Chapter 6

Analysis of miRNA function in neurons

The goal of this thesis on a very basic level is to determine the function of miRNAs in neurons. Are neuronal miRNAs important? What is their repertoire of targets? Do these miRNAs share target pools? What can be learned about the overall function of these miRNAs from studying detected targets? Within this chapter I will attempt to answer these questions using the data derived during my research and previously published literature.

Previously, individual miRNAs were shown to regulate expression from dozens to hundreds of genes (Stark et al., 2003; Enright et al., 2003; Farh et al., 2005; Lim et al., 2005; Giraldez et al., 2006; Baek et al., 2008; Selbach et al., 2008), however functions of miRNAs are usually viewed through a prism of a handful of validated targets. Additionally, roles of miRNAs in the differentiation of neural progenitors are well established (Introduction, section 1.1.3). However after differentiation, the functions of a majority of miRNAs are poorly understood. Chapter 5 described the identification of hundreds of putative direct targets for six miRNAs (**miR-124**, **miR-434-3p**, **miR-143**, **miR-145**, **miR-25** and **cel-miR-67**) in committed primary neuronal cultures. Two of these miRNAs, miR-124 and miR-434-3p, were either highly expressed or upregulated in the development of primary forebrain cultures (Chapter 3, section 3.3). Therefore, they were expected to have endogenous functions in neurons. On the other hand, other miRNAs were either downregulated in the development of cultures, or absent from the mouse genome altogether (Chapter 3, section 3.3). Therefore, no specifically neuronal function was expected to be associated with these miRNAs in primary neuronal cultures. In this chapter I will compare and analyse lists of targets of these six miRNAs, and attempt to identify func-

tions associated with miRNA mediated regulation as a whole and functions that may be associated specifically with neuronal miRNAs.

The analysis of intersections between lists of targets, together with the analysis of enrichment of functionally annotated gene categories (GO and KEGG categories (terms) (Ashburner et al., 2000; Kanehisa et al., 2008)) in the lists of miRNA targets, is described in the first section (section 6.1) of this chapter. In the second section (section 6.2), I will explore the connection between functions of neuronal miRNAs and stress responses and highlight the importance of these processes for neuronal biology.

6.1 Characterisation of identified miRNA targets

6.1.1 Significant intersections between targets of different miRNAs

It was reasoned that between the lists of targets of six miRNAs intersections that were bigger than expected, would indicate related targeting repertoires, while smaller than expected intersections would suggest distinct repertoires. Assessment of the significance of an intersection between elements of two lists can be done using the hypergeometric test, where the lists are viewed as samples from a larger set of elements, which is referred to as the test universe. Selection of an appropriate test universe is critical for this type of analysis: an unsuitably large universe artificially enhances, while a restricted universe reduces the significance of intersections. As the targets of miR-124 were in good agreement with previously published results (Chapter 5, section 5.2.1), data derived from the miR-124 experiment (the 6DIV bidirectional transfection, see Chapter 5, section 5.2.1) was judged to be reliable, and 10,821 genes detected[1] in this experiment were used as the test universe.

Using the single gene universe from the miR-124 experiment, the hypergeometric test for enrichment showed significant intersections between the majority of the lists of miRNA targets (Figure 6.1). For example, targets of miR-124, miR-434-3p and miR-145 intersected significantly with the lists of targets of all of the other five miRNAs (enrichment P-values varied from 0.0383 to $6.41e-10$). Interestingly, the intersection of the targets of miR-124, miR-434-3p and miR-145 with the targets of cel-miR-67, which is not expressed in the mouse, was two or more times bigger than expected by chance alone (enrichment P-values were 0.00054, 0.0057 and $7.04e-07$). Overall, a cross comparison of the targets of

[1] Using the standard Illumina detection call threshold $P < 0.01$, see Methods (section 2.7)

the six miRNAs revealed 12 significant intersections with $P < 0.05$ (Figure 6.1). The observation of multiple significant intersections could not be explained by the increased false discovery rate associated with multiple testing, because nine results remained significant ($P < 0.05$) even with the strictest adjustment method (Bonferoni correction).

Significant intersections were also identified between the putative direct targets identified in this thesis, and miRNA targets identified elsewhere. For example, miR-124 targets derived from the experiment conducted by Makeyev and colleagues, which overexpressed miR-124 in the CAD cell line[2] (Makeyev et al., 2007), had a significant intersection with transfection targets of five out of six miRNAs from this thesis (Figure 6.1). Importantly, the intersection of the CAD miR-124 targets with targets of the non-mouse cel-miR-67 was also significant ($P < 0.0006$, see Figure 6.1). The latter indicated that the significant intersections where not only a feature of neuronal miRNAs, but also of miRNA mediated gene expression regulation as such.

6.1.2 Explaining the intersection: A hypothesis of a pool of transcripts primed for miRNA mediated regulation

One explanation for the multiple significant intersections could be artefacts in detection of differential expression. For example, due to technical biases, differential expression might have been detectable for only a small, common subset of transcripts in each of the transfection experiments. In such case, intersections between the lists of miRNA targets would be bigger than expected by chance alone, because identification of targets was itself reliant on detection of differential expression. An alternative explanation of significant intersections would be the existence of a pool of transcripts that is primed for miRNA mediated regulation. In other words, a higher than expected intersections could be explained by the fact that not all transcripts that were expressed and contained the seed matching sites were equally likely to be downregulated upon overexpression of miRNAs (I will sometimes refer to this proposition as a "pool of targets hypothesis"). To distinguish between these two explanations (i.e. artefacts versus the pool of targets hypothesis), targets of miRNAs obtained in this thesis were compared to targets identified with an experiment, which did not rely on microarrays or on detection of differential expression by other means.

[2]See Chapter 5 (section 5.3.2) for the description of derivation of putative direct miR-124 targets from the CAD cell experiment (Makeyev et al., 2007).

One such experiment is the Ago HITS-CLIP experiment conducted by Chi and colleagues (Chapter 5, section 5.3.3) (Chi et al., 2009). In this experiment Chi and colleagues identified putative direct targets of miR-124 and of 19 other miRNAs, which were all abundant in the P13 mouse neocortex (Chi et al., 2009). The putative miRNA targets were defined as transcripts that were bound by Ago, and which had a hexamer seed matching site near the Ago binding site for one of the 20 miRNAs. I will refer to the combined set of these putative targets as the Ago HITS-CLIP set[1].

There was a significant intersection between the Ago HITS-CLIP set and transfection targets of five out of six miRNAs identified in this thesis (Figure 6.1). This result was unexpected, because only one miRNA, miR-124, was investigated both in this thesis and in the HITS-CLIP experiment. Therefore the intersection could have been explained by the bias in the seed based target identification only for miR-124 (only transcripts with miR-124 seed matching sites can be defined as targets, which restricts possible selection of genes from the universe[2]). None of the other six miRNAs from this thesis shared the seed region (either hexamer or heptamer, see Introduction, section 1.2.1) with miRNAs from the HITS-CLIP study. Nevertheless, multiple significant intersections with the Ago HITS-CLIP were observed (Figure 6.1). Moreover, a significant intersection was identified between the Ago HITS-CLIP set and targets of cel-miR-67. This intersection was 87 genes (out of the 350 targets of cel-miR-67 present in the test universe), which was approximately 1.7 times more than expected by chance alone ($P < 7.35e-07$, see Figure 6.1). This was, perhaps, the strongest support for the pool of targets hypothesis, because targets of cel-miR-67 can be viewed as a sample of transcripts that could be regulated by miRNAs in primary neurons unbiased by previous evolutionary selection.

These observations implied that regardless of the method used for detection of putative direct miRNA targets, miRNAs in the neuronal systems (i.e. the P13 neocortex in the HITS-CLIP study, and primary forebrain cultures in this thesis) appeared to target a common subset of genes. Therefore, the pool of targets hypothesis was considered to be a likely explanation of the observed intersections between the lists of targets of different miRNAs.

[1]In experiments performed in this thesis, on average $\approx$ 1,500 genes encoding transcripts of the Ago HITS-CLIP set were detected as expressed (using the standard Illumina detection call threshold $P < 0.01$, see Methods, section 2.7).

[2]It should be noted that even for miR-124 the observed significant intersection was unlikely to had been explained by the bias in the seed based definition of targets. In Chapter 5 (section 5.3.3), I showed that the intersection between miR-124 targets identified in the thesis and in the HITS-CLIP experiment was significant ($P < 1.43e-12$) within the universe of only the genes that were encoding transcripts with the miR-124 seed matchings sites.

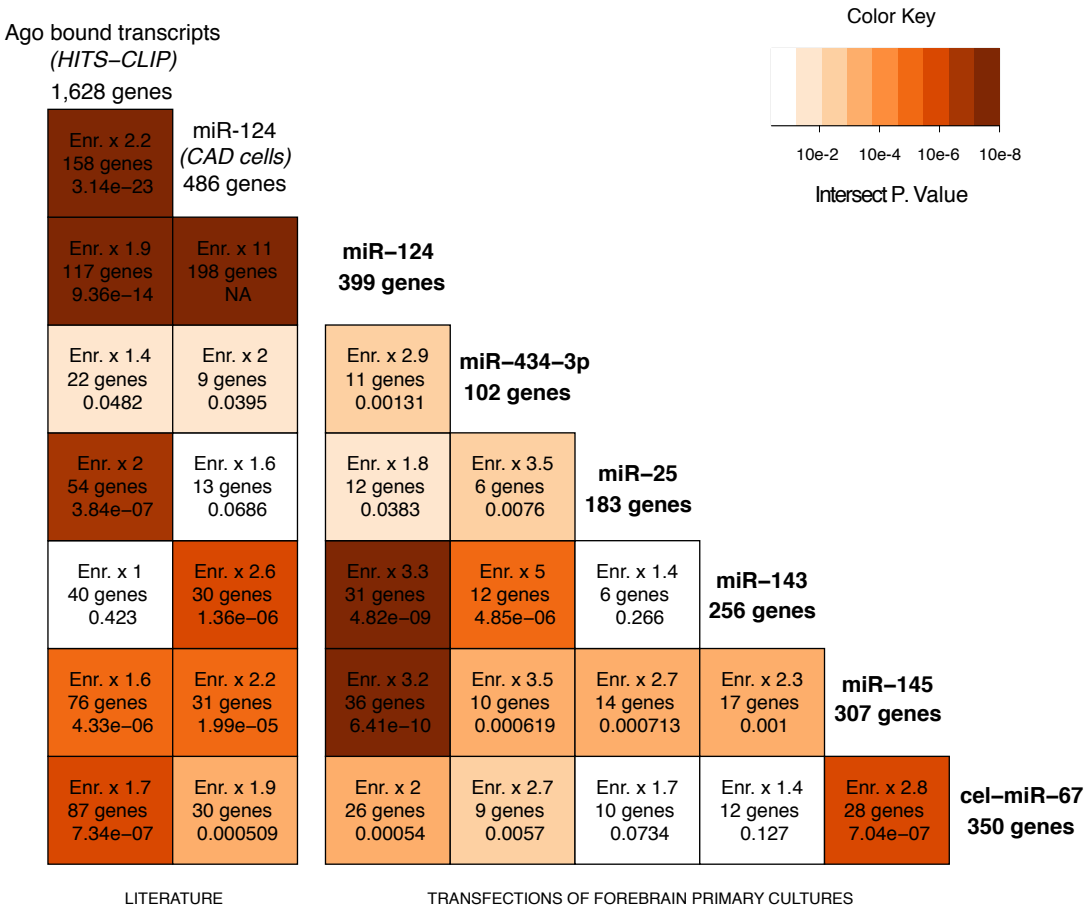

Figure 6.1: Intersections of the lists of putative direct miRNA targets.
Names of the genelists together with the total number of genes in the lists, belonging to the test universe is given on the sides of the boxes. The text inside of the boxes provides information about the intersection of the list that correspond to top and right sides of the boxes: 1) Fold enrichment (*Enr.*) the number of times (×) more than expected by chance alone; 2) The number of genes in the intersection between the two lists; 3) The hypergeometric P-value of the enrichment. The color of boxes corresponds to the hypergeometric P-value according to the color-scheme of the *Color Key*. The test universe (*Gene Universe*) for all tests was the complete set of 10,821 genes detected in the miR-124 bidirectional transfection experiment (6DIV, see Chapter 5, section 5.2.1).

