Small regulatory RNAs

Seinen, Erwin

Chapter 6

Summarizing discussion

Introduction

During the last decade, RNA interference (RNAi) has become an important tool to investigate the function of genes and the gene-gene interaction between them. By adding double stranded RNA (dsRNA) constructs, the RNAi pathway induces downregulation of the expression of a gene [1]. This is done by administering dsRNAs with sequence homology against the target gene. The dsRNAs inflict a cascade of reactions ending in the destruction of the messenger RNA (mRNA) originating from the target gene, and as a consequence the protein is not formed. By studying the phenotype induced by downregulation of the mRNA, the function of the chosen gene is elucidated. Unfortunately, RNAi may also have collateral effects, of which the so called off-targets are the most important. Off-target effects can occur because of the presence of sequence similarity between the target genes and other, genes.

RNAiSelect; a novel bioinformatics tool that profiles off-targets

With time, knowledge grew about how RNAi works and what the limitations are in using them for experiments [2]. One of these limitations is that off-targets may occur when an RNAi molecule has similarity to other regions on the genome than its intended target [3]. Bioinformatics plays a key role in circumventing these off-targets by analysing the sequence of the RNAi molecule and reporting about the uniqueness amongst genes [4,5,6]. Still, one important property of RNAi was ignored by the current available tools: The RNAi machinery can not only target mature RNA sequences but can also target intron containing pre-messenger RNA and other nuclear located RNAs [7,8,9]. This has major consequences for the generality of the already existing bioinformatics tools, as these are only considering mature messenger RNA (mRNA) present in the cytoplasm. The knowledge that the RNAi machinery can also have a nuclear localization, indicates that there are many more sequences that can be the target of regulatory breakdown.

Our tool “RNAiSelect” (chapter 2) responds to these observations by allowing a much more thorough analysis including the complete genome to find off-targets. RNAiSelect is based on a novel algorithm that requires only seconds to complete a comprehensive search of 21-nt against the complete genome while allowing up to 3 mismatches. By analysing the complete genome, pre-mRNA sequences are also considered as well as other known

(i.e. miRNAs) or unknown regions of the genome. This approach certainly has ample false-positive hits caused by sequences that do have similarity to the RNAi molecule under investigation, but do not inflict a real biological effect. **However, it is important to realize that the purpose of RNAiSelect is not to identify off-targets *per se*, but to select the best RNAi molecule candidates amongst the many that are possible.** By analysing all candidate siRNAs that may be created for a particular target gene, the ones with the least number of predicted potential off-targets will be the most preferable ones and from which the most suitable according to efficiency properties may be chosen. In addition, the number of mismatches against predicted potential off-targets as well as the type of mismatch is considered resulting in a general score that also aids in choosing the best siRNA with the least number of potential off-targets.

We have validated our approach using publically available micro-arrays. By analysing the off-target profiles of the used dsRNAs and after comparing the output with the actual regulatory differences on the arrays, we concluded that RNAiSelect indeed has a predictive value regarding which sequences can contribute to off-targets. Although the array showed numerous marginal expression differences considering the complete array, a significant effect was observed for the set of predicted off-targets. The knowledge that small changes at the mRNA level can have large effects on protein expression levels underscores the significant effect we have found in chapter 2 [10].

Fortunately, in our genome-wide analysis of all potential siRNA sequences using RNAiSelect, we have found that the majority of genes have potent siRNAs with relatively few predicted potential off-targets (up to 80% less than average). When performing a genome-wide screen using dsRNA molecules, RNAiSelect will be particularly useful to identify these clean RNAis and assists in the decision which of them are most “clean” and therefore most promising to proceed with. When designing a RNAi experiment, one can choose a single siRNA (21 nt in length) instead of a large dsRNA (approximately between 250 and 800 nt in length) to prevent as many potential off-targets as possible. In addition, the outcome (especially micro-array data) of already performed RNAi experiments can now be re-evaluated using RNAiSelect and it can be checked whether reported RNAi-induced effects are *bona fide* effects or can be explained by off-target effects.

