LETTERS

Prevalence of off-target effects in *Drosophila* RNA interference screens

Yong Ma[1], Adrian Creanga[1]†, Lawrence Lum[1]† & Philip A. Beachy[1,2]†

RNA interference (RNAi) in both plants and animals is mediated by small RNAs of approximately 21–23 nucleotides in length for regulation of target gene expression at multiple levels through partial sequence complementarities[1,2]. Combined with widespread genome sequencing, experimental use of RNAi has the potential to interrogate systematically all genes in a given organism with respect to a particular function[3–9]. However, owing to a tolerance for mismatches and gaps in base-pairing with targets[10–12], small RNAs could have up to hundreds of potential target sequences in a genome[13,14], and some small RNAs in mammalian systems have been shown to affect the levels of many messenger RNAs besides their intended targets[15,16]. The use of long double-stranded RNAs (dsRNAs) in *Drosophila*, where Dicer-mediated processing produces small RNAs inside cells, has been thought to reduce the probability of such 'off-target effects' (OTEs)[5]. Here we show, however, that OTEs mediated by short homology stretches within long dsRNAs are prevalent in *Drosophila*. We have performed a genome-wide RNAi screen for novel components of Wingless (Wg) signal transduction[17] in *Drosophila* S2R+ cells, and found few, if any, legitimate candidates. Rather, many of the top candidates exert their effects on Wg response through OTEs on known pathway components or through promiscuous OTEs produced by tandem trinucleotide repeats present in many dsRNAs and genes. Genes containing such repeats are over-represented in candidate lists from published screens, suggesting that they represent a common class of false positives. Our results suggest simple measures to improve the reliability of genome-wide RNAi screens in *Drosophila* and other organisms.

We selected S2R+ cells in preference to several other cell lines (Supplementary Fig. 1) for a high-throughput RNAi screen for novel Wg pathway components, using a 96-well format and a transcriptionally responsive firefly luciferase reporter that contains seven consecutive copies of consensus TCF-binding sites (SuperTopFlash[18]; typically 50–400-fold induction). A control reporter constitutively expressing *Renilla* luciferase was used to normalize for transfection efficiency and to serve as a general indicator of the health of transfected cells[8]. A dsRNA library consisting of >20,000 dsRNAs targeting >95% of the annotated *Drosophila* transcriptional units[19] was screened as pools of three independent dsRNAs. Two hundred and fifty-four pools in the initial screen were identified as reducing Wg-stimulated reporter activities below an arbitrary threshold (Supplementary Fig. 2). Individual dsRNAs within these pools were subsequently screened, with pooled dsRNAs as controls. Five known pathway components, *armadillo* (*arm*), *arrow* (*arr*), *legless* (*lgs*), *pangolin* (*pan*) and *pygopus* (*pygo*), were identified from these pools, validating the efficacy of our screening approach. Seventy dsRNAs that reproduced the effects of their parent pools were selected for further analysis, and these dsRNAs were synthesized *de novo* to ensure uniform yield and quality; their effects on fold induction (ratio of reporter activities in the presence and absence of Wg) spanned the entire spectrum of possible outcomes (Fig. 1 and Supplementary Fig. 3). Exclusion of dsRNAs previously reported to cause growth arrest or cell death[19] left seven (*HDC04705*, *CG31374*, *CG32465*, *CG6834*, *l(1)G0003*, *CG8538* and *CG12993*) that reduced fold induction by more than 2.5 fold; cDNAs for six of these were readily available.

