university of groningen

University Medical Center Groningen

University of Groningen

RNAi-induced off-target effects in Drosophila melanogaster

Seinen, E.; Burgerhof, J.G.; Jansen, R.C.; Sibon, O.C.

Published in:
Briefings in Functional Genomics

DOI:
10.1093/bfgp/elr017

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Seinen, E., Burgerhof, J. G., Jansen, R. C., & Sibon, O. C. (2011). RNAi-induced off-target effects in Drosophila melanogaster: frequencies and solutions. *Briefings in Functional Genomics*, *10*(4), 206-214. https://doi.org/10.1093/bfgp/elr017

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 10-10-2024

BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 10. NO 4. 206–214 doi:10.1093/bfgp/elr017

RNAi-induced off-target effects in *Drosophila melanogaster*: frequencies and solutions

Erwin Seinen, Johannes G.M. Burgerhof, Ritsert C. Jansen and Ody C.M. Sibon

Advance Access publication date 19 May 2011

Abstract

Genes can be silenced with short-interfering RNA molecules (siRNA). siRNAs are widely used to identify gene functions and have high potential for therapeutic treatments. It is critical that the siRNA specifically targets the expression of the gene of interest but has no off-target effects on other genes. Although siRNAs were initially considered to be exclusively active on mature mRNAs in the cytoplasm, additional studies have shown that siRNAs are present in the nucleus as well, suggesting that pre-mRNA sequences containing introns and other untranslated regions can also be targeted. In this study, we investigated the extent to which off-targets may occur in *Drosophila melanogaster* by looking at mature mRNA sequences and pre-mature RNA sequences separately. First, an in *silico* approach revealed that, based on sequence similarity, numerous off-targets are predicted to occur in RNAi experiments. Second, existing microarray data were used to investigate a possible effect of the predicted off-targets based on analysis of *in vitro* data. We found that the occurrence of off-targets in both mature and pre-mature RNA sequences in RNAi experiments can be extensive and significant. Possibilities are discussed how to minimize off-target effects.

Keywords: *Drosophila; RNAi; off-target; on-target; dsRNA; microarray*

INTRODUCTION

In experiments based on RNA interference (RNAi) technology, off-target effects can be induced by unintended cross-hybridization between short-interfering RNA molecules (siRNAs) and endogenous RNA sequences other than the targeted sequences [1–5]. Off-targets obscure the functional interpretation of gene silencing experiments [1] and should therefore be avoided as much as possible. Potential hybridizations between siRNAs and mature mRNAs, are generally identified by *in silico* approaches using sensitive tools specifically designed to predict possible RNAi off-targets [6–8]. To our knowledge, these existing tools only consider the mature mRNA sequences, but not promoter and intron sequences. However, studies have shown that siRNAs can also act in the nucleus [9–14] where they target promoters [14] and introns [9]. Other on line tools are available that enable the design of effective RNAi constructs [15–17]. Annotated genome libraries can be searched (using BLAST) for off-targets at the genome-wide level [18]. Recently, we described a highly sensitive method that enables the identification of sequence similarity between short stretches of contiguous nucleotides with up to three mismatches [19], with a speed and sensitivity that outperforms BLAST-related tools. This method enables the identification of potential off-targets at a genome-wide level (including homologous sequences containing G:U

Corresponding author. Ody C.M. Sibon, Section of Radiation & Stress Cell Biology, Department of Cell Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. Tel: +(0)503632559; Fax: +(0)503632913; E-mail: o.c.m.sibon@med.umcg.nl

Erwin Seinen received his master's degree in 2005. He will obtain his PhD 11 April 2011 and he is the founder and manager of the Dutch company Bio-ITech.

Johannes G.M. Burgerhof is a statistician at the University Medical Center Groningen.

Ritsert C. Jansen is full professor at the Groningen Bioinformatics Center, University of Groningen, Haren, The Netherlands and at the Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands.

Ody C.M. Sibon is full professor at the University Medical Center Groningen, Department of Cell Biology in the Netherlands.

© The Author 2011. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

wobble mismatches which exhibit a very high binding energy [20, 21]). Here, we used this method to investigate the occurrence of predicted genome-wide off-targets of dsRNA sequences derived from six genes. Our analysis of these dsRNAs revealed that based on sequence similarity, there is a high prevalence of predicted off-targets. Subsequently, existing microarray data were analyzed generated with the dsRNAs to investigate whether the set of predicted off-targets are indeed more downregulated compared to the total set of transcripts presented on the array. These results demonstrated that the predicted off-targets (present in pre-mature and in mature mRNA) are significantly more downregulated as compared to genes outside the set of predicted off-targets. Our results show that off-target effects can obscure RNAi experiments and possible solutions are discussed as how to circumvent these unwanted effects.