A solution to the overlooked problem of off-targets shared by independent dsRNAs targeting the same gene

It is currently not possible to predict with a high level of accuracy whether a specific RNAi molecule does provoke a potent on-target effect and it is also not possible to predict whether an off-target effect, identified by bioinformatic tools will indeed provoke a measurable biological effect. A method to circumvent the usage of bioinformatics tools is to use two independent and non-overlapping siRNA constructs to downregulate the same gene of interest. The usage of this type of controlled-experiment is reasonable because the assumption is that two completely different dsRNAs will have identical on-target effects but will have a completely different genome wide off-target profile. Any identical biological effect provoked by the distinct dsRNA constructs is considered as a *bona fide* on target effect.

Unfortunately, this assumption has proven to be very wrong (chapter 3). Statistical analysis and our results show that within the *Drosophila* genome it is highly likely for distinct 21 nt sequences derived from the same gene to have closely mapped sequence similarity elsewhere on the genome. In other words, if one slices a Drosophila gene in pieces of 21-nt and a sequence similarity analysis is performed, many of the individual 21-nt pieces will find homology within the same 'other' *Drosophila* gene (i.e. potential off-target gene). **Indeed, our results using actual genome data show that most genes have many of these sequences that map to the same potential off-target. We therefore concluded that an identical biological effect induced by randomly choosing distinct dsRNAs derived from one gene is not per definition a *bona fide* on-target effect.** Moreover, our results show that even when 3 or more distinct dsRNAs are used, there is still a significant possibility of overlapping off-targets which means that "blindly" picking dsRNA constructs is almost never suffice to counter off-target effects.

Despite this knowledge, the use of a double controlled RNAi experiment is appealing as it may indeed prevent many false observations. RNAiSelect (presented in chapter 3) is useful in finding the best predicted siRNA candidate combinations that have most likely no shared off-targets based on sequence similarity. In summary, we present a method to identify 2 distinct dsRNAs from a gene of choice that do not show any potential off-target overlap, -based on sequence similarity- by performing a thorough potential

off-target overlap analysis. This tool is freely available at http://www.rnaiselect.info/dsrna and may be used by the *Drosophila* community where dsRNAs are generally used for gene down-regulation studies.

Pantethine rescues dPANK depleted Drosophila S2 cells involved in CoA metabolism

The procedure of using RNAi in researching a human disease model in *Drosophila* has proven to be very useful to unravel the mechanisms behind the disease [11]. RNAiSelect has been used in chapter 4 to generate clean and specific dsRNAs to model the human disease PKAN (Pantothenate kinase–associated neurodegeneration) in *Drosophila*. PKAN is a rare neuronal disease caused by a mutation in the human PANK2 gene.

PANK2 is evolutionary conserved amongst many species and is essential in the biochemical pathway that converts vitamin B5 into Coenzyme A (CoA) [12]. *Drosophila* has proven to be particularly suitable as a model for PKAN, as it shares many characteristics that are reported in human PKAN patients carrying a mutation in PANK2. These phenotypes comprise increased loss of locomotor function, neurodegeneration, and a decreased lifespan.

With RNAi, a cell based PKAN *Drosophila* model (S2) was created. A biochemically measurable effect of this knock-down mutant cell line is the severely decreased levels of CoA. In addition, a phenotypical effect is measured through decreased cell-count in time. By adding the chemical substrate pantethine to the growth medium, CoA levels and cell numbers were restored comparable to wild type S2 cells. Thus our research suggests that an alternative route may exist leading from pantethine to CoA, independent from pantothenate kinase. This hypothesis is further supported by our findings that any residual *dPank/fbl* kinase activity does not on itself contribute to the phosphotransferase conversion of Pantethine, because (1) western blotting demonstrated the nearly complete knock-down of dPank/fbl and (2) the addition of increasing concentrations of vitamin B5 did not have any significant effect on the mutant phenotype.