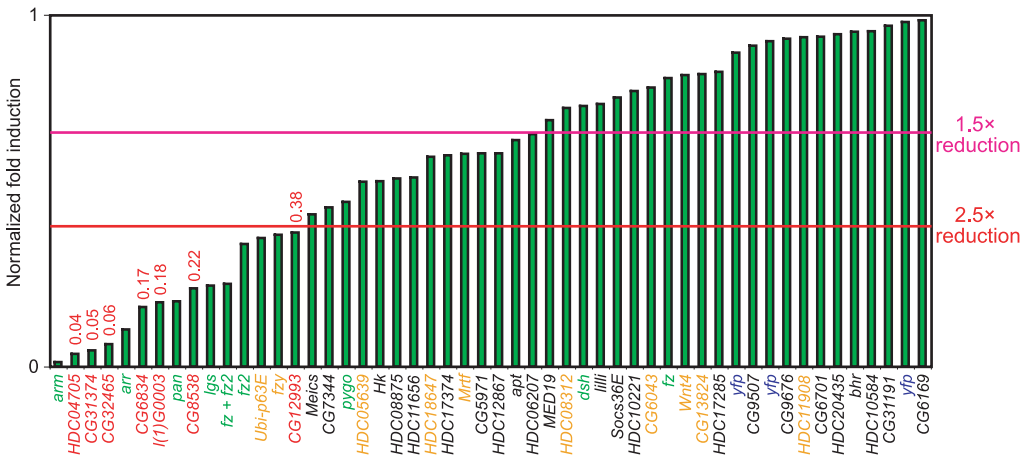

Figure 1 | Candidates for positively acting Wg pathway components. Forty re-synthesized library dsRNAs reduced normalized fold induction (223-fold average). The seven most promising candidates are labelled in red with the effects of their dsRNAs on fold induction indicated above their respective bars. Known pathway components are labelled in green, three independent *yfp* controls in blue, and growth and viability genes that were excluded from the next stage of analysis in orange.

[1]Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, [2]Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. †Present addresses: Department of Cell Biology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, L4.234, Dallas, Texas 75390-9039, USA (L.L.); Department of Developmental Biology and Institute for Stem Cell Biology and Regenerative Medicine, B300 Beckman Building, 279 Campus Drive West, Stanford, California 94305-5329, USA (A.C., P.A.B.).

To confirm the roles of these six genes in Wg signal transduction, additional segments of dsRNA covering the entire lengths of available cDNAs were tested. Notably, the only dsRNAs effective in inhibiting Wg response were those that overlapped with particular regions of the original dsRNAs from the library (Fig. 2a). The dsRNAs that were effective in S2R+ cells, however, were also effective in cl-8 cells (Fig. 2a), confirming that their effects are specific and are not restricted to a particular cell type. To test the possibility that ineffective dsRNAs were not efficiently targeting their cognate genes, expression constructs without untranslated regions (UTRs) were epitope-tagged for those genes with available full-length open reading frames (ORFs): *CG31374*, *CG6834*, *CG8538* and *CG12993*. As expected, neither UTR dsRNAs nor control *yfp* dsRNA affected the levels of proteins transiently overexpressed in S2R+ cells (Fig. 2b). All but one of the remaining 22 dsRNAs significantly reduced levels of their cognate proteins, including those that did not affect the Wg reporter assay. A series of mapping experiments further indicated that the Wg pathway effects of all seven library dsRNAs reside exclusively within 30–40-base-pair (bp) regions (Fig. 2c), suggesting that OTEs mediated by these shorter sequences might be responsible. Consistent with such a hypothesis, expression of the four candidates carrying translationally silent changes to render them resistant to dsRNA treatments failed to rescue Wg reporter activities under these dsRNA treatments (Supplementary Fig. 4).

The 30-bp effective sequence present in *CG32465* was analysed in finer detail by creation of six dsRNA pools, in each of which a group of five consecutive positions was replaced by a mixture of the three base pairs other than that in the original sequence (Fig. 3a). Four dsRNA pools lost the inhibitory ability of the original dsRNA, thus identifying a 20-bp region containing the sequences required for this effect. In each of 20 additional dsRNA pools tested, a single base pair within the effective region was replaced, and a 16-bp region responsible for the OTE was identified by marked reductions in effectiveness of dsRNA pools. A BLAST homology search with this 16-bp sequence identified a perfect match to *arm* mRNA sequence. Notably, all the other six effective dsRNA sequences also contain distinct short homologies to *arm* (Supplementary Fig. 5a), indicating that their effects on Wg reporter activities may result from OTEs on *arm*.

To investigate this possibility, S2R+ cells treated with 30–40-bp dsRNAs (or a 248-bp dsRNA for *CG12993*) were subsequently transfected with a control or Wg expression construct. The levels of *arm* mRNA and Arm protein were reduced in stimulated (Wg-transfected; Fig. 3b) or unstimulated (control-transfected; Supplementary Fig. 6a) cells, consistent with action directed against *arm* expression. In addition, quantitative real-time RT–PCR showed that all seven dsRNAs reduced mRNA levels of the endogenous Wg signalling response target, *naked*[20] (Supplementary Fig. 6b). Each of the six short homologous sequences from *arm* was also inserted immediately after the stop codon of the firefly luciferase ORF under the control of *actin5C* promoter to generate a series of reporter constructs (Supplementary Fig. 5b). All but *l(1)G0003* short dsRNA significantly reduced the activities of their respective reporters, indicating that the short *arm* homologies are sufficient to mediate the observed OTEs (Fig. 3c).