METHODS

Computer hardware and software

Genomic data (build 45–43 b) were downloaded from the Ensembl website (www.ensembl.org). The data from Ensemble and its derived seed tables were processed, searched and stored in a MySQL database, version 5.0, running on top of Ubuntu 6.06. The on-line available RNAiSelect program (http://www.RNAiSelect.info/) was written in C#.NET and runs on Microsoft Internet Information Server 6.0, running inside a VMware virtual machine and connected to the MySQL database server. Both database and application were hosted on a single system with dual XEON 5140 2.33 GHz processors and 16 GB of 667 ECC memory.

RNAiSelect algorithm

The algorithm RNAiselect was used as previously described [19]. In short: the RNAiSelect algorithm was specifically designed for finding relationships between short nucleotide sequences. It has a high performance and usability for any short-sequence study, including siRNA (off-)targets or miRNA docking sites. The algorithm is based on the following assumption:

> *'An example sequence TTTTAATTTGGGCCCGGG consists of 18 nt and may be split into two 9-nt child sequences; TTTTAATTT and GGGCCCGGG. By plain observation, we know that the sequence GGGCCCGGG is exactly 9-nt separated from TTTTAATTT in the original sequence.'*

For the RNAiSelect algorithm to work, we first wrote a program that generates a seed table which holds the exact *Drosophila* genomic location(s) for every possible 9-nt sequence (4^9, or 262.144 sequences). This seed table is then used to rapidly look-up any sequence of 9 nt for its positions in the genome. By searching 9-nt subsequences of the whole-query sequence for consecutive matches of locations, it will find hits larger than 9 nt without performing actual DNA comparisons. This following example, in layman code, shows how to find an 18-nt sequence in the genome by first splitting the sequence into its two 9-nt subsequences and comparing these sequences with the available index table with a word size of 9.

(1) SPLIT QUERY SEQUENCE(18 nt) INTO *dnacode_left(9 nt)* AND *dnacode_right(9 nt)*
(2) EXTRACT LOCATIONS FROM *index_table* FOR dnacode_left AND STORE IN *seedtable_left*
(3) EXTRACT LOCATIONS FROM *index_table* FOR dnacode_ *right* AND STORE IN *seedtable_*right
(4) SELECT ALL HITS WHERE (LOCATIONS *seedtable_left + 9) EQUALS (LOCATIONS seedtable_right)*

This example merely demonstrates how to find an exact 18-nt hit not allowing any mismatches. However, users can allow mismatches by expanding the seed searches by variations in such a way that all possible combinations will be found. We thus included variations of the 9-nt sub-sequences and then compared the distance relationship between the original locations of the seed hits, which has to be exactly 9. This may considerably increase the number of seed searches, but because these are relatively cheap in terms of processing time, the overall performance is very high while it guarantees that every possible alignment is evaluated.

Validation by microarray analysis

Microarray data were obtained via the EMBL-EBI on-line repository (http://www.ebi.ac.uk/) [22]. We have used the microarray data with the IDs E-MEXP-202 and E-GEOD-2623, which are based on at least two technical replicates per

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

condition. The knock-down genes for E-MEXP-202 were involved in the Nonsense-mediated mRNA decay pathway [23] and E-GEOD-2623 is a set of microarrays obtained after RNAi to target a zinc-finger protein [24]. The sequences of the dsRNAs used is provided in the Supplementary Data. The downloaded raw CEL-data were imported into ArrayAssist 5.5.1 and PLIER normalized. Because our analysis requires all information available at the microarrays, we used a specific approach that allows the evaluation of transcript levels within a large group. First, all transcripts of the whole micro array derived from the dsRNA experiments were divided in two groups: one group representing all upregulated transcripts and another group representing all downregulated transcripts as compared with the control array. The control array contains the expression profile of cells subjected to dsRNA against GFP. This control excludes that observed changes in expression profiles are induced by the introduction of dsRNA itself. A transcript is considered as upregulated in case the mean value of its expression (calculated from the five technical replicates) subtracted by the mean value of the expression levels of this transcript on the control arrays was positive and a transcript is considered to be downregulated in case the mean value of its expression (calculated from the five technical replicates) subtracted by the mean value of the expression levels of this transcript on the control arrays was negative. Please note that for this specification of groups no cut-off values were used. Although this analysis is certainly not appropriate for single probe analysis, this approach does make it possible to gather sufficient information to identify a general trend within the chosen groups as compared to the background. All ratio comparisons were subjected to statistical analysis (see below). For all six experiments the ratio of down versus upregulated transcripts was defined and referred to as the background ratio (presented in Figure 2). RNAiSelect was used to define the predicted set of off-targets and within these sets, the ratio of down versus upregulated transcripts was determined. This ratio was compared with the background ratio (presented in Figure 2).