6.1.3 Context-dependent nature of the pool: Over 20% of targets were induced by the transfection procedure itself

Having observed significant intersections between the lists of putative direct targets of different miRNAs, I proceeded to functional characterisation of these lists. If targets of more than one miRNA were associated with the same function, it would suggest existence of global functions of miRNA mediated regulation. To associate functions with the lists of miRNA targets, I evaluated enrichment of KEGG pathways (Kanehisa et al., 2008) in all lists (Methods, section 2.10).

To make KEGG enrichment analysis fully comprehensive, test universes were defined individually for each of the lists of targets. The universes were defined as all genes detectably expressed in the experiment that lead to generation of each list[1]. By pairing the universe with the corresponding transfection experiments ensured that all targets would be within the universe to be included in the analysis. Enrichment of KEGG pathways was assessed in the lists with the hypergeometric test (Methods, section 2.10), results of which depend on sizes of the lists tested (i.e. it is harder to obtain significant P-values for smaller lists of targets). Therefore, to make a comparison between the lists, for each of the lists an arbitrary cutoff was set at the 25 most enriched KEGG pathways, and all 25 pathways were considered irrespective of the enrichment P-value.

Analysis of KEGG term enrichment in the Ago HITS-CLIP target set (i.e. *in vivo* miRNA targets in P13 mouse neocortex), revealed that the set was enriched in the genes that were unlikely to be expressed constitutively. For example, three out of four most enriched KEGG pathways in the Ago HITS-CLIP set were "Long-term potentiation", "Regulation of actin cytoskeleton" and "Axon guidance" (Supplementary Data, Table A.22). It is possible that the 20 highly expressed in neurons miRNAs, which targets comprised the Ago HITS-CLIP set, acted as buffers of expression of the genes from the aforementioned pathways at the times when these pathways were induced (e.g. during neuronal plasticity). Such hypothesis agrees with previous reports of miR-124 (Rajasethupathy et al., 2009) and miR-134 (Gao et al., 2010) to reduce the plasticity, and a report of miR-134 to reduce the size of the synapse (Schratt et al., 2006) (functions of miRNAs as inhibitors of neuronal plasticity is described in the Introduction, section 1.1.3). This

[1]The exception to this was the Ago HITS-CLIP set, which was produced in the external experiment and where the full set of genes expressed was not known. Instead, the universe of genes from the miR-124 transfection experiment was used for testing KEGG enrichment in the Ago HITS-CLIP set. For all other experiments, all genes detected, using the standard Illumina detection call threshold $P < 0.01$ (Methods, section 2.7) were used as gene universes.

reasoning suggested that miRNAs exogenously added in the transfections could have also inhibited the genes that were induced during the experiments.

To test the hypothesis that the exogenously added miRNAs (i.e. transfected miRNAs) inhibited inducible pathways, I first identified genes that were induced during the transfection experiments. To achieve this, the expression profiles of mock transfected cultures (i.e. cultures treated with the transfection reagent, but without RNA added to it) were compared to the expression profiles of matched untransfected cultures (Methods, section 2.5). This comparison revealed that 1,293 genes were upregulated ($P < 0.05$) by the treatment with the transfection reagent (these genes will be referred to as the "induced by the transfection reagent" set). Analysis of KEGG pathway enrichment in this set[2] showed that the genes induced by the transfection reagent were enriched not in the pathways involved in normal neuronal function, but in pathways related to diseases and stresses. For example, the most highly enriched KEGG pathway was "p53 signalling pathway", while the pathway "Metabolism of xenobiotics by cytochrome P450" was also among the top 25 most highly enriched pathways. Additionally, of the 25 most enriched terms in the induced by the transfection reagent set, 12 were related to cancer or other diseases (Supplementary Data, Table A.21). Enrichment of stress and disease related pathways in the induced by the transfection set suggested that a significant fraction the induced genes was involved in offsetting the adverse effects of the transfection reagent.

Next, I tested if transfected miRNAs inhibited the induced by the transfection reagent genes. Using genes that were detected in the mock transfection experiment (using the standard Illumina detection call $P < 0.01$, see Methods, section 2.7) as the gene universe, I found that the intersections of the induced by the transfection reagent set and of the targets of five out six miRNAs was statistically significant (Figure 6.2). Interestingly, targets of the two neuronal miRNAs, miR-124 and miR-434-3p, were the most enriched in the genes induced by transfection reagent (3.3 times more than expected by chance alone for miR-124 ($P < 2.35e-41$) and 3.8 time more for miR-434-3p ($P < 3.56e-14$)). When the intersections were assessed outside of the gene universe, expression of 34.3% and 37.5% of miR-124 and miR-434-3p targets in total was found to have been induced by the transfection reagent. Also, the "p53 signalling pathway", which was the most highly enriched pathway in the genes induced by transfection (Supplementary Data, Table A.21) was among top 25 most enriched pathways in the targets of both miR-124 and miR-434-3p (Supplementary Data, Table A.43). If targets of all six miRNAs were combined,

[2]Genes detected using the standard Illumina detection call threshold $P < 0.01$ (Methods, section 2.7) in the mock transfection experiment were used as a gene universe for these tests.

producing a unique list of 1,512 genes, then 22.2% (337 genes) were identified as induced by the transfected reagent.

Furthermore, 14 out of the 25 most enriched KEGG pathways in the induced by the transfection reagent set were also in the top 25 most enriched pathways in the targets of one or more miRNAs (Supplementary Data, Table A.15 to A.20). Figure 6.3 shows this recurrent enrichment of KEGG pathways in targets of a selection of three miRNAs[3] (miR-124, miR-434-3p and cel-miR-67), and genes induced by the transfection reagent. Multiple disease related pathways (e.g. "Pathways in cancer") were found to be induced by the transfection reagent and targeted by several miRNAs. In fact, the KEGG term "Pathways in cancer" was among 25 most enriched terms in the targets of five miRNAs: miR-124, miR-434-3p, miR-25, miR-143 and cel-miR-67 (Figure 6.3 and Table 6.1). Identification of this and other cancer-related KEGG pathways as enriched in miRNA targets was in agreement with a large body of evidence that showed significance of miRNA mediated regulation in development of various types of tumors (Volinia et al., 2010).

In summary, results presented in this section showed that the significant intersection of lists of targets of different miRNAs (see section 6.1.2) could be due to different miRNAs having converged on inhibition of a common set of genes. According to the proposed pool of targets hypothesis, this set of genes encoded a pool of transcripts primed for miRNA mediated regulation. In transfection experiments conducted in this thesis, I found that genes induced by the transfection reagent contributed significantly to this pool of primed targets. Functional characterisation of miRNA targets and genes induced by transfection, showed that multiple transfected miRNAs inhibited pathways that were induced by the transfection reagent (i.e. many disease and stress associated pathways). Targets of the two neuronal miRNAs, miR-124 and miR-434-3p, were the most enriched in the genes induced by the transfection reagent. On the contrary, targets of an oncogenic miRNA, miR-25, and of a non-mouse miRNA, cel-miR-67, were least enriched in the genes induced by the transfection reagent. These observations suggested that targeting repertoire of neuronal miRNAs could have specifically evolved to buffer the expression of genes that can be induced by adverse treatments of neurons (such as the treatment by the transfection reagent).

These collected observations suggest that miRNA mediated regulation as a whole converges on inhibition of genes that are upregulated in the system, i.e. miRNAs can act

[3]These three miRNAs were selected to demonstrate that pathways induced by the transfection reagent contributed to targets of both mouse neuronal miRNAs (miR-124 and miR-434-3p), and a non-mouse miRNA (cel-miR-67). Recurrent enrichment of KEGG pathways in targets of these unrelated miRNAs supports the hypothesis of a pool of transcripts primed for miRNA mediated regulation.

as a buffer against deviation of the transcriptome from equilibrium of the differentiated state. In accordance with this proposition, in experiments that were performed in this thesis, transfected miRNAs acted to reduce the perturbation caused by the transfection reagent. In this function as a buffer of transcriptional changes, miRNAs can be particularly important for neurons, because these cells constantly receive a flux of stimuli, each of which, potentially, can alter gene expression and make the transcriptome to deviate from the *status quo*. The latter is in agreement with published reports of exogenous miRNAs reducing the plasticity of neurons (Gao et al., 2010; Rajasethupathy et al., 2009), while removal of the endogenous miRNA increases plasticity (Gao et al., 2010; Konopka et al., 2010).

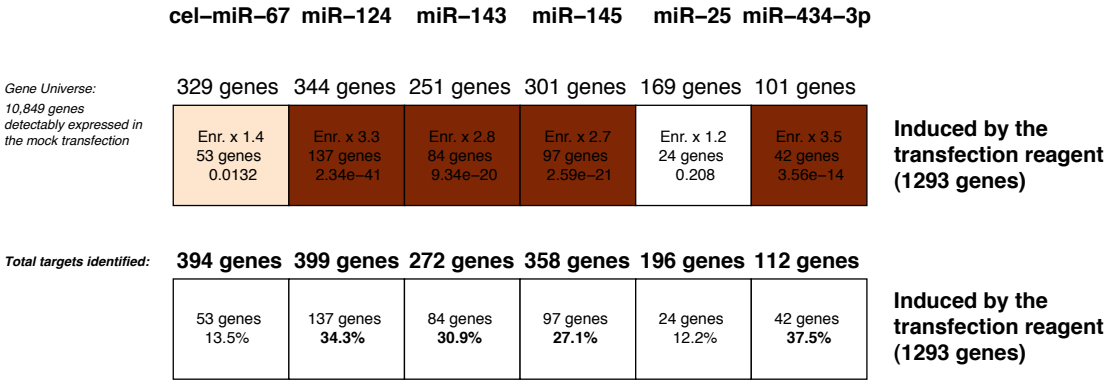

Figure 6.2: Intersection of genes induced by the transfection reagent and miRNA targets. Each of the boxes shows intersection between two gene lists: 1) A list of miRNA targets (the corresponding names of miRNAs are labelling the tops of the boxes); 2) The list of genes induced by the genes induced by the transfection reagent (labelling the right sides of the boxes). The upper row of boxes displays information about the intersections between the lists that were limited by the gene universe (10,849 genes detected with the standard Illumina detection call $P < 0.01$ (Methods, section 2.7) in the mock transfection experiment). Information about the intersections is presented in the same way (and coloring is according to the same color-scheme) as in Figure 6.1. The bottom row of clear boxes shows intersections of complete lists (i.e. not restricted by the universe of genes detected in the mock transfection experiment). The bottom row of boxes provides the following information: 1) The total number of genes in the intersections; 2) The percent of the intersections in the lists of miRNA targets.