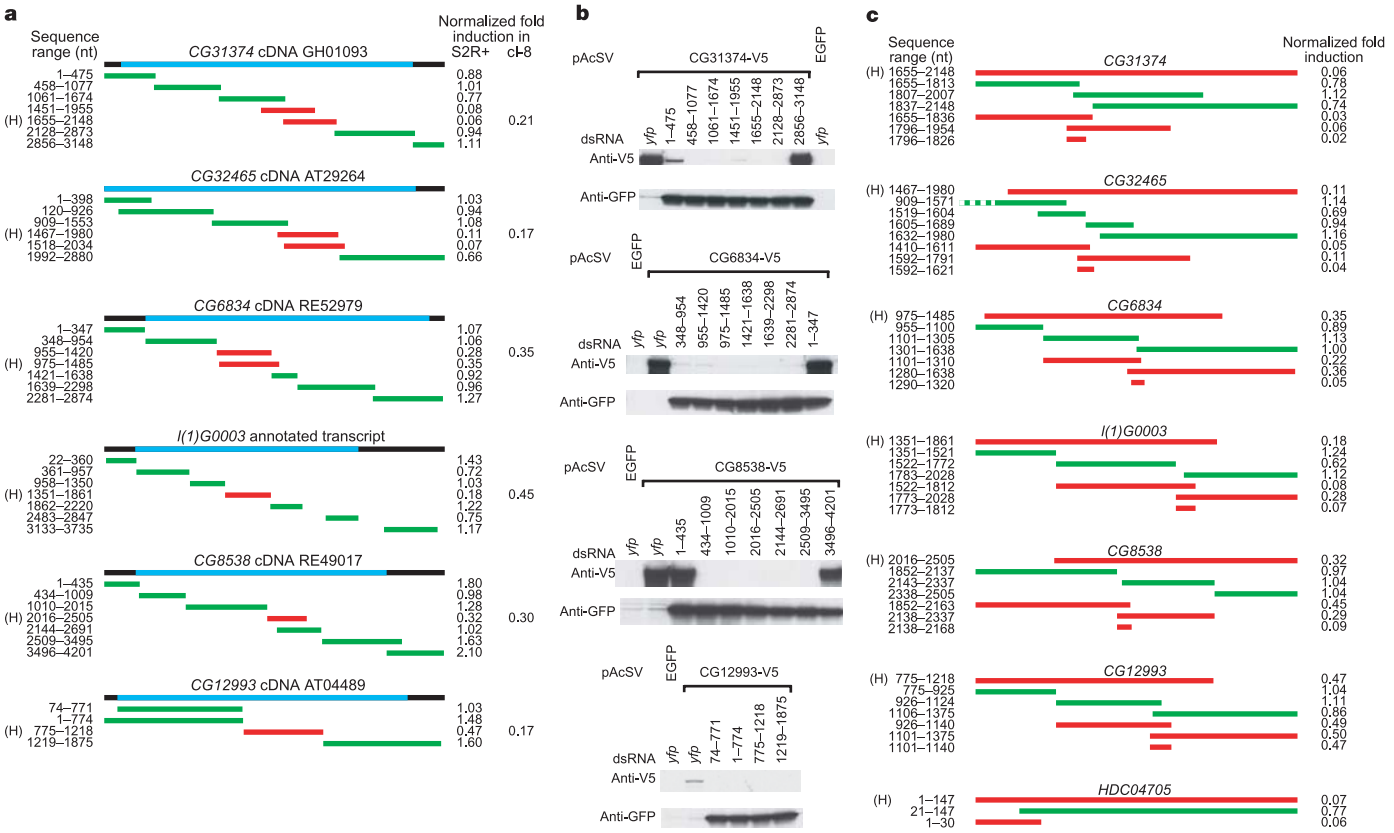

Figure 2 | Suppression of Wg pathway activity by the top seven candidate dsRNAs maps to short ~30–40-bp fragments. a, Only library dsRNAs and some overlapping dsRNAs suppressed Wg reporter activities. cDNAs (annotated transcript for *l(1)G0003*) are represented with ORFs in blue and UTRs in black. dsRNAs with effects similar to those of their corresponding library dsRNAs (marked by H) are represented by red lines; those having no effect or much weaker effects by green lines. The sequence range covered by each dsRNA relative to the respective template is shown on the left; the normalized fold induction produced by each dsRNA is on the right. **b**, V5-tagged expression constructs and a control EGFP construct were transiently transfected into S2R+ cells together with the indicated dsRNAs. All dsRNAs except for *CG8538* (1–435) were effective in knocking down expression of their cognate proteins. **c**, The effective sequences within all seven library dsRNAs are further mapped to small regions of ~30–40 bp in length.

The clear OTEs of the top seven candidate dsRNAs prompted us to ask whether potential OTEs could also account for the effects of other candidate dsRNAs on Wg reporter activities. BLAST searches were performed for the 22 dsRNAs that reduced the normalized fold induction more than 1.5 fold (Fig. 1). The sequences of 19 dsRNAs contain short homologies (<30 bp) to known pathway components (Supplementary Table 1) for which RNAi produced significant inhibitory effects in the pilot experiment (*arm*, *arr*, *frizzled 2* (*fz2*), *lgs*, *pan* and *pygo*; see Supplementary Fig. 1). The three dsRNAs without BLAST-detectable homologies to these known components were classified as 'growth and viability' genes in a previous study[19], raising the question of whether their apparent effects on Wg reporter activities are the result of direct interference with Wg signalling response or indirect effects secondary to their cytotoxicity. Independent dsRNAs that do not overlap with library dsRNAs were tested for an additional seven candidates with readily available cDNAs. With the exception of *fizzy* (*fzy*), one of the growth and viability genes, these non-overlapping dsRNAs did not reproduce the inhibitory effects of their library dsRNAs in a Wg reporter assay (Supplementary Table 2). This behaviour is consistent with the action of the library dsRNAs on Wg reporter activities through OTEs on pathway components. We note that most primary candidates from a genome-wide screen for novel components of the Hedgehog (Hh) signalling pathway also behaved in a similar fashion, suggestive of OTEs (A.C. and P.A.B., data not shown).