Statistical analysis

Tables 2 and 3 show a detailed chi-square analysis of the six individual dsRNA experiments from the off-target data predicted by RNAiSelect, with the exception of CG1559 in Table 3 where a two-tailed binomial test was used due to a low transcript number. Table 2 confirms that when considering both introns and exons, the six different dsRNAs show a significant number of off-targets by comparing the observed number of downregulated transcripts with the microarray background ($\alpha = 0.05$). Table 3 shows that when only introns are considered, three out of six analyses still show a significant number of downregulated transcripts ($\alpha = 0.05$). These analyses suggest that intron-based off-targets may indeed occur.

RESULTS

An *in silico* approach based on sequence similarity predicts a high incidence of off-targets when dsRNAs are used to silence gene expression

In *D. melanogaster*, dsRNA molecules of 200–800 bp are commonly used to induce downregulation of genes. From a specific dsRNA, several siRNAs will be formed through the endogenous RNA interference (RNAi) machinery and each siRNA in theory has its own set of potential off-targets. This implies that dsRNA constructs may have multiple off-targets. To investigate this further, we used a previously described method [19] (see Methods section) to analyze the prevalence of off-targets of six specific dsRNA constructs that were used in RNA experiments and for which microarray data are publicly available (see Methods section). Because the dsRNAs are split into siRNAs of ~21 nt by the RNAi machinery, we first created a list of all possible siRNA sequences that can be obtained from the cDNA sequences of each of these six dsRNA constructs. This complete list was subsequently reduced using established scoring rules to exclude 21-bp siRNAs that are most likely non-active (Supplementary Data S1). Our analysis predicted that each dsRNA construct can induce numerous off-target effects based on sequence similarity (Figure 1) and an average of 83 potential off-targets per dsRNA was observed with 0% of them containing zero mismatches, 4–10% containing one or two mismatches, and 90–96% of them containing three mismatches. This *in silico* analysis suggests that these dsRNAs can have multiple off-targets.