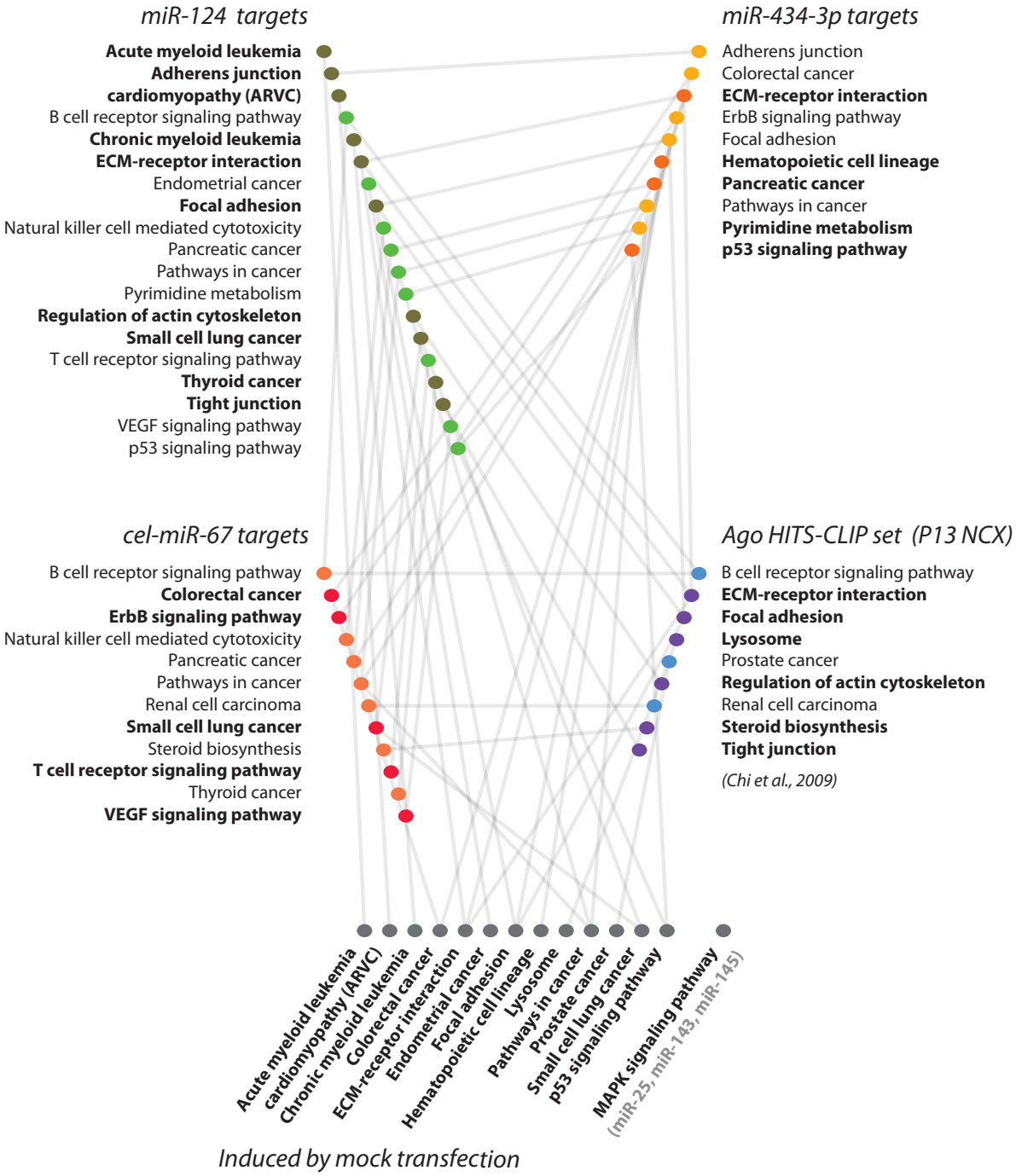

Figure 6.3: Recurrence in top 25 most enriched KEGG terms.
The text in italic is the gene lists in which KEGG enrichment was assessed. The text in regular font shows the lists of KEGG pathways that were among the top 25 most enriched pathways in more than one of these gene lists (with the exception of "MAPK signalling pathway", which was within the top 25 most enriched pathways in the targets of miR-143, miR-145 and miR-25). The pathways in bold, were enriched with the P-value < 0.05 (Methods, section 2.10).
The pathways that are enriched in multiple gene lists are connected with the grey lines. The full list of miRNA targets from within the following pathways is in the Supplementary Data: "ECM-receptor interaction" (Table A.39), "Gap junction" (Table A.40), "ErbB signaling pathway" (Table A.41), "Tight junction" (Table A.42), "p53 signaling pathway" (Table A.43), "Regulation of actin cytoskeleton" (Table A.44), "Focal adhesion" (Table A.45), "MAPK signaling pathway" (Table A.46), "VEGF signaling pathway" (Table A.47), "Toll-like receptor signaling pathway" (Table A.48).

Symbol	Description
cel-miR-67 (**Ranked 17, P ≈ 0.101**)	
Appl1	adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 1
Bcl2	B-cell leukemia/lymphoma 2
Cycs	cytochrome c, somatic
Fzd3	frizzled homolog 3 (Drosophila)
Kras	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
Pias3	protein inhibitor of activated STAT 3
Pik3cb	phosphatidylinositol 3-kinase, catalytic, beta polypeptide
Ralb	v-ral simian leukemia viral oncogene homolog B (ras related)
Rarb	retinoic acid receptor, beta
Rb1	retinoblastoma 1
Tpm3	tropomyosin 3, gamma
miR-124 (**Ranked 22, P ≈ 0.11**)	
Ccnd1	cyclin D1
Col4a1	collagen, type IV, alpha 1
Erbb2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived [...]
Fadd	Fas (TNFRSF6)-associated via death domain
Itgb1	integrin beta 1 (fibronectin receptor beta)
Lamc1	laminin, gamma 1
Nras	neuroblastoma ras oncogene
Rela	v-rel reticuloendotheliosis viral oncogene homolog A (avian)
Skp2	S-phase kinase-associated protein 2 (p45)
Smad3	MAD homolog 3 (Drosophila)
Stat3	signal transducer and activator of transcription 3
Tcf7l1	transcription factor 7-like 1 (T-cell specific, HMG box)
Traf3	TNF receptor-associated factor 3
miR-143 (**Ranked 23, P ≈ 0.161**)	
Birc5	baculoviral IAP repeat-containing 5
Egfr	epidermal growth factor receptor
Fadd	Fas (TNFRSF6)-associated via death domain
Mmp2	matrix metallopeptidase 2
Pdgfb	platelet derived growth factor, B polypeptide
Pdgfra	platelet derived growth factor receptor, alpha polypeptide
Smad2	MAD homolog 2 (Drosophila)
Smo	smoothened homolog (Drosophila)
miR-145 (Ranked 26, P ≈ 0.222)	
Birc5	baculoviral IAP repeat-containing 5
Cycs	cytochrome c, somatic
Gli3	GLI-Kruppel family member GLI3
Ikbkg	inhibitor of kappaB kinase gamma
Nras	neuroblastoma ras oncogene
Pdgfra	platelet derived growth factor receptor, alpha polypeptide
Ptch1	patched homolog 1
Traf6	TNF receptor-associated factor 6
Wnt7b	wingless-related MMTV integration site 7B
miR-25 (**Ranked 24, P ≈ 0.111**)	
Fgf10	fibroblast growth factor 10
Fgf12	fibroblast growth factor 12
Igf1r	insulin-like growth factor I receptor
Mapk8	mitogen-activated protein kinase 8
Pik3cb	phosphatidylinositol 3-kinase, catalytic, beta polypeptide
Pik3r2	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 2 (p85 beta)
Wnt5a	wingless-related MMTV integration site 5A
miR-434-3p (**Ranked 20, P ≈ 0.135**)	
Birc5	baculoviral IAP repeat-containing 5
Egfr	epidermal growth factor receptor
Fgf13	fibroblast growth factor 13
Stat3	signal transducer and activator of transcription 3
Tgfbr2	transforming growth factor, beta receptor II

Table 6.1: A list of the miRNA targets found within the "Pathways in cancer" KEGG pathway.
The text in parenthesis shows the rank of the enrichment of the "Pathways in cancer" among all the KEGG pathways and the P-value of that enrichment. The text is in bold if the enrichment was ranked within the top 25 most enriched pathways.

6.1.4 Context dependent nature of published miR-124 targets

The HITS-CLIP experiment (Chi et al., 2009) and the transfection experiments in this thesis were performed in related neuronal systems (P13 mouse neocortex and primary forebrain cultures). Therefore it was expected that similarities would be found between miRNA targets identified in both of these studies. Indeed, miR-124 targets identified in this thesis had a significant inetersection with the Ago HITS-CLIP as a whole (Figure 6.1), and specifically with the miR-124 HITS-CLIP targets (which were a part of the Ago HITS-CLIP set, see Chapter 5, section 5.3). Moreover, multiple KEGG pathways, which were overrepresented in the targets identified by miRNA transfection experiments and in genes induced by the transfection reagent alone, were also overrepresented in the Ago HITS-CLIP set (Figure 6.3).

Despite these similarities between the miRNA targets identified with transfections of primary cultures and the targets identified with the HITS-CLIP method, a line of evidence showed that transfection targets and the HITS-CLIP targets were focused around different aspects of miRNA function. In section 6.1.3, I described that in transfection experiments in this thesis project, a significant proportion of targets of five out of six miRNAs (including those of miR-124) was induced by the treatment with the transfection reagent (Figure 6.2). Interestingly, targets of miR-124 that were identified in published transfections of HeLa (Lim et al., 2005) and CAD cell lines (Makeyev et al., 2007) also had significant intersections with the set of genes induced by the transfection reagent (Figure 6.4). However, the HITS-CLIP miR-124 targets had approximately four orders of magnitude less significant P-value for the intersection, while the intersection of the whole Ago HITS-CLIP set and the genes induced by the transfectecion reagent was not not significant (Figure 6.4). This difference between the two types of targets (transfection versus HITS-CLIP) is important, because in this thesis the transfection targets of neuronal miRNAs (miR-124 and miR-434-3p) were comprised to a large and significant extent of the genes induced by the transfection reagent with an enrichment 3 times bigger than expected by chance alone, see Figure 6.2. Therefore, the transfection targets can be viewed as focusing on miRNAs function in disease and stress, because the "induced by the transfection reagent" set was enriched in disease and stress related pathways (Supplementary Data, Table A.21). At the same time, the HITS-CLIP targets are likely to focus on miRNA function in normal neurons, because the Ago HITS-CLIP set was highly enriched in several specifically neuronal pathways (Supplementary Data, Table A.22), and not enriched in the induced by the transfection reagent set itself (Figure 6.4).

These results support the proposition of a context dependent function of miRNAs as inhibitors of the inducible genes – in different transfection experiments miRNAs converge on inhibition of the genes induced in the transfection experiment, and it is not the case in the transfection-free experiment. This conclusion has direct implications for future studies aiming at identifying targets of miRNAs, as it is important to realise that miRNAs can act on different genes in different experimental contexts.