The prevalent OTEs observed in our Wg screen prompted us to examine candidates from a similar screen published earlier[21]. Preliminary BLAST searches identified short homologies between many candidate dsRNAs and known pathway components (data not shown). The dsRNA of one such candidate, *CG5402*, seemed to target both *lgs* and *pygo* (Supplementary Fig. 7), consistent with the observed RNAi phenotype. The effective sequences of two other candidates, *CG4136* and *warts* (*wts*), were mapped to regions containing simple tandem repeats of the trinucleotide CAN (N indicates any base, Fig. 4a and Supplementary Fig. 8). A large proportion of reported negative pathway components contains such repeats; our observations suggest that the effects of these dsRNAs on uninduced reporter activities may depend on the duration of dsRNA treatment. Thus, whereas 2-day treatment mildly reduced basal reporter activities, 5-day treatment more dramatically produced the opposite effect, consistent with their proposed negative roles in Wg signal transduction. In contrast to SuperTopFlash activity, levels of the control *Renilla* luciferase reporter were reduced, and this effect was especially pronounced after 5 days of dsRNA treatment, leading to the apparent increase in normalized pathway activity. Thus, extending the time of dsRNA treatment, while presumably allowing for greater turnover of targeted protein, unfortunately also increases the chances of observing OTEs.

To examine further the effects of dsRNAs containing CAN repeats, artificial dsRNAs containing 13 repeats of CAA, CAG, CAC or CAU were produced and tested in the Wg reporter assay. Whereas the $(CAC)_{13}$ and $(CAU)_{13}$ dsRNAs produced only mild effects on basal reporter activities and on control *Renilla* luciferase activities, strong effects were produced by the $(CAA)_{13}$ and $(CAG)_{13}$ dsRNAs, thus essentially recapitulating the effects observed for *CG4136* and *wts* dsRNAs (Fig. 4b). Although the apparent Wg pathway effects of CAN-repeat-containing dsRNAs could, in part, occur through CAN-repeat-containing pathway components such as *dishevelled* (*dsh*), *shaggy* (*sgg*), or *nejire* (*nej*), specific RNAi of these components does