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

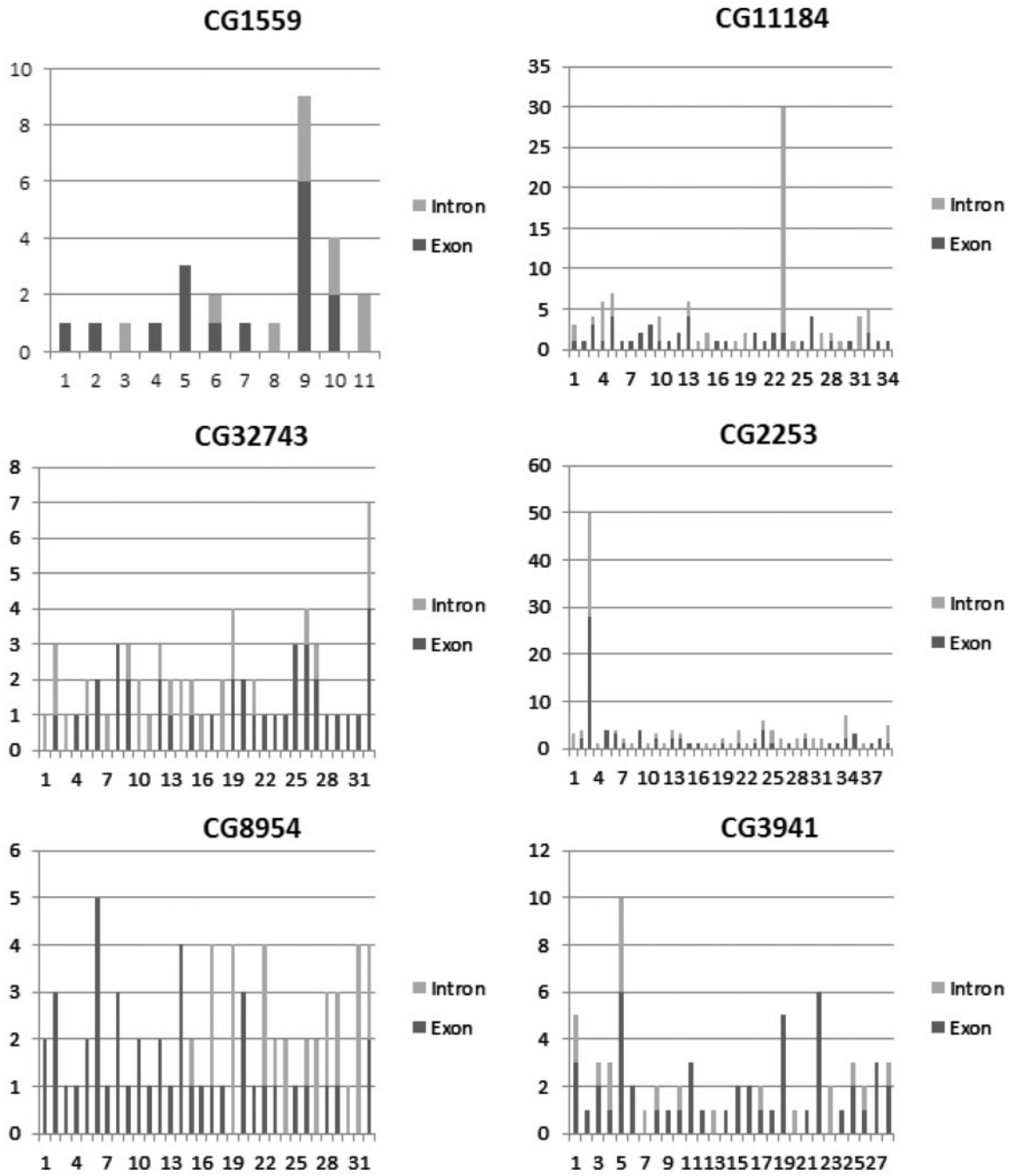

Figure 1: *In silico* analysis predicts a high prevalence of off-targets in RNAi experiments. A previously described tool was used [19] to identify possible off-targets—based on sequence similarity—at the genome-wide level. On the *x*-axis, all the predicted siRNAs derived from the original dsRNA strand using the rules stated in the Supplementary Data are represented as numbers (from the CG1559 dsRNA construct 11 siRNAs were predicted, from the CG11184 dsRNA construct 34 siRNAs were predicted etc). On the *y*-axis, the number of off-targets (that may contain up to three mismatches) found per siRNA is displayed. Dark gray areas represent the number of off-targets found within exons; light gray areas represent the number of off-targets found in introns.

Downloaded from btg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

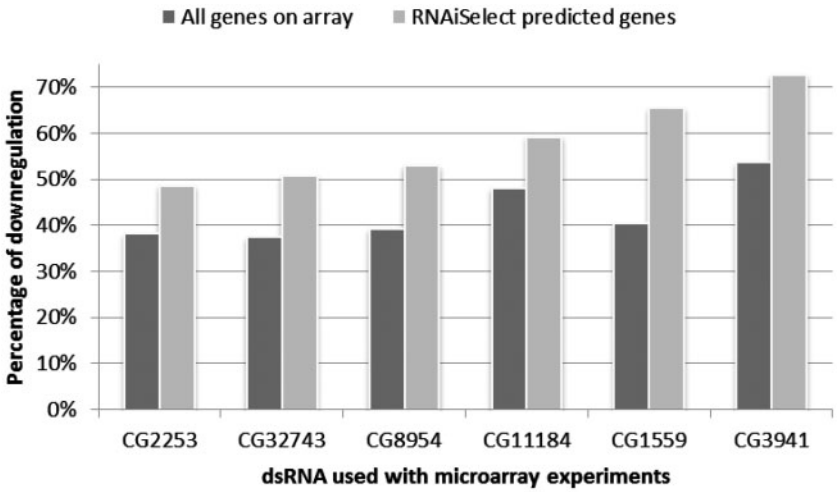

Figure 2: Analysis of existing microarray data reveal that within the set of predicted off-targets, an enrichment is observed of downregulated transcripts after dsRNA treatment. To validate the list of potential off-targets found by our analysis, we compared the number of transcripts downregulated on the whole array with the number of predicted off-targets that were actually downregulated. This was performed for six independent experiments. The dark gray bars represent the percentage of downregulated genes of each experiment on the whole array. The light gray bars represent the percentage of predicted off-targets that are downregulated. This analysis shows that the off-target set predicted by our analysis contains an increased fraction of downregulated genes compared to genes randomly selected from the total array ($P < 0.01$ on average).