Pantethine has also been administered to *Drosophila* fly mutants, which rescues all tested aspects of their neurodegenerative phenotype and increased their life span. Moreover, it was demonstrated that in a human cell

model for PKAN, pantethine also works protective. Together our results show that a well-controlled RNAi experiment in Drosophila S2 cells leads to the identification of a lead compound, that may be at the base of a possible treatment for a so far non-treatable disease.

Identifying miRNA targets in human cells to understand Hodgkin Lymphoma

In addition to the possibility of our algorithm to search for full length sequence similarities, short 6-nt 'seed' sequences that can result in possible miRNA-like effects can be searched for. The latter is also useful for identifying or confirming naturally occurring miRNA targets that regulate endogenous genes. This is particularly interesting for research concerning the role of miRNAs in developmental regulation, but also to study the effect of malfunctioning miRNAs as they occur in several cancerous diseases like Hodgkin Lymphoma (HL; a cancer originating from leucocytes).

Chapter 5 describes a high-throughput approach to identify endogenous miRNA targets of untreated human cells in which specific miRNAs are upregulated. The approach uses a combination of wet experiments (in vitro/vivo studies) and RNAiSelect (in silico) capabilities to find sequence similarities. While other methods have been established to find miRNA targets, none of them can be done in a wild type background because these published methods make use of cells that overexpress synthetic miRNAs and flag-tagged miRNA related proteins which require a significant modulation of the cells [13,14,15,16,17]. These approaches clearly influence the physiological state of the cells under investigation and will affect the outcome of the experiments. This underscores the need for a better system. Our high throughput approach (as described in chapter 5) is not altering physiological conditions and is accomplished by using antibodies against Argonaute (Ago2; which is the catalytic component of the RISC complex that binds and cleaves targeted mRNAs by complementary binding with the miRNA seed region) in combination with immuno-precipitation of this RISC complex. The precipitated complex contains mRNAs that were bound to RISC and are therefore candidate miRNA targets. Using this method, miRNA targets bound to Ago2 can be isolated from untreated and physiologically relevant cell or tissue samples and identified by nucleotide sequencing.

A proof of principle has been performed by using 2 different HL cell lines that have several characteristic miRNAs upregulated (miR-17/20/93/106). First, RISC from untreated HL cells was isolated and attached mRNAs were isolated, sequenced and identified. Next, the HL cell lines have been transfected with anti-miRNAs (i.e. anti-miR-17) and again pools of mRNAs that were attached to Ago2 have been isolated. By cross-referencing both pools, mRNAs that are bound to RISC in the unaltered cells but not in transfected cells could be isolated and were candidates for being actual targets of the miRNAs under investigation.

Although this pool is large and contains many false positives due to non-specific or non-related binding of mRNAs to RISC, it is a perfect start for RNAiSelect to condense the list of candidate genes to a validated list based on seed complementarity to the miRNAs under inverstigation. By using RNAiSelect, a list of mRNAs containing highly significant enrichment of seed-sequences against the miRNAs under investigation has been isolated from the large pool of mRNA that were found in the Ago2 immune-precipitate. Because these mRNAs are both found in RISC and have multiple occurrences of the seed sequences against miRNAs that are known to be highly active, it is very likely for them to be actual targets. Indeed, by using luciferase assays against 8 chosen genes from the isolated mRNA set predicted to be downregulated by miRNA-17, all of them are confirmed to be upregulated upon miRNA-17 inhibition. Notably, 2 genes have also been assayed using the same method, which showed significant enrichment in the Ago complex, but appeared to lack the seed region of miRNA-17 after analyzing with RNAiSelect. Upon miRNA-17 inhibition, the correlation with this miRNA could not be confirmed consisting with the lacking seed region. This observation demonstrates that initial large pools of mRNAs could effectively be reduced with RNAiSelect by reducing false positives and therefore provides the necessary sensitivity of the novel high-throughput approach to identify miRNA targets.