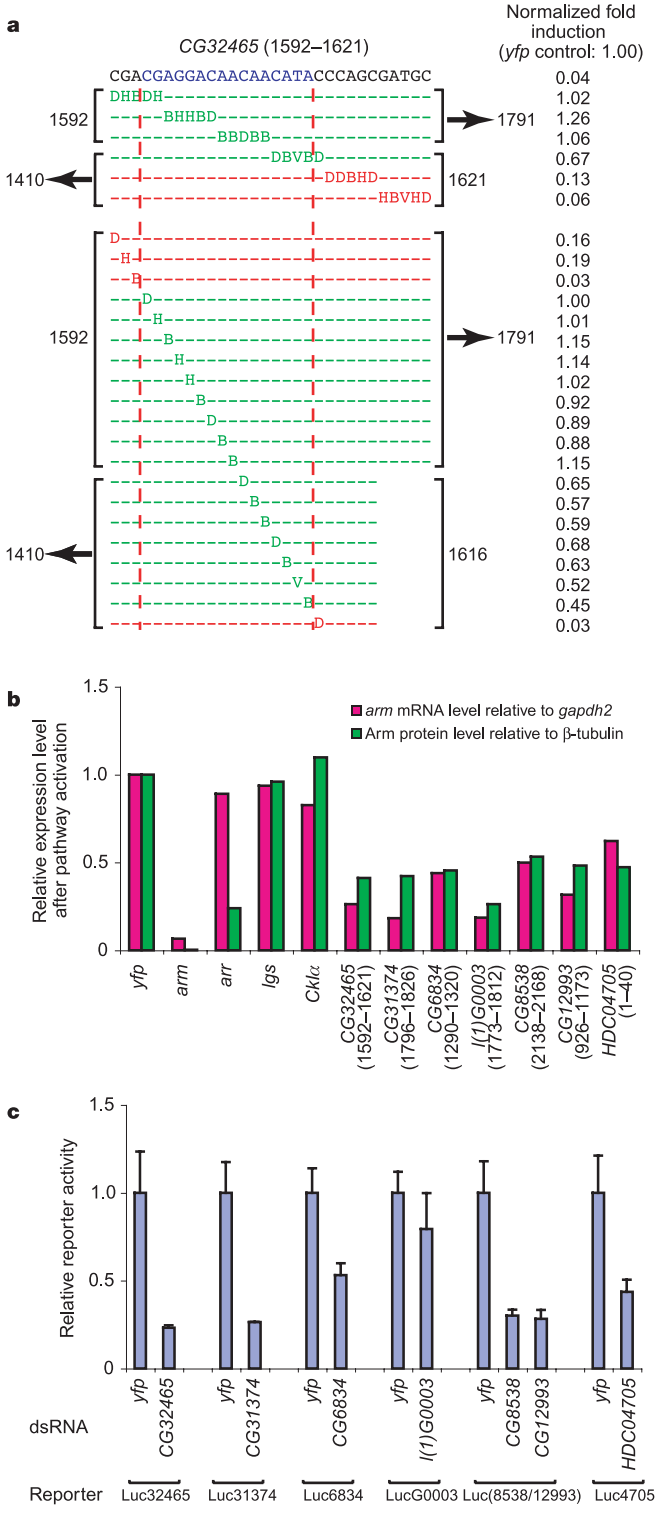

Figure 3 | Short homologies to *arm* account for the activity of top candidates. **a**, Sequence requirement for the Wg pathway effect of *CG32465* dsRNA. Mutational analysis of the 30-bp effective segment within *CG32465* initially identified a 20-bp region (with a series of 5-bp alterations) and subsequently 16 consecutive base pairs (with single-base-pair alterations) as essential for an inhibitory effect on Wg reporter activity. Altered bases within the DNA templates for each pool of dsRNAs are indicated according to the IUPAC code. Numbers flanking the sequences designate the end points of each dsRNA; sequences not shown are represented by arrows. Pools retaining or losing inhibitory effects are shown in red and green, respectively. **b**, Short dsRNAs (and a longer one for *CG12993*) reduced the levels of *arm* mRNA and Arm protein upon pathway activation. RNA levels were determined by quantitative real-time RT–PCR, and protein levels were quantified by NIH Image software analysis of a western blot. **c**, When each OTE luciferase reporter (see text) and a control *Renilla* luciferase reporter were co-transfected into S2R+ cells, five of the six short *arm* sequences were sufficient to mediate the RNAi effects of their respective short dsRNAs. Error bars represent standard deviations of triplicate experiments.

not reduce control *Renilla* luciferase activities (data not shown). The apparent cytotoxicity of CAN repeats thus appears to be due to disrupted expression of other CAN-repeat-containing genes or alternatively, to disruption of other unidentified cellular processes.