Analysis of existing microarray data reveals that an enriched fraction of predicted off-targets is downregulated in RNAi experiments

Next, we used existing biological data to further evaluate the predicted off-targets. Hereto, we analyzed the six independent sets of published microarray data (see Methods section) obtained with the dsRNA constructs. Because our analysis requires all information available at the microarrays, we used the following approach that allows the evaluation of transcript levels within a large group. First, all transcripts of the whole microarray derived from the dsRNA experiments were divided in two groups: one group representing all upregulated transcripts and another group representing all downregulated transcripts as compared with the control array. The control array contains the expression profile of cells subjected to dsRNA against GFP. This control excludes that observed changes in expression profiles are induced by the introduction of any non-specific dsRNA construct. For each dsRNA construct under evaluation at least two independent arrays were available (for CG1559, CG11184, CG32743, CG2253 and for CG8954 $n = 2$; for CG3941 $n = 3$). Internal replicas for each gene on the arrays were used when available (varying between 1 and 5). This mean value was subtracted by the mean value of the expression levels of the transcript on the control arrays. When this subtraction was positive, the transcript was considered to be upregulated and when the subtraction was negative the transcript was considered to be downregulated. Please note that for this specification of groups no cut-off values were used. Although this approach is certainly not appropriate to analyze the expression of single genes, this approach does make it possible to gather sufficient information to identify a general trend within the chosen groups as compared to the background. All ratio comparisons were subjected to statistical analysis (see Methods section). For all six experiments, the ratio of down- versus upregulated transcripts was defined and referred to as the background ratio (presented in Figure 2). For the CG3941 arrays, we analyzed whether there was a correlation between the length of a genes and the probability of being downregulated (Supplementary Data). This analysis demonstrated that based on the complete array there is no difference in length between all downregulated genes versus all non-downregulated genes. Next RNAiSelect [19] was used to define the predicted set of off-targets and within these sets, the

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

Table 1: Intronal sequences may contribute to off-target effects

	CG2253 ($n=142$) (%)	CG32743 ($n=73$) (%)	CG8954 ($n=70$) (%)	CG11184 ($n=93$) (%)	CG1559 ($n=26$) (%)	CG3941 ($n=70$) (%)
Intron + exon (A + B)	10	13	13	14	25	18
Exon only (A)	13	9	14	5	22	13
Intron only (B)	8	19	11	20	30	32

The set of predicted potential off-targets (predicted Set A+B) was split into one set containing exon sequences (predicted Set A) and another set containing intron sequences and sequences overlapping intron/exon boundaries (predicted Set B). Enrichment of downregulated genes within the predicted off-targets compared to the microarray background is presented for each set. All the analyzed microarrays showed an increased downregulated fraction in the intronic set (including the boundaries) compared to the background, three of which were statistically significant (see Methods section). The number of predicted off-targets per dsRNA construct for the six independent experiments is indicated with *n*.

Table 2: Statistical analysis for both intron and exon data in the microarray experiments

Gene	CG2253		CG32743		CG8954		CG11184		CG1559		CG3941	
H0: $\pi =$	0.38		0.37		0.39		0.48		0.4		0.53	
Expression	+	−	+	−	+	−	+	−	+	−	+	−
Expected	88.0	54.0	46.0	27.0	43.3	27.7	55.6	51.4	15.6	10.4	32.9	37.1
Observed	73	69	36	37	34	37	41	66	9	17	20	50
Chi-square	6.761		5.865		5.132		8.025		6.981		9.543	
P-value	0.009		0.015		0.023		0.005		0.008		0.002	

For each dsRNA, we analyzed the predicted off-targets. The second row presents the percentage of genes that are downregulated on the complete microarray. The third row presents the number of predicted off-targets. This number is divided in upregulated and downregulated genes (expected values; fourth row) based on the microarray background (H0; first row). The fifth row presents the actual number of upregulated (+) and downregulated (−) genes within the set of predicted off-targets (observed values). Statistical analysis of these data shows that for every experiment the predicted set of off-targets contains a significant larger fraction of downregulated genes as compared to the complete microarray.

Table 3: Statistical analysis for intron data in the microarray experiments

Gene	CG2253		CG32743		CG8954		CG11184		CG1559		CG3941	
H0: $\pi =$	0.38		0.37		0.39		0.48		0.4		0.53	
Regulation	+	−	+	−	+	−	+	−	+	−	+	−
Expected	43.3	26.7	18.8	11.2	17	11	32.3	29.7	6	4	9.4	10.6
Observed	38	32	13	17	14	14	20	42	3	7	3	17
Chi-square	1.768		4.978		1.424		6.156		[binomial]		11.594	
P-value	0.184		0.026		0.233		0.013		0.1096		0.001	

As in Table 2, we have analyzed the predicted off-targets for each dsRNA, except we now filtered for intron targeted regions. For each dsRNA, we further analyzed the predicted off-targets. The second row shows the percentage of genes that are downregulated on the complete microarray. The third row presents the number of predicted off-targets. This number is divided in upregulated and downregulated genes (expected values; fourth row) based on the microarray background (H0; first row). The fifth row presents the actual number of upregulated (+) and downregulated (−) genes within the predicted set off-targets (observed values). Statistical analysis of these data shows that for every experiment the predicted set of off-targets contains a significant larger fraction of downregulated genes as compared to the complete microarray.

ratio of down- versus upregulated transcripts was determined. This ratio was compared with the background ratio (presented in Figure 2).

