through chemical transfection of nucleic acids, which led to an excess of miR-124. Therefore, even though targets identified in this thesis were in good agreement with these two experiments, the question remained whether the thesis targets were relevant to miR-124 activity outside cell cultures and transfection paradigms.

Lists of targets of miR-124 and 19 other highly expressed in brain miRNAs were recently identified using the HITS-CLIP method in an innate P13 mouse neocortex (Chi et al., 2009). The HITS-CLIP did not involve transfections, but was a combination of UV cross-linking of Ago proteins (a key component of miRNA silencing complex, see Introduction, section 1.1.2) and nucleic acids, and immunoprecipitation of bound RNA followed by a high-throughput RNA sequencing. The sequences, which also harboured the seed matching sites for one of the 20 miRNAs, corresponded to putative direct targets of these miRNAs[1] (the *in vivo* targets).

A significant overlap (approximately 2.1 times bigger than was expected by chance alone, $P < 1.44e-12$) was detected between HITS-CLIP miR-124 putative direct targets and miR-124 targets identified in this thesis (Figure 5.12). This suggested that a significant proportion of miR-124 targets identified in this thesis were direct miR-124 *in vivo* targets.

Summary of section 5.3

This chapter described the identification of putative direct miRNA targets in primary neuronal cultures of six different miRNAs: miR-124, miR-434-3p, miR-25, miR-143, miR-145 and of cel-miR-67. The targeting repertoire of miR-124 was previously studied, thus it was possible to compare published results to the list of miR-124 targets identified in this thesis. A good agreement was found in comparisons between the thesis targets and three published studies (Lim et al., 2005; Makeyev et al., 2007; Chi et al., 2009). This indicated that methods of this work to identify putative direct miRNA targets were valid and generated reproducible results.

---

[1]The targets identified with the HITS-CLIP method were available for download from the authors' website http://ago.rockefeller.edu/.