We also noted apparent effects of all four types of $(CAN)_{13}$ repeats on Hh induction of normalized reporter activities in Hh signalling assays in cl-8 cells (Fig. 4c), suggesting that the effects of CAN repeats may not be restricted to Wg reporter assays. Indeed, although CAN repeats are only present in ~5% of dsRNAs in the library, CAN-repeat-containing dsRNAs constitute up to 60% or more of the candidate genes listed in several published screens that used this library (Supplementary Table 3). Such enrichment for a large number and high proportion of dsRNAs containing CAN repeats suggests that candidate lists from these and other screens should be treated with caution.

Our genome-wide RNAi screen for positive components of the Wg signalling pathway in *Drosophila* S2R+ cells has identified few novel candidates. We have also reached a similar conclusion with this and other dsRNA libraries in screens for novel components of the Hh pathway (ref. 8 and A.C. and P.A.B., unpublished data). We do not exclude the possibility that *bona fide* pathway components remain to be discovered[22–24], but our results suggest that the number of pathway components not yet identified in mutagenesis screens is limited. Moreover, our study unequivocally demonstrates that long dsRNAs, although remaining highly effective in targeting gene expression, also have the potential to generate phenotypically significant OTEs like those documented for short dsRNAs in mammalian systems[15,16,25,26]. These OTEs are associated with unique sequences such as the *arm* homologies we noted in our Wg candidates or with repetitive simple sequences such as CAN repeats, which may produce pleiotropic effects through promiscuous targeting of related repeats in many gene products. A further complication of dsRNAs containing CAN repeats is that they can cause cytotoxicity, leading to identification of many genes as cell viability hits[19] and consequent exclusion from further consideration as candidates in subsequent screens (see Supplementary Table 3).

Our experience suggests several simple measures that should produce more reliable lists of candidates in RNAi screens that use long dsRNAs. First, libraries should be designed to avoid sequences present in multiple genes, thus preventing the identification of false-positives through promiscuous OTEs. Second, phenotypic effects should be confirmed with more than one non-overlapping dsRNA for each candidate identified. Such testing with multiple dsRNAs would also have the benefit of reducing false positives that can arise due to noise inherent in screens of large numbers of items. Of course, RNAi-based screens at best merely provide a starting point in the identification and further mechanistic study of genetic elements with roles in a biological process of interest.