These data show that the percentages of downregulation in the predicted sets are 10.48, 13.38, 13.73, 13.71, 25.10 and 17.91% (on average 15.48%) enriched compared to a representative random set of the total array ($P=0.009$, $P=0.015$, $P=0.023$, $P=0.005$, $P=0.008$, $P=0.002$) in studies 1–6, respectively (Figure 2, see also Supplementary Data for an MA plot of the CG3941 data including the predicted off-targets highlighted in yellow). Although this analysis is not suitable to conclude whether specific observed

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

differences within the predicted set of genes are true-off targets or not, this analysis reveals a general trend and suggests that there is an enrichment of downregulated transcripts in the set predicted off-targets as compared to the total array. For the above analysis, off-targets present in exonal and in non-translatable sequences were pooled. Next, we analyzed the occurrence of off-targets in these different target domains separately.

The predicted potential off-targets, were divided in a subset containing exon sequences (Set A) and in another subset containing intron sequences, including sequences overlapping intron/exon boundaries (Set B). The data set containing the intron sequences (Set B) showed the fraction of downregulated genes for all dsRNAs to be enriched compared to the whole data sets, and significantly enriched for three out of six dsRNAs ($P = 0.184$, $P = 0.026$, $P = 0.233$, $P = 0.013$, $P = 0.1096$, $P = 0.001$ in studies 1–6, respectively (Table 1, see Methods section). This finding is consistent with previous findings that there is RNAi activity in the nucleus and specific pre-mRNAs might be exposed to silencing [9, 12]. Examples of predicted off-targets that are strongly downregulated are listed in Supplementary Table S1, many of which are intronic and functionally unrelated to the target gene (Supplementary Table S2 and Table S3 for a list of all predicted siRNAs of the CG3941 dsRNA including their fold downregulation compared to the control array). Although, the 'suspicious' examples as presented in Supplementary Table S1 are predicted off-targets, unrelated to the target gene and strongly downregulated, additional 'wet' experiments (see also Discussion section) are required to investigate whether these effects are due to true off-target effects or are due to biological responses induced by downregulation of the target gene.

DISCUSSION

Our analysis reveals that off-targets may indeed occur in RNAi experiments in which dsRNA constructs are used. These results are consistent with others demonstrating off-target effects in *Drosophila* but also in mammals [1, 4, 25]. There are two explanations why a non-targeted transcript is downregulated as a result of an RNAi experiment: (i) The transcript is a true off-target of the used RNAi constructs, (ii) The downregulated on-target gene triggers a cascade of regulatory effects which result in downregulation of seemingly unrelated gene products. The latter explanation complicates the validation of off-target prediction algorithms and because of this possibility a significantly downregulated transcript does not necessarily represent an off-target effect. It must be noted that the biological consequences induced by the downregulation of the on-target gene will influence the expression of the genes on the whole array (this includes the set of predicted off-targets), whereas off-target effects will only influence the expression values of the genes present within the set of predicted off-targets. Therefore, to enable a proper off-target validation, we compared the percentage of downregulated genes in the predicted set with the percentage of downregulated genes from the total set of genes from the same microarray data. For our analysis, we did not use a specific threshold or cutoff value but we divided the transcripts in two groups: downregulated and not downregulated as compared to the control array. We used this approach for the following reasons. First, our aim was not to identify individual transcripts to be downregulated but instead we were interested in a general trend. Second, it has previously been demonstrated that RNAi can induce off-target effects resulting in <2-fold reduced expression, while still inducing strong protein reduction and subsequently biological effects [26]. Third, by using no threshold, this allows a proper statistical analysis with a large number of transcripts, which would not be possible when filtered groups were used. In addition, by maximizing the sensitivity, small significant expression changes that might have real biological effects are not overlooked. With these considerations, we validated our methods on six experimental data sets from *D. melanogaster*. *In silico*, we predicted the potential off-targets of specific double-stranded RNAs (dsRNAs) and empirically showed that predicted off-target genes were significantly more frequently silenced than other genes. We believe that considering our results, intron containing off-target effects and homologies up to three mismatches should not be ignored. Our analysis reveals predicted off-targets including the number of mismatches between the siRNA and the potential off-target sequence. An assumption is that the off-target RNAi activity is reversely proportional with the number of mismatches. Unfortunately, we were unable to make any statistical distinction between the number of mismatches and the likelihood of being an off-target. This can be explained by the

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

fact that there is a relative low amount of predicted off-targets with less than three mismatches and the majority of predicted off-targets (90–96%) do contain three mismatches.