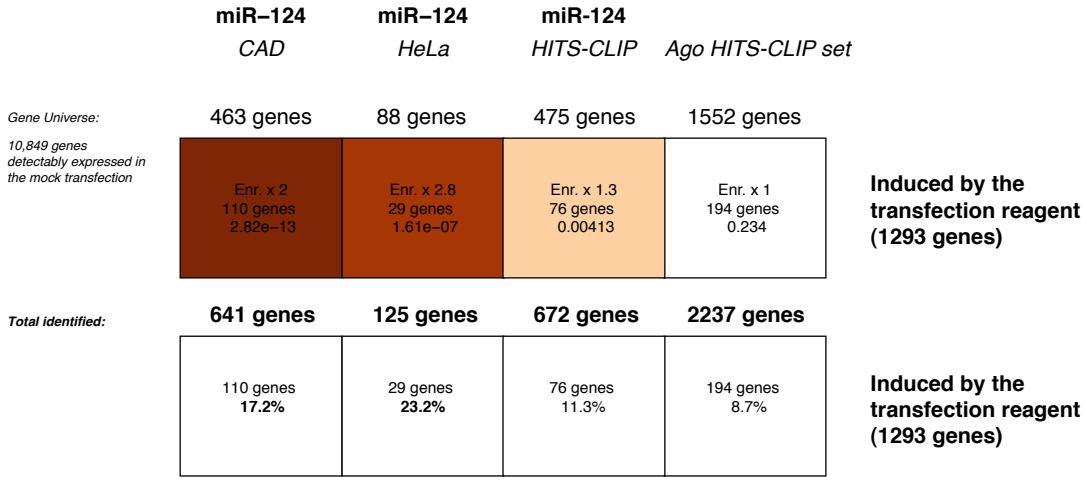

Figure 6.4: Intersection of published miR-124 targets and genes induced by the transfection reagent.
Each of the boxes shows the intersection between two gene lists: 1) A list of published miRNA targets (miR-124 targets or the complete list of the Ago HITS-CLIP targets of 20 most highly expressed miRNAs, see text). The corresponding names of miRNAs are labelling the tops of the boxes; 2) The list of genes induced by the mock transfection (labelling the right sides of the boxes). The upper row of boxes displays information about the intersections between the lists that were limited by the gene universe (10,849 genes detected with the standard Illumina detection call $P < 0.01$ (Methods, section 2.7) in the mock transfection experiment). Information about the intersections is presented in the same way (and colored according to the same color-scheme) as in Figure 6.1. The bottom row of clear boxes shows intersections of complete lists (i.e. not restricted by the universe of genes detected in the mock transfection experiment). The bottom row of boxes provides the following information: 1) The total number of genes in the intersections; 2) The percent of the intersections in the lists of published miRNA targets.

6.1.5 Recurrent enrichment of GO terms was in agreement with the pool of targets hypothesis

The hypothesis that a pool of transcripts was primed for miRNA mediated regulation was formulated in section 6.1.2, and it was based on the observation of significant intersections between transfection targets identified for unrelated miRNAs. In agreement with this hypothesis, several KEGG pathways (i.e. cell adhesion, cell signaling and stress related pathways, see sections 6.1.3) were recurrently enriched in the lists of targets of different miRNAs (Figure 6.3). In this section, the pool of targets hypothesis is further supported by the observation that several large Gene Ontology (GO) terms (Ashburner et al., 2000) were recurrently enriched in targets of unrelated miRNAs (Methods, section 2.10).

There are currently tens of thousands of annotated GO terms and the relationship between them is complicated (http://www.geneontology.org/). Therefore, to ease interpretation of GO enrichment across the lists of miRNA targets, the survey of GO enrichment was based on a selection of several relatively large representative GO terms (more that 100 genes in each, "Cellular compartment" and "Biological process" types of GO terms (Ashburner et al., 2000)). This selection of GO terms was based on 40 most enriched terms in the iduced by the transfection reagent and the Ago HITS-CLIP sets. Selection of the representative terms was made from these two sets, because the mock transfection and the HITS-CLIP experiments uncovered aspects of miRNA function in two different contexts: in neurons under the transfection stress and in neurons in the normal state (section 6.1.3). The selected terms are listed in the legend of Figure 6.5 and the complete list of the 40 most highly enriched GO terms in the induced by the transfection reagent, the Ago HITS-CLIP sets and also in the targets of the six miRNAs, are shown in Supplementary Data, Tables A.23 to A.38.

As expected based on the pool of targets hypothesis, most of the terms representative of GO enrichment in the induced by the transfection reagent and the Ago HITS-CLIP sets were found to be enriched in targets of one or more of the six miRNAs (Figure 6.5). For example, the most consistently enriched GO term was "signalling" (enriched in targets of four out of six miRNAs, and also in the mock transfection and the Ago HITS-CLIP sets). This was in agreement with results of KEGG enrichment analysis, where various signaling pathways were frequently observed among the 25 most enriched pathways in targets of different miRNAs (Figure 6.3). Several other selected GO terms were also enriched in targets of more than one miRNAs. Of "Cellular compartment" terms, these were "extracellular region", "cytoplasm", "endoplasmic reticulum", "plasma

membrane", "Golgi apparatus" and "cytoskeleton" (Figure 6.5a). Of the "Biological process" terms, these were "multicellular organism development", "cell adhesion", "cell cycle", "transport" and "cell communication" (Figure 6.5b).

Analyses presented in this section supports the pool of targets hypothesis, by showing that a representative selection of GO termes was recurrently enriched in targets of different miRNAs. This result was in agreement with observations of the recurrent enrichment of KEGG pathways (Figure 6.3), the enrichment of lists of targets in a set of genes induced by the transfection reagent (Figure 6.2), and multiple significant intersections between the target lists themselves (Figure 6.1). Additionally, the analysis of GO enrichment further described functions that were associated with miRNA targets in primary neurons. These functions were inhibition of genes from categories "signalling", "transport", "cytoskeleton" and etc. The nature of these categories suggests that a significant proportion of miRNA targets is likely not to be constitutively expressed, but to be inducible in certain contexts. Therefore, the results presented in this section are consistent with the proposition of miRNAs to buffer the expression of inducible genes.

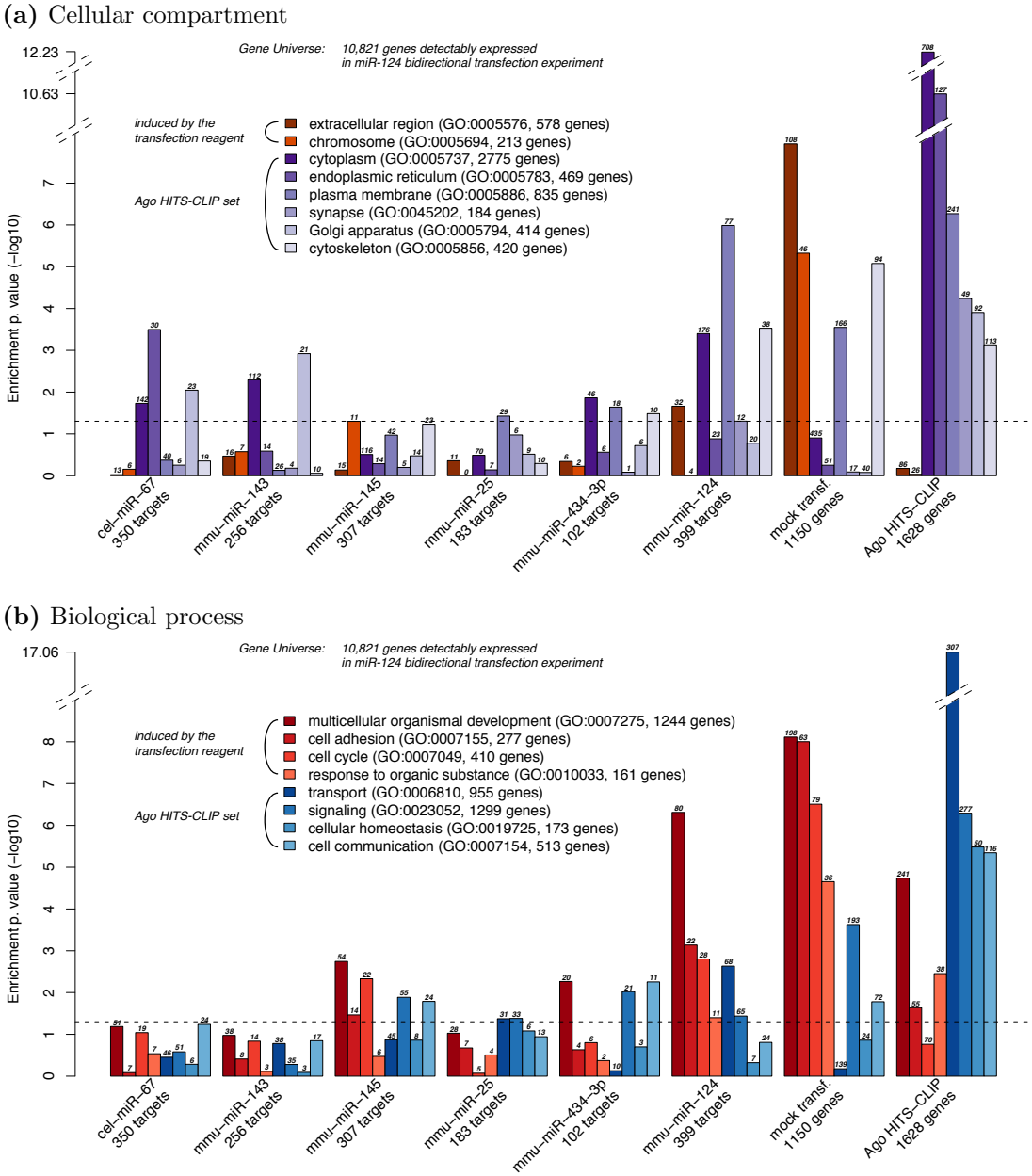

Figure 6.5: Enrichment of Gene Ontology (GO) terms in miRNA targets.
The y-axes show the enrichment P-value of the GO terms (log_{10}) that were selected from the top 40 most enriched GO terms in the genes induced by the transfection reagent of in the Ago HITS-CLIP set. The selected GO terms are listed in the plot areas. The colors that are attributed to the selected GO terms correspond to the colors of the bars for each of the gene lists (as specified on the x-axes). 6.5a – the enrichment of GO terms of the "Cellular compartment" type; 6.5b – the enrichment of GO terms of the "Biological process" type. The total number of genes from each of the selected GO terms that were identified in the corresponding list of targets is given at the top of the bars.

6.1.6 Enrichment of GO terms highlighted the importance of miR-124 and miR-434-3p

Significant intersections between the targets of unrelated miRNAs were described in previous sections. This overlap could be interpreted as an indication that all miRNAs were equally important functionally for neurons. In this section analysis of GO enrichment was used to explore the functional importance of the six transfected miRNAs, which leads to a conclusion that the two neuronal miRNAs (miR-124 and miR-434-3p) were more functionally important than other miRNAs investigated in this thesis.

In a study by Huang and colleagues on improving computational seed based miRNA target predictions, an assumption was made that functional miRNA targets "should have more consistent Gene Ontology annotations than random subsets of the sequence-based predictions" (Huang et al., 2007). A target prediction method, which was supported by the conclusion drawn from this assumption[1], was validated experimentally on the example of let-7 targets in human retinoblastoma (Huang et al., 2007). Therefore, the total number of GO enriched categories was assumed to be indicative of the functional significance of a gene list.