This thesis has also given new insights into how miRNA-mRNA interactions occur. Studies from the Bartel group have shown that 8-mer seed sequences show a higher and more reliable occurrence of interaction, suggesting a preference for these 8-mer interactions [18]. Our results however show that although 8-mer sites in 3'-UTR regions definitely are good indicators to identify miRNA-specific targets, they are not obligatory for effective targeting. In reality, 5 out of 9 targets of the miR-17 seed

family were confirmed with the luciferase assay, but did not contain 8-mer sites as opposed to the 6-mers.

Summarizing conclusion

Gene expression regulation by RNAs is part of a complex machinery for which recently most has been unraveled. Roughly this topic can be divided in endogenous RNA regulation and exogenous RNAi research and examples of both of them have been under investigation and discussed in detail in this thesis. First bioinformatics tools were designed to be able to evaluate results obtained with manipulating regulation by RNA (chapter 2 and 3) and subsequently these tools were implemented in 'wet' experiments (chapter 4 and 5). For exogenous siRNAs or derivatives thereof, novel methods have been devised that aid in designing clean RNAi experiments to identify a lead compound suitable to develop a future therapy for a so far non-curable disease. For the endogenous regulation, a novel method has been presented to deduce natural miRNA targets in order to interpret and further direct 'wet's experiments highly efficiently. The results presented in this thesis underscore the importance of merging bioinformatics in medical relevant life science research..

References

1. Misquitta L, Paterson BM (1999) Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): A role for nautilus in embryonic somatic muscle formation. Proceedings of the National Academy of Sciences of the United States of America 96: 1451-1456.
2. Maine EM (2001) RNAi As a Tool for Understanding Germline Development in Caenorhabditis elegans: Uses and Cautions. Developmental Biology 239: 177-189.
3. Jackson A, Bartz S, Schelter J, Kobayashi S, Burchard J, et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. Nature biotechnology 21: 635-637.
4. Yamada T, Morishita S (2005) Accelerated off-target search algorithm for siRNA. Bioinformatics 21: 1316.

5. Qiu S, Adema C, Lane T (2005) A computational study of off-target effects of RNA interference. Nucleic Acids Res 33: 1834-1847.
6. Naito Y, Yamada T, Matsumiya T, Ui-Tei K, Saigo K, et al. (2005) dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. Nucleic Acids Res 33: W589-591.
7. Lin S-L, Kim H, Ying S-Y (2008) Intron-mediated RNA interference and microRNA (miRNA). Frontiers in bioscience : a journal and virtual library 13: 2216-2230.
8. Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes & development 18: 2873-2878.
9. Bosher J, Dufourcq P, Sookhareea S, Labouesse M (1999) RNA Interference Can Target Pre-mRNA: Consequences for Gene Expression in a Caenorhabditis elegans Operon. Genetics 153: 1245-1256.
10. Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between Protein and mRNA Abundance in Yeast. Molecular and Cellular Biology 19: 1720-1730.
11. Botas J (2007) Drosophila researchers focus on human disease. Nat Genet 39: 589-589.
12. Zhou B, Westaway SK, Levinson B, Johnson MA, Gitschier J, et al. (2001) A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. Nat Genet 28: 345-349.
13. Easow G, Teleman AA, Cohen SM (2007) Isolation of microRNA targets by miRNP immunopurification. RNA 13: 1198-1204.
14. Zhang L, Ding L, Cheung TH, Dong M-Q, Chen J, et al. (2007) Systematic Identification of C. elegans miRISC Proteins, miRNAs, and mRNA Targets by Their Interactions with GW182 Proteins AIN-1 and AIN-2. Molecular cell 28: 598-613.
15. Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, et al. (2007) A biochemical approach to identifying microRNA targets. Proceedings of the National Academy of Sciences 104: 19291-19296.
16. Hendrickson DG, Hogan DJ, Herschlag D, Ferrell JE, Brown PO (2008) Systematic Identification of mRNAs Recruited to Argonaute 2 by Specific microRNAs and Corresponding Changes in Transcript Abundance. PLoS ONE 3: e2126.
17. Landthaler M, Gaidatzis D, Rothballer A, Chen PY, Soll SJ, et al. (2008) Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA 14: 2580-2596.
18. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell 27: 91-105.