In our analysis, we searched for predicted off-targets considering 21 bp in length. When 19 bp are considered, which is a relevant length [27], our previously published algorithm [19] predicts 5669 (for the CG2253 dsRNA); 3956 (for the CG32743 dsRNA); 4255 (for the CG8954 dsRNA); 6125 (for the CG11184 dsRNA); 1435 (for the CG1559 dsRNA); and 2456 (for the CG3941 dsRNA) potential off-targets. Although this set is too large to analyze whether there is enrichment of downregulation in the predicted off-target set on the arrays these numbers do further underscore that off-targets (containing three mismatches or less) should not be ignored.

Several tools exist that enable the design of effective RNAi constructs for specific genes of interest. Examples of these tools are: E-RNAi [28], DEQOR [29], siR [30] and SiExplorer [31]. Recently, a tool was developed that enables the design of a complete RNAi library of any annotated genome of interest (RNAiNext [18]). These tools do contain methods that prevent off-targets by avoiding sequences with homologies including introns and UTR's [18], and these methods are mostly based on BLAST. Our previously described tool [19] as we used here to identify off-targets is more sensitive because it enables the identification of all sequences of 21 bp in length, present in the complete genome, that posses a similarity with up to three mismatches. Our analysis does not allow to indentify the most effective RNAi construct, to identify true off-targets, nor to identify the true mechanism behind it (i.e. miRNA like seed complementarity), however our analysis does demonstrate using a sensitive approach that off-target effects should not be ignored.

Several strategies can be followed to perform and control RNAi experiments optimally and keep off-target effects to a minimum. One strategy is to perform two (or more) independent RNAi experiments using two non-overlapping dsRNAi constructs that target the same on-target gene. The idea behind this approach is that two non-overlapping dsRNAi constructs possess sequence similarity to their own specific off-targets but in the meantime will target the same on-target gene. As a result, an induced phenotype shared by the two independent dsRNAi constructs most likely is not an off-target effect but can be considered as a bona fide on-target effect. Although this sounds reasonable, we have recently demonstrated that for every gene (within a set of 99 randomly chosen genes) there are at least two non-overlapping sequences present within its cDNA, that share identical off-targets, based on sequence similarity [19]. As a possible solution to this, we presented a tool to identify a pair of dsRNA constructs that do not share identical off-targets. For every gene of interest within the *Drosophila* genome, these pairs can be searched for [19] (www.RNAiSelect.info). To truly validate the observed biological effects of gene silencing by RNAi technology, it can be tested whether the observed biological effect is rescued by expression of the gene of interest from a related species. Ideally, the sequence of the related species is similar enough to rescue the RNAi-induced phenotype but different enough to be resistant to the used RNAi constructs. Recently a universal method was described for cross-species RNAi rescue in *Drosophila* and it was shown that RNAi induced phenotypes in *D. melanogaster* can be rescued by expression of *Drosophila pseudoobscura* orthologs of the targeted genes [32]. When two 'clean' pairs of independent dsRNA constructs, chosen by RNAiSelect induce a similar phenotype and this shared phenotype is rescued by the cross-species approach, the RNAi-induced phenotype is thoroughly validated and is most likely a bona fide consequence of downregulation of the on-target gene.

SUPPLEMENTARY DATA

Supplementary Data are available online at http://bfgp.oxfordjournals.org/.

Key Points

- Off-target effects can obscure RNAi experiments.
- We provide *in silico* evidence for off-target effects.
- We provide *in vitro* evidence for off-target effects.
- We discuss solutions to the overlooked problem of off-target effects.

FUNDING

This work was supported by a Vidi grant from the Dutch organization for Scientific Research to O.S.