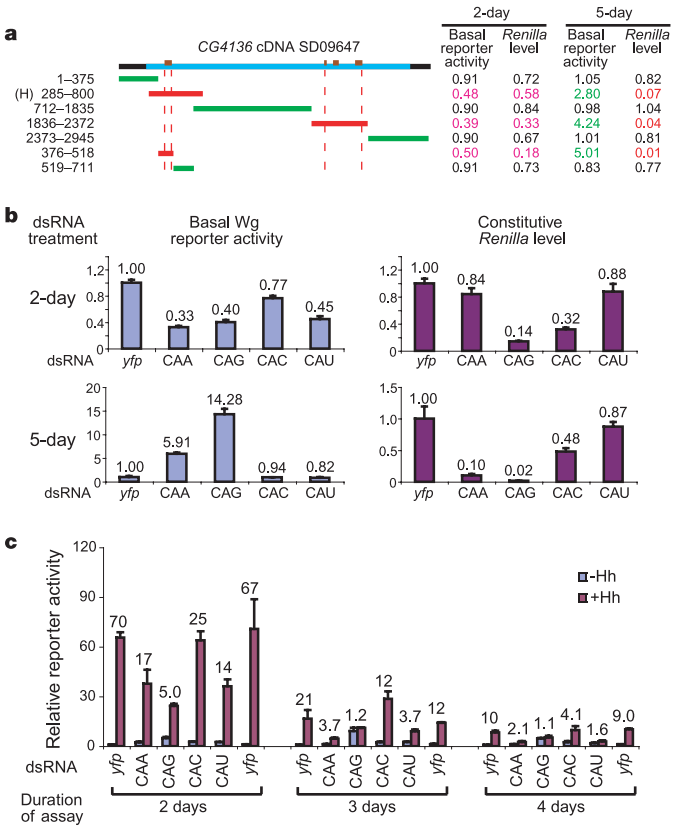

Figure 4 | Effects of CAN-repeat-containing dsRNAs in Wg and Hh reporter assays. a, The effective sequence of *CG4136* dsRNA contains multiple stretches of CAN repeats (defined as ≥6 CAN repeats and shown as short brown lines above the cDNA schematic). Two-day and five-day dsRNA treatments have opposite effects on normalized basal reporter activity, but five-day treatment greatly enhanced the reduction of control *Renilla* luciferase reporter activity. **b**, $(CAA)_{13}$ and $(CAG)_{13}$ dsRNAs exhibited marked effects on normalized basal reporter activities and control *Renilla* luciferase activities, reminiscent of the effects observed with *CG4136* dsRNA. Error bars represent standard deviations of triplicate experiments. **c**, All four $(CAN)_{13}$ dsRNAs increased basal Hh reporter activities in cl-8 cells, whereas $(CAA)_{13}$, $(CAG)_{13}$ and $(CAU)_{13}$ dsRNAs reduced the reporter activities upon Hh stimulation, resulting in an apparent decrease of fold induction (shown above each pair of bars). Note that the longer the treatment was, the less responsive the cells appeared. Error bars represent standard deviations of triplicate experiments.

METHODS

Reagents. The antibodies used in this study are mouse anti-V5 monoclonal antibody (Invitrogen), rabbit anti-Myc antibody A-14 (Santa Cruz Biotechnology), mouse anti-Arm monoclonal antibody N2 7A1, anti-β-Tubulin monoclonal antibody E7 (Developmental Studies Hybridoma Bank), and rabbit anti-GFP antibody (Molecular Probes). Rabbit anti-CG31374 antibodies were generated against AA(1–147) (Spring Valley Lab). The luciferase activities in the cell lysates were measured with Dual-Luciferase Reporter Assay System or Dual-Glo Luciferase Assay System (Promega). Quantitative real-time PCR was performed with iQ SYBR green reagents (Bio-Rad). *Drosophila* cells were transfected with Effectene (Qiagen).

Wingless reporter assay. S2R+ cells were seeded in 96-well (initial screen) or 24-well (secondary screen and all subsequent experiments) plates on day 0. On day 1, six wells of cells were transfected with a single transfection mix containing SuperTopFlash reporter, control *Renilla* luciferase reporter, and various dsRNAs as indicated. On day 2, Wg-expressing and control S2 cells were each added into three wells of transfected S2R+ cells. The luciferase activities of the cell lysates were measured with a BMG FLUOstar OPTIMA luminometer on day 3.

Additional methods can be found in Supplementary Information.

Received 10 May; accepted 17 August 2006.
Published online 10 September 2006.

1. Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).
2. Zamore, P. D. & Haley, B. Ribo-gnome: the big world of small RNAs. *Science* **309**, 1519–1524 (2005).
3. Moffat, J. & Sabatini, D. M. Building mammalian signalling pathways with RNAi screens. *Nature Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
4. Armknecht, S. *et al.* High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392**, 55–73 (2005).
5. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nature Rev. Genet.* **7**, 373–384 (2006).
6. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
7. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
8. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
9. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
10. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).

11. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11,** 599–606 (2004).
12. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3,** e85 (2005).
13. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
14. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769–773 (2005).
15. Birmingham, A. *et al.* 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods* **3,** 199–204 (2006).
16. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21,** 635–637 (2003).
17. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20,** 781–810 (2004).
18. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13,** 680–685 (2003).
19. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303,** 832–835 (2004).
20. Zeng, W. *et al. naked cuticle* encodes an inducible antagonist of Wnt signalling. *Nature* **403,** 789–795 (2000).
21. DasGupta, R., Kaykas, A., Moon, R. T. & Perrimon, N. Functional genomic analysis of the Wnt-wingless signaling pathway. *Science* **308,** 826–833 (2005).
22. Mosimann, C., Hausmann, G. & Basler, K. Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with β-catenin/Armadillo. *Cell* **125,** 327–341 (2006).
23. Banziger, C