I have estimated the total number of enriched GO categories in targets of the six miRNAs from this thesis (as in Huang *et al.*, the type of "Biological process", size > 5 genes). This parameter was also estimated for the Ago HITS-CLIP set (section 6.1.2), where 267 GO categories were enriched at P-value threshold of < 0.05 (Figure 6.6a). This provided a benchmark for the scope of GO enrichment that may be expected in a broad set of functional miRNA targets. A similar number of GO categories (247 categories) was enriched in miR-124 transfection targets. At the same time, the number of GO categories enriched in targets of all other miRNAs (miR-434-3p, miR-143, miR-145, miR-25 and cel-miR-67) was more than two times smaller, i.e. around 100 categories and less. At a stricter P-value threshold of 0.001, there were still over twice as many GO categories enriched in miR-124 targets (32 categories) compare to miR-434-3p targets (14 categories), while targets of all other miRNAs had only one to three categories enriched (Figure 6.6b). It should be noted that this result could not be completely explained by differences in the number of identified targets for different miRNAs and resulting behaviour of the hypergeometric test (it is more difficult to obtain a significant results for smaller lists): at P-value threshold of 0.001 there were over three times as many GO categories enriched in

[1] with the focus on a "Biological process" type of GO categories of the size bigger than 5 genes (Ashburner et al., 2000)

miR-434-3p targets than in targets of any of the non-neuronal miRNAs (miR-25, miR-143, miR-145 and cel-miR-67), although the number of putative direct targets of miR-434-3p was the smallest of all miRNAs. The result was also not explained completely by the optimal timepoint to detect miR-124 and miR-434-3p targets being 6DIV, while for other miRNAs it was 4DIV (Chapter 5): analysis of 240 targets of miR-124 determined from a 4DIV experiment (a suboptimal timepoint for miR-124 target identification, see Chapter 5, section 5.2.1) showed enrichment of eight GO categories (P-value < 0.001), which was over twice as many enriched categories as was in targets of any of the non-neuronal miRNAs.

It should be pointed out that the results presented in this section are not, on their own, a definitive proof of greater functional significance of the two neuronal miRNAs for primary neurons, but they do provide additional support to the same conclusion being drawn from other results discussed in this chapter. For example, genes induced by the transfection reagent (i.e. the transfection stress) were the most significantly enriched in targets of miR-124 and miR-434-3p (Figure 6.2). Additionally, in the next section I will describe these two miRNAs as efficient inhibitors of genes that were induced in the brain by other two types of stresses (section 6.2.1). Together, these findings suggested that the neuronal miRNAs were more efficient than other miRNAs at inhibiting functionally related groups of genes (such as genes associated with GO terms, or genes co-expressed upon a treatment of the cultures).

Summary of section 6.1

Transfected miRNAs converged on inhibiting expression of a shared set of targets (which I referred to as a pool of targets primed for miRNA mediated regulation). Genes that were induced during the transfection experiments contributed significantly to the shared pool of targets: miRNA targets of five out of six miRNAs were enriched in genes induced by the transfection reagent. I also found that these genes were significantly enriched in published miR-124 transfection targets (Lim et al., 2005; Makeyev et al., 2007). These results suggest that miRNAs inhibit inducibly expressed genes, such as genes induced by the transfection reagent. Therefore targets of miRNAs (hence, their function) are defined, to a significant extent, by the experimental context.

In agreement with this proposition, targets identified by HITS-CLIP, which is a transfection-free method (Chi et al., 2009), were enriched in pathways related to a normal neuronal function rather than those affected by transfection. For example, KEGG path-

(a) Counts of the enriched terms ($P < 0.05$) **(b)** Counts of the enriched terms ($P < 0.001$)

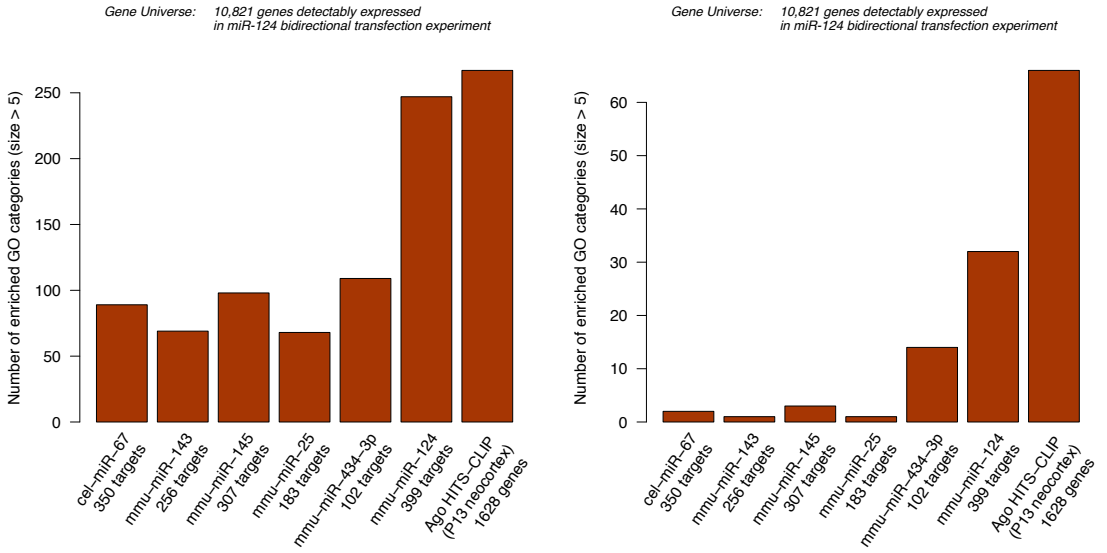

Figure 6.6: Counts of GO categories ("Biological process", size > 5 genes) enriched in miRNA targets.

ways "Long-term potentiation", "Regulation of actin cytoskeleton" and "Axon guidance" were three of the four most enriched pathways in the complete Ago HITS-CLIP set of targets. On the other hand, genes induced by the transfection reagent were enriched in disease and stress related pathways, and miRNA targets identified by transfection were also enriched in multiple disease and stress related pathways. Therefore, targets of miRNAs identified in transfection experiments may be informative of miRNA function in diseases and stresses, while HITS-CLIP targets will elucidate function in normal neurons.

Finally, neuronal miRNAs (miR-124 and miR-434-3p) were identified as having greater functional significance for primary neurons than non-neuronal miRNAs. Targets of neuronal miRNAs were more enriched in genes co-expressed in neurons than the targets of the non-neuronal miRNAs. For example, of all miRNAs, targets of neuronal miRNAs were the most significantly enriched in genes induced by the neuronal response to the transfection reagent, and also in genes associated with GO terms. Therefore, it is possible that neuronal miRNAs evolved to buffer genes that are inducible in neurons.

6.2 The function of miRNAs in neurons and the brain

6.2.1 In transfection experiments miRNAs downregulated stress inducible genes

Genes induced by transfection were enriched in KEGG pathways associated with diseases and stresses (section 6.1.3). For example, the KEGG pathways "p53 signaling pathway" (Supplementary Data, Table A.43), "Toll-like receptor signaling pathway" (Supplementary Data, Table A.48) and "pathways in cancer" (Table 6.1) were enriched among targets of several miRNAs. To test if miRNAs can inhibit a wide spectrum of genes induced by stress (i.e. not only the genes associated with KEGG terms), I obtained from published literature two lists of genes that were induced by the adverse treatments of the brain. These two published experiments were: an injection of kainate into the mouse hippocampus (Akahoshi et al., 2007) and ageing of the human brain (Lu et al., 2004).

Whole transcriptome microarray profiling data was available for the kainate injection (Akahoshi et al., 2007) and mock transfection experiments[1] (section 6.1.3), and I used Sylamer (Methods, section 2.8) to test if the innate biases in distribution of miRNA seed matching sites could be observed in these experiments. If stress upregulated genes encoded transcripts that were enriched in miRNA seed matching sites, that would suggest that there was scope for buffering of these genes by endogenous miRNAs. Such significant enrichment of miR-124 seed matching sites was observed in transcripts upregulated by both the mock transfection and kainate injection experiments, while in the latter a bias for miR-434-3p was also observed (Figure 6.7). One possible explanation to these biases can be that under normal conditions (i.e. before the kainate stress or the transfection) expression of the inducible transcripts is moderated by the endogenous miR-124 and miR-434-3p.

For the next step, I assessed whether lists of genes induced by the stress were significantly downregulated by transfected miRNAs. Genes induced by mock transfection and by the kainate stresses were derived from microarray expression profiling data[2] and I also obtained the list of genes induced by ageing of the human brain directly from a publisher's website[3]. The lists of genes induced by three types of stresses (the transfection, kainate

[1] Mock transfection experiment was performed and analysis as a part of this thesis project (section 6.1.3), and the raw microarray profiling data for the kainate injection experiment (Akahoshi et al., 2007) was available from Gene Expression Omnibus (GEO) database (Sayers et al., 2010), GEO ID GSE6388.

[2] as genes upregulated with differential expression $P < 0.05$ (Methods, section 2.7).

[3] http://www.nature.com/nature/journal/v429/n6994/suppinfo/nature02661.html. Mouse homologs of these human genes were obtained from HomoloGene Version 65 (Sayers et al., 2010).

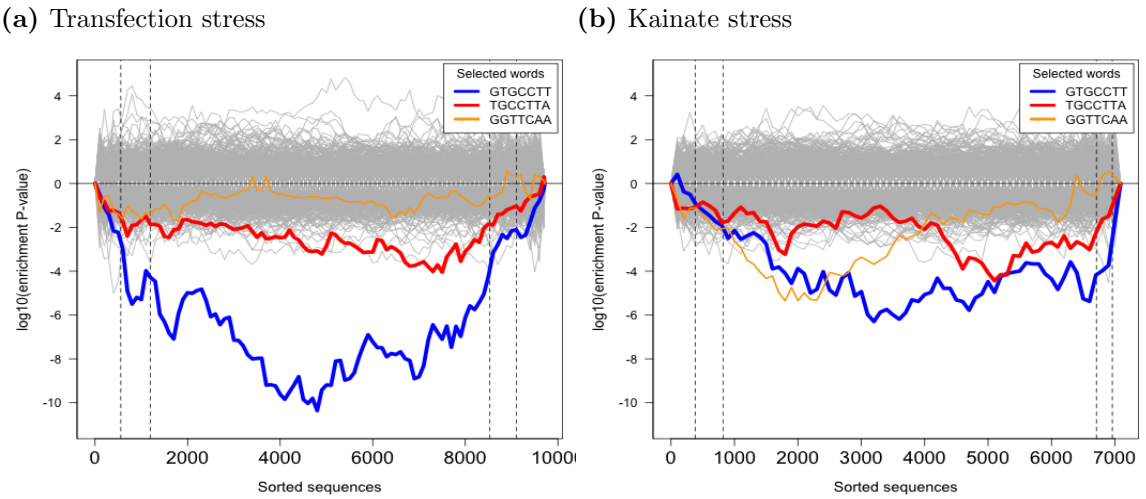

Figure 6.7: The innate miRNA seed matching site distribution biases in transfection and kainate stresses.
The x-axes represent 3′UTRs corresponding to genes sorted **from the most downregulated on the left to the most upregulated on the right** according to the fold change t-statistics in: 6.7a – the mock transfection experiment (the transfection stress); 6.7b – the kainate injection experiment (the kainate stress). The vertical dashed lines mark the P-value cutoffs (0.01 and 0.05) on both sides of the ranked gene lists. The y-axes represent the hypergeometric P-values ($-|log_{10}$(P-value)| if depletion, $+|log_{10}$(P-value)| if enrichment) for occurrence biases of 878 distinct seed matching sites (7(2) and 7(1A)-types) for a complete list of mature mouse miRNAs, which is 581 distinct miRNAs according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008). Blue and red lines show enrichment profiles of 7(2) and 7(1A)-type seed matching sites for miR-124, the orange line – 7(2)-type seed matching site for miR-434-3p. 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 parameters of Sylamer analysis (van Dongen et al., 2008) is in Methods (section 2.8).

and ageing stress) were frequently significantly downregulated in the miRNA transfection experiments of this thesis. The most significant relative downregulation of all three lists of the stress induced genes was achieved by neuronal miRNAs, miR-124 and miR-434-3p (Figure 6.8). Also, overexpression of miR-145 significantly downregulated genes induced in the ageing, and miR-143 – the kainate stress induced genes, while both miRNAs downregulated genes induced by the transfection stress (Supplementary Data, Figure A.7). Interestingly, neither miR-25 nor cel-miR-67 led to a significant downregulation of any of the three stress induced sets. This analysis suggested that although downregulation of stress induced genes may be a functional feature of miRNA mediated regulation as a whole, the neuronal miRNAs, such as miR-124 and miR-434-3p, may be specifically adapted for this function in the brain and neurons.

Another hypothesis generated by the analysis of genes induced by the three types of stresses, was that the stress response in neurons and in the brain may be canalised, or in

other words, the stress response is reproducible[4]. This conclusion was drawn from the observation of a significant similarity between changes in gene expression that was observed upon each of these stresses, despite their unrelated nature. The similarity manifested itself in significant intersections of the genes that were induced by the stresses[5] (Figure 6.9). The capability of several miRNAs to downregulate genes that were induced by the three unrelated stresses suggests that miRNAs may be involved in canalisation of the stress response. In agreement with this hypothesis, genes upregulated in all three stresses were targets of several miRNAs, including miR-124 (7 targets, which was ≈ 12 times more than expected by chance alone, $P < 7.6e-09$) and miR-434-3p (3 targets, ≈ 18 times more than expected by chance alone, $P < 0.00057$). The genes that were induced by all three stresses (unrestricted by the gene universe there were 19 of such genes) and their targeting by different miRNAs is listed in Table 6.2.

In summary, inhibition of genes induced by the stresses was identified as a feature of several miRNAs overexpressed in the transfection experiments. At the same time, the targeting repertoires of neuronal miRNAs (miR-124 and miR-434-3p) are likely to be more specifically adapted for inhibition of genes induced by stress of the brain and neurons, than random miRNA targets (such as the targets of cel-miR-67, for example). Moreover, a signature of the innate activity of miR-124 and miR-434-3p was discovered in the stress induction experiments, suggesting that both endogenous miR-124 and miR-434-3p may act as buffers of stress induced genes under normal circumstances. Genes that were shared between the three stresses were found to be highly enriched in targets of both miR-124 and miR-434-3p (≈ 12 and ≈ 18 times more than expected by chance alone), suggesting that miRNA mediated buffering of expression of these genes may be particularly important.

[4]Canalisation is the term that describes the reproducibility of the outcome of a developmental gene expression program, despite mutations of individual genes in the program (Siegal and Bergman, 2002; Hornstein and Shomron, 2006). I use this term to describe reproducibility of the outcome of the stress response gene expression program, despite the stress being triggered by unrelated factors.

[5]A strict test universe, consisting of 7,646 genes detectably expressed in development of both hippocampal and forebrain cultures (Chapter 3, section 3.1) was used to estimated the hypergeometric P-values for these intersections.

Symbol	Description	miR-124	miR-434-3′	miR-25	miR-143	miR-145	cel-miR-67
Anxa3	annexin A3		+		+		
Anxa5	annexin A5	+					
Prdx6	peroxiredoxin 6					+	
Ddr1	discoidin domain receptor family, member 1						
Cav1	caveolin 1, caveolae protein	+					
Cyp1b1	cytochrome P450, family 1, subfamily b, polypeptide 1						
Gfap	glial fibrillary acidic protein						
Gja1	gap junction protein, alpha 1		+				
Sdc2	syndecan 2		+	+			
Lamp2	lysosomal-associated membrane protein 2	+					
Myo10	myosin X						
Ntrk2	neurotrophic tyrosine kinase, receptor, type 2	+				+	+
Pmp22	peripheral myelin protein 22					+	
Tgif1	TGFB-induced factor homeobox 1						
Itpkb	inositol 1,4,5-trisphosphate 3-kinase B	+					
Pon2	paraoxonase 2				+	+	
Litaf	LPS-induced TN factor	+			+		
Tsc22d4	TSC22 domain family, member 4	+			+		
Wwtr1	WW domain containing transcription regulator 1						

Table 6.2: miRNA targeting of genes induced in three stress types.
The fist two columns of the table *Symbol* and *Description* give the official name symbols and descriptions of the 19 genes that were induced by all three stresses (the ageing, kainate and transfection stresses). The remaining columns indicate presence of these genes in the target lists of: miR-124 (Chapter 5, section 5.2.1); miR-434-3p (Chapter 5, section 5.2.3); miR-25 (Chapter 5, section 5.2.2); miR-143 (Chapter 5, section 5.2.2); miR-145 (Chapter 5, section 5.2.2); cel-miR-67 (Chapter 5, section 5.1.1).

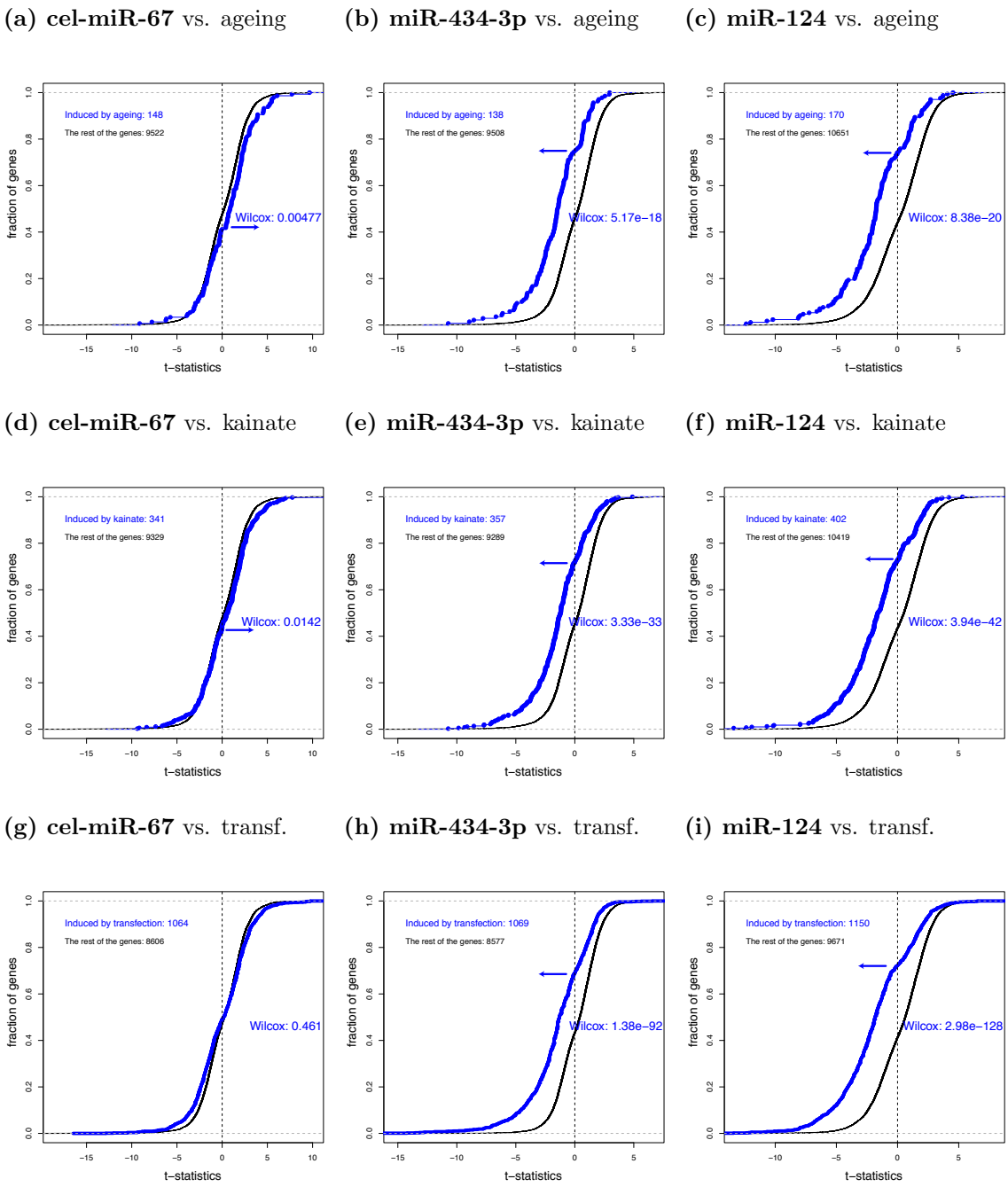

Figure 6.8: miR-124 and miR-434-3p downregulated stress induced genes.
The y-axes show the cumulative fraction of genes, the x-axes show the fold change t-statistics (Methods, section 2.7). Genes significantly induced (differential expression $P < 0.05$) by one of the three stresses (the ageing, kainate or transfection stresses) are shown as the blue line/points. The rest of the genes (except 0.01% most highly up- and downregulated genes, which were not plotted for the purpose of better scaling) is shown as the black lines. The text in the plot areas shows: 1) The number of genes *induced by* a stress that were expressed in the miRNA transfection experiments; 2) The number of other expressed genes (*The rest of the genes*); 3) The Wilcoxon test P-value for the difference in medians of the fold change t-statistics for the stress induced genes and the rest of the genes (*Wilcox*). The blue arrows show the direction of the shift in experiments where the Wilcoxon test P-value was significant ($P < 0.05$). The titles of the subfigures show: The names of the perturbed miRNAs (in bold) and the name of the stress experiment where the stress induced genes were identified.

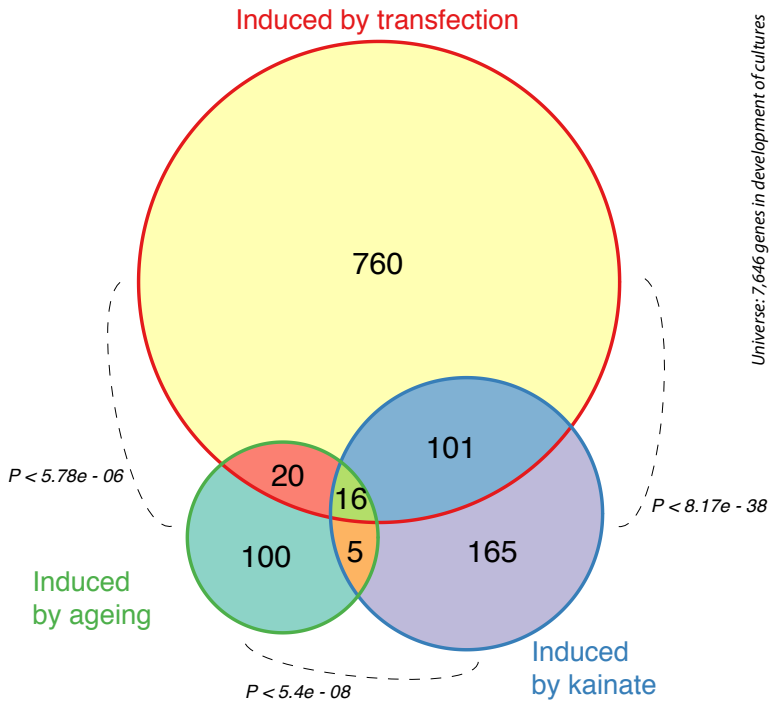

Figure 6.9: Significant intersections of genes induced in three stresses of neurons and the brain.
The Venn diagram shows the number of genes upregulated in three types of stress (the ageing, kainate and transfection and that are present in the test universe (*Gene Universe*). The test universe is 7,646 genes that were detectably expressed (using the standard Illumina detection call $P < 0.01$, see Methods, section 2.7) both in hippocampal and forebrain cultures development (Chapter 3, section 3.1). The hypergeometric test P-values are shown for each of the intersections between pairs of the lists.

6.2.2 Synaptic genes linked to neurological disease were enriched in miR-124 targets and in stress induced genes

The previous section showed that miRNAs can downregulate genes inducible by stresses in the brain and neurons. To gain an insight into how biologically important this miRNA mediated regulation of stress inducible genes may be, I evaluated the enrichment of neuronal genes linked to neurological diseases among genes that were induced by stresses and regulated by miRNAs. It has recently been demonstrated that 199 human genes encoding components of the synaptic proteome were genetically linked to neurological diseases (Bayés et al., 2011), of which 153 had mouse homologs expressed in primary forebrain cultures[1] (Chapter 3). This set of 153 genes (disease-linked genes) was assumed to be a list of neuronal genes, function of which was likely to be biologically important for the brain and neurons.

Next, I assessed whether the disease-linked genes were enriched in genes identified as stress induced. Indeed, enrichment was detected for all three lists of stress induced genes, which were discussed in the previous section (i.e. the kainate, the ageing and the transfection stresses, see section 6.2.1). Of the genes upregulated upon ageing of the brain, there were 11 disease-linked genes (approximately 4.4 times more than expected by chance alone[2], $P < 4.34e-05$), and among the genes upregulated by kainate injection there were 14 disease-linked genes induced (2.6 times more than expected, $P < 0.001$). The disease-linked genes were not significantly enriched in the genes induced by transfection (i.e. the set of 1,293 genes upregulated in mock transfected samples with P-value for differential expression < 0.05, see section 6.1.3). However, if genes induced by the transfection reagent were defined with a more stringent differential expression cutoff ($P < 0.01$), then the enrichment of disease linked genes became significant: 15 disease-linked genes were among the stricter set of genes induced by the transfection reagent (approximately 1.7 times more than expected by chance alone, $P < 0.0272$). In total, out of 153 disease-linked genes expressed in primary cultures, 31 were induced by one or more of the three stresses (Table 6.3), which was approximately 2.1 times more than expected by chance alone ($P < 3.67e-05$). These observations show that genes induced by different stresses

[1]The list of mouse homologs was retrieved directly from a publisher's website http://www.nature.com/neuro/journal/v14/n1/full/nn.2719.html#/supplementary-information.

[2]The gene universe that was used in to estimate and test the significance of enrichment of disease-linked genes in other gene lists was a complete list of genes expressed in primary forebrain cultures (9,826 genes).

were biologically important for neurons, because mutations in these genes were more frequently linked to neurological diseases than genes on average.

Similarly, the importance of miRNA regulation also manifested itself in the enrichment of disease-linked genes among miR-124 targets. Of the 153 disease-linked genes, twelve were identified as putative direct targets of miR-124 (Chapter 5, section 5.2.1), 2.4 times more than expected by chance alone ($P < 0.0041$). Interestingly, of the twelve targets of miR-124 that were linked to neurological diseases, seven were induced by one or more stresses (shown in bold in Table 6.3), which is approximately 2.9 times more than expected by chance alone, $P < 0.00281$. This observation indicated that genes, which were both biologically important for neurons (i.e. disease-linked genes) and at the same time stress inducible, were significantly more likely to be under miR-124 mediated regulation, than genes on average.

Table 6.3: Synaptic genes linked to neurological diseases and upregulated in stresses. The table provides a list of human synaptic genes that were induced by at least one of the three stresses (ageing, kainate or transfection) and linked to a neurological disease. *Symbol* – the Approved Gene Symbol (human); *Induced by* – description of a stress condition which induced the gene: *A* – the ageing stress; *K* – the kainate stress; *M* – the trasfection stress. Genes in **bold** were both induced in stresses and targeted by miR-124 (homologs of the targets identified in this thesis); *OMIM Disease Description* – the desease to to which the gene is linked in OMIM.

Symbol	Induced by	OMIM Disease Description
ALDH2	M	ALCOHOL SENSITIVITY, ACUTE
ALDH4A1	**A**	**HYPERPROLINEMIA, TYPE II; HPII**
APOE	K,M	ALZHEIMER DISEASE 2
APOE	K,M	LIPOPROTEIN GLOMERULOPATHY; LPG
APOE	K,M	MACULAR DEGENERATION, AGE-RELATED, 1; ARMD1
APOE	K,M	SEA-BLUE HISTIOCYTE DISEASE
C3	K,M	MACULAR DEGENERATION, AGE-RELATED, 9; ARMD9
CNTNAP2	M	CORTICAL DYSPLASIA-FOCAL EPILEPSY SYNDROME
CNTNAP2	M	AUTISM, SUSCEPTIBILITY TO, 15; AUTS15
CRYAB	A,M	ALPHA-B CRYSTALLINOPATHY
CST3	K	AMYLOIDOSIS VI
CST3	K	MACULAR DEGENERATION, AGE-RELATED, 11; ARMD11
DCX	K	LISSENCEPHALY, X-LINKED, 1; LISX1
DTNA	K	NONCOMPACTION OF LEFT VENTRICULAR MYOCARDIUM WITH CONGENITAL HEART
DTNA	K	NONCOMPACTION OF LEFT VENTRICULAR MYOCARDIUM, FAMILIAL ISOLATED, AUTOSOMAL
ENO3	K	GLYCOGEN STORAGE DISEASE XIII, GSD13

Continued on the next page

Synaptic genes linked to neurological diseases and upregulated in stresses.

Symbol	Induced by	OMIM Disease Description
ETFB	**M**	**MULTIPLE ACYL-CoA DEHYDROGENASE DEFICIENCY; MADD**
GFAP	A,K,M	ALEXANDER DISEASE
GJA1	A,K,M	ATRIOVENTRICULAR SEPTAL DEFECT; AVSD
GJA1	A,K,M	HYPOPLASTIC LEFT HEART SYNDROME
GJA1	A,K,M	SYNDACTYLY, TYPE III
GJA1	A,K,M	OCULODENTODIGITAL DYSPLASIA; ODDD
GNAI2	**K,M**	**VENTRICULAR TACHYCARDIA, FAMILIAL**
GPX1	M	GLUTATHIONE PEROXIDASE DEFICIENCY, HEMOLYTIC ANEMIA POSSIBLY DUE TO, INCLUDED
GRIA3	**M**	**MENTAL RETARDATION, X-LINKED 94; MRX94**
HSPB1	M	CHARCOT-MARIE-TOOTH DISEASE, AXONAL, TYPE 2F
HSPB1	M	NEURONOPATHY, DISTAL HEREDITARY MOTOR, TYPE IIB; HMN2B
HSPB8	**M**	**CHARCOT-MARIE-TOOTH DISEASE, AXONAL, TYPE 2L**
HSPB8	**M**	**NEURONOPATHY, DISTAL HEREDITARY MOTOR, TYPE IIA; HMN2A**
MYO6	K	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL 22; DFNA22
MYO6	K	DEAFNESS, CONGENITAL NEUROSENSORY, AUTOSOMAL RECESSIVE 37; DFNB37
NDRG1	A	CHARCOT-MARIE-TOOTH DISEASE, TYPE 4D; CMT4D
NDUFA2	M	LEIGH SYNDROME; LS
PC	K	PYRUVATE CARBOXYLASE DEFICIENCY
PLEC1	**K**	**EPIDERMOLYSIS BULLOSA SIMPLEX WITH MUSCULAR DYSTROPHY**
PLEC1	**K**	**EPIDERMOLYSIS BULLOSA SIMPLEX WITH PYLORIC ATRESIA**
PLEC1	**K**	**EPIDERMOLYSIS BULLOSA SIMPLEX, OGNA TYPE**
PLP1	A	SPASTIC PARAPLEGIA 2, X-LINKED; SPG2
PLP1	A	PELIZAEUS-MERZBACHER DISEASE; PMD
PTPN11	**K**	**NOONAN SYNDROME 1; NS1**
PTPN11	**K**	**LEOPARD SYNDROME 1**
RDX	A	DEAFNESS, AUTOSOMAL RECESSIVE, 24; DFNB24
SLC4A4	A	RENAL TUBULAR ACIDOSIS, PROXIMAL, WITH OCULAR ABNORMALITIES AND MENTAL
TPP1	K	CEROID LIPOFUSCINOSIS, NEURONAL, 2; CLN2
VCAN	A	WAGNER SYNDROME 1; WGN1

Continued on the next page

Synaptic genes linked to neurological diseases and upregulated in stresses.

Symbol	Induced by	OMIM Disease Description
WFS1	A	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL 6; DFNA6
WFS1	A	WOLFRAM SYNDROME 1; WFS1
WNK1	A	PSEUDOHYPOALDOSTERONISM, TYPE II; PHA2

6.2.3 miR-124 in development: Reduction of variability in gene expression

In this thesis I have identified several aspects of miR-124 mediated regulation that are changing with progression of development of the cultures. For example, transcripts with 3′UTRs that are not depleted of miR-124 seed matching sites, were found to be upregulated relatively late in development of cultures (from 4DIV to 8DIV, see Chapter 3, Figure 3.10). At the same time, the effect of miR-124 overexpression on gene expression in primary neuronal cultures consistently increased as development progressed, and was maximal in the experiment at the 6DIV timepoint (Chapter 5, section 5.2.1). Together with these developmental changes, genes that were induced by three types of stresses were significantly downregulated by miR-124 (Figure 6.8), therefore I investigated expression of stress induced genes in development of cultures.

Previously, miRNAs were suggested to be involved in canalisation of developmental gene expression programs (Hornstein and Shomron, 2006), therefore I investigated variability of stress induced genes in development of primary cultures. The standard deviation of intensities of microarray probes corresponding to these genes (the mapping of the probes is described in Methods, section 2.7) between the replicates within one developmental timepoint was used as a measure of the variability of gene expression.

Standard deviation in intensities of probes for stress induced genes were compared to that of the rest of the genes[1]. Additionally, as an internal control, I estimated standard deviations in intensities of probes for genes, expression of which was reduced in the stress[2]. A progressive increase in standard deviation (i.e. variability) of stress induced

[1]To avoid a bias from correlation of the standard deviation and the level of expression, the analysis was confined to top 25% most highly expressed genes, where no such correlation was observed (Supplementary Data, Figures A.8 and A.9).

[2]Genes that were reduced in the mock transfection and kainate experiments were defined as genes that were downregulated with differential expression $P < 0.05$. Genes that were reduced by ageing were obtained directly from a publisher's website (http://www.nature.com/nature/journal/v429/n6994/suppinfo/nature02661.html). Mouse homologs of these human genes were obtained from HomoloGene Version 65 (Sayers et al., 2010).

genes was observed in development of both hippocampal and forebrain cultures (Figure 6.10). Importantly, variability of the stress reduced genes did not increase in all cases but one[3] (Figure 6.10).

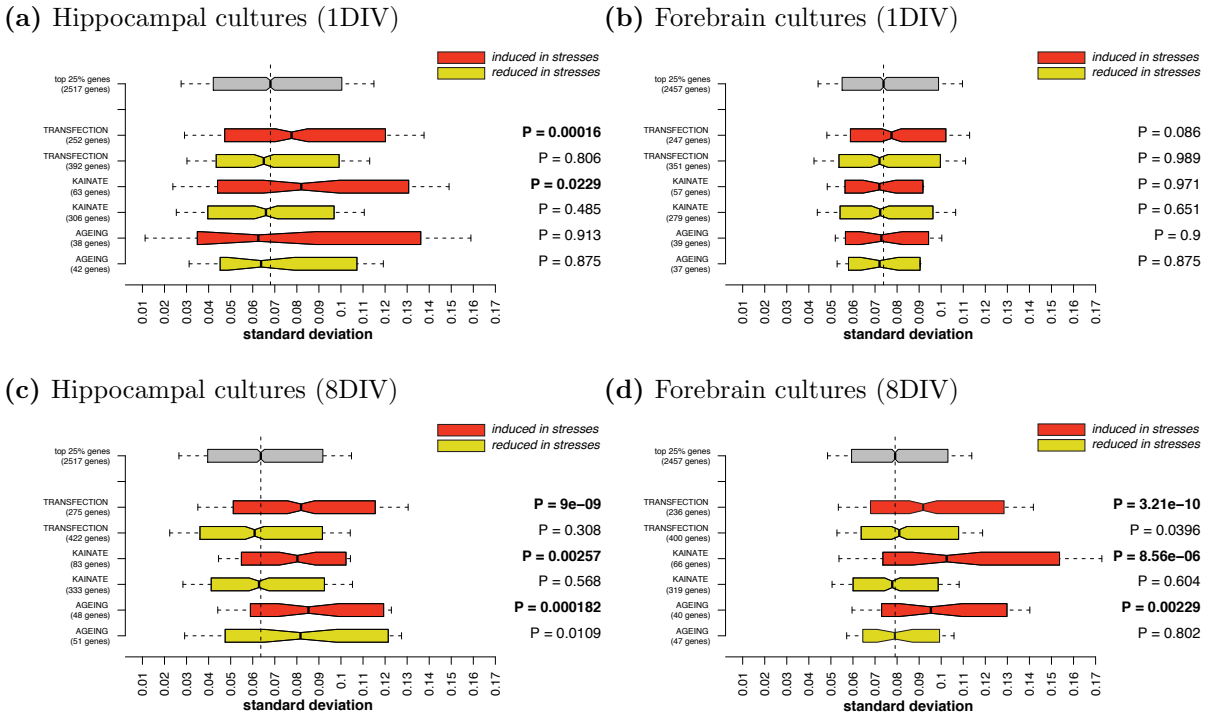

Figure 6.10: Variability in expression of stress induced genes increased with progression of development.
The x-axes show the standard deviation of gene expression between biological replicates within one developmental timepoint, the y-axes show the names of the gene lists and the number of the genes from the lists that were selected from the top 25% most highly expressed genes in the experiment (the numbers are in parentheses). The grey boxes correspond to the distribution of standard deviations of all 25% most highly expressed genes, all other boxes correspond to the genes belonging to the stress genelists (as indicated on the y-axes and according to the color-scheme). The notches in the boxes correspond to the median value of the distribution of standard deviations of the genes in the genelists. The left and the right sides of the boxes correspond to the first and the third quartiles. The whiskers extend to no more than 0.25 times the interquartile range (IQR), or to the most extreme data-point, if it is closer to the median than 0.25 IQR. The Wilcoxon test P-values are for the difference between the medians of the distributions of the genes within the gene lists (confined to the 25% most highly expressed genes) and the rest of the genes in the top 25%.

Next, I investigated whether miR-124 could play a role in controlling this variability of the stress induced genes. In order to do this, the standard deviation of probe intensities was estimated between replicates within the following treatments: transfection with the

[3]In hippocampal cultures at the 8DIV timepoint, variability in expression of genes reduced by ageing was significantly higher than for the rest of the genes. This difference, however, was smaller than that of the genes induced by ageing.

mimics of miR-124, transfection with the inhibitor of miR-124, transfection with the mimic of cel-miR-67 and a mock transfection. In samples transfected with a mimic of miR-124 the variability of the stress induced genes was not significantly different from the rest of the genes[4] (Figure 6.11a). On the other hand, in mock transfected samples or samples transfected with an inhibitor of miR-124 the variability of stress induced genes remained significantly higher than of the rest of the genes (Figures 6.11b and 6.11c). Transfection of a mimic of a non-mouse miRNA, cel-miR-67[5], at the same developmental timepoint and with the same protocol (Methods, section 2.5), did not reduce the variability in expression of the stress induced genes (Figure 6.11d).

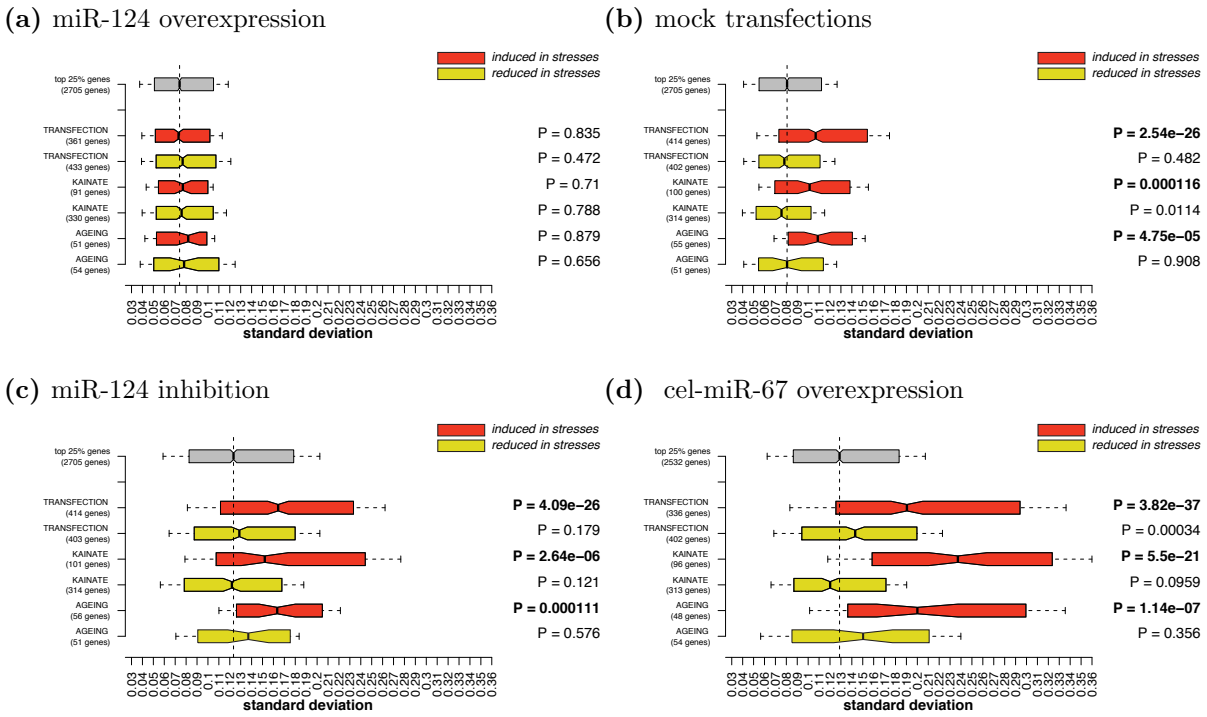

Figure 6.11: Variability in expression of stress induced genes was reduced by miR-124 overexpression.
The x-axes show the standard deviation of gene expression between biological replicates within one treatment (the treatments are named in the titles to the subfigures). See the legend to Figure 6.10 for the description.

[4]As in the case of the developmental profiling, the analysis was confined to the top 25% most highly expressed genes, where no significant correlation between the level and standard deviation of gene expression was observed in miR-124 and cel-miR-67 transfection experiments (Supplementary Data, Figures A.10 and A.11).

[5]Of the two cel-miR-67 overexpression experiments at 6DIV (Methods, section 2.5), the data from the experiment "B" was used for analysis in this section, as in the experiment "B" bigger effects on differential gene expression were observed than in the experiment "A" (compare Figure 5.1d and Figure 5.1e).

In summary, in relatively mature primary neurons (8DIV) variability in expression of genes induced by three types of stress (transfection, kainate and ageing stresses) was found to be higher than for other genes. Overexpression of miR-124 was found to reduce this variability (Figure 6.11). Therefore, it is possible that endogenous miR-124 can also act to limit variability in expression of stress induced genes in mature neurons.

Summary of section 6.2

The analyses described in this section showed that exogenously introduced miRNAs inhibited genes that were induced in three types of stress to both neurons and the brain (the transfection, kainate and ageing stress). This inhibition was most significant in the case of the two neuronal miRNAs (miR-124 and miR-434-3p), while it was not significant upon transfection of the non-mouse miRNA, cel-miR-67, or the oncogenic miRNA, miR-25. Therefore, targeting repertoires of endogenous miRNAs may have specifically evolved to inhibit genes that can be induced by the stresses of their cognate tissues and cell types.

Sylamer analysis of occurrence biases of miRNA seed matching sites in 3′UTRs of transcripts in the mock transfection experiment (transfection stress) revealed the innate signature of miR-124 mediated regulation (seed matching sites for miR-124 were enriched in 3′UTRs of transcripts upregulated by the treatment with the transfection reagent). A very similar distribution of miR-124 seed matching sites was observed upon the kainate stress. Additionally, enrichment of seed matching sites for miR-434-3p was also revealed. One possible explanation of these observations, is that in normal circumstances endogenous miR-124 and miR-434-3p can act as buffers of genes inducible by stress.

The stress induced genes targeted by miR-124 were enriched in synaptic genes and linked to neurological diseases. Therefore, the role of miR-124 in regulation of expression of stress induced genes, is likely to be of importance for neuronal biology.

In more mature primary neurons (8DIV), variability in expression of genes induced by the stresses was higher than that of the rest of the genes. Interestingly, overexpression of miR-124 suppressed this variability, while overexpression of cel-miR-67 did not. I suggest that endogenous neuronal miRNAs also regulate variability of stress inducible genes. This proposition is speculative, however it can be tested experimentally (Discussion, section 7.4).

Returning to the questions posed at the beginning of this chapter, I believe my results show that neuronal miRNAs have an important function in maintaining equilibrium in neuronal gene expression. I analysed hundreds of targets for miRNAs in neurons, and

demonstrated that targets of different miRNAs converged on the pool of transcripts that was, to a signficant extent, induced by the transfection procedure itself. At the same time, I have found that targeting repertoires of two neuronal miRNAs, miR-124 and miR-434-3p, seem to be adapted for inhibition of incucible in neuorns and the brain genes. Also, miR-124 and miR-434-3p appear to be endogenous buffers of neuronal transcription. This is important for regulation of stress response and has implications for disease. In summary, my data leads me to conclude that neuronal miRNAs act to maintain gene expression of their targets within steady state boundaries.