Downloaded from bfg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011

References

1. Kulkarni M, Booker M, Silver S, *et al*. Evidence of off-target effects associated with long dsRNAs in Drosophila melanogaster cell-based assays. *Nat Methods* 2006;**3**:833–8.
2. Moffat J, Reiling JH, Sabatini DM. Off-target effects associated with long dsRNAs in Drosophila RNAi screens. *Trends Pharmacol Sci* 2007;**28**:149–51.
3. Fedorov Y, Anderson EM, Birmingham A, *et al*. Off-target effects by siRNA can induce toxic phenotype. *RNA* 2006; **12**:1188–96.
4. Ma Y, Creanga A, Lum L, Beachy P. Prevalence of off-target effects in Drosophila RNA interference screens. *Nature* 2006;**443**:359–63.
5. Jackson A, Burchard J, Schelter J, *et al*. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 2006;**12**:1179–87.
6. Iyer S, Deutsch K, Yan X, Lin B. Batch RNAi selector: a standalone program to predict specific siRNA candidates in batches with enhanced sensitivity. *Comput Methods Programs Biomed* 2007;**85**:203–9.
7. Naito Y, Yamada T, Matsumiya T, *et al*. dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucleic Acids Res* 2005; **33**:W589–91.
8. Naito Y, Yamada T, Ui-Tei K, *et al*. siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res* 2004;**32**:W124–9.
9. Bosher J, Dufourcq P, Sookhareea S, Labouesse M. RNA interference can target pre-mRNA: consequences for gene expression in a Caenorhabditis elegans operon. *Genetics* 1999;**153**:1245–56.
10. Langlois M-A, Boniface C, Wang G, *et al*. Cytoplasmic and nuclear retained DMPK mRNAs are targets for RNA interference in myotonic dystrophy cells. *J Biol Chem* 2005;**280**: 16949–54.
11. Matzke MA, Birchler JA. RNAi-mediated pathways in the nucleus. *Nat Rev Genet* 2005;**6**:24–35.
12. Robb GB, Brown KM, Khurana J, Rana TM. Specific and potent RNAi in the nucleus of human cells. *Nat Struct Mol Biol* 2005;**12**:133–7.
13. Weinberg MS, Barichievy S, Schaffer L, *et al*. An RNA targeted to the HIV-1 LTR promoter modulates indiscriminate off-target gene activation. *Nucleic Acids Res* 2007;**35**: 7303–12.
14. Morris K, Simon WLC, Jacobsen S, Looney D. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 2004;**305**:1289–92.
15. Chaudhary A, Srivastava S, Garg S. Development of a software tool and criteria evaluation for efficient design of small interfering RNA. *Biochem Biophys Res Commun* 2011;**404**: 313–20.
16. Min W-P, Ichim T, Kim YJ. *RNA Interference*, Vol. 623. New York: Humana Press, 2010, pp. 81–92.
17. Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics* 2009;**10**:392.
18. Horn T, Sandmann T, Boutros M. Design and evaluation of genome-wide libraries for RNA interference screens. *Genome Biol* 2010;**11**:R61.
19. Seinen E, Burgerhof JGM, Jansen RC, Sibon OCM. RNAi Experiments in D. melanogaster: Solutions to the Overlooked Problem of Off-Targets Shared by Independent dsRNAs. *PLoS ONE* 2010;**5**:e13119.
20. Xu D, Landon T, Greenbaum NL, Fenley MO. The electrostatic characteristics of G.U wobble base pairs. *Nucleic Acids Res* 2007;**35**:3836–47.
21. Holen T, Moe S, Sorbo J, *et al*. Tolerated wobble mutations in siRNAs decrease specificity, but can enhance activity in vivo. *Nucleic Acids Res* 2005;**33**:4704–10.
22. Brazma A, Parkinson H. ArrayExpress service for reviewers/ editors of DNA microarray papers. *Nat Biotech* 2006;**24**: 1321–2.
23. Rehwinkel J, Letunic I, Raes J, *et al*. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA* 2005;**11**:1530–44.
24. Page AR, Kovacs A, Deak P, *et al*. Spotted-dick, a zinc-finger protein of Drosophila required for expression of Orc4 and S phase. *EMBO J* 2005;**24**:4304–15.
25. Echeverri CJ, Perrimon N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* 2006;**7**: 373–84.
26. Aleman L, Doench J, Sharp P. Comparison of siRNA-induced off-target RNA and protein effects. *RNA* 2007; **13**:385–95.
27. Ge Q, Ilves H, Dallas A, *et al*. Minimal-length short hairpin RNAs: the relationship of structure and RNAi activity. *RNA* 2010;**16**:106–17.
28. Arziman Z, Horn T, Boutros M. E-RNAi: a web application to design optimized RNAi constructs. *Nucleic Acids Res* 2005;**33**:W58–8.
29. Henschel A, Buchholz F, Habermann B. DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res* 2004;**32**:W113–20.
30. Shah J, Garner H, White M, *et al*. sIR: siRNA information resource, a web-based tool for siRNA sequence design and analysis and an open access siRNA database. *BMC Bioinformatics* 2007;**8**:178.
31. Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi activity. *Nucleic Acids Res* 2007;**35**:e27.
32. Kondo S, Booker M, Perrimon N. Cross-species RNAi rescue platform in Drosophila melanogaster. *Genetics* 2009; **183**:1165–73.

Downloaded from btg.oxfordjournals.org at Rijksuniversiteit Groningen on September 5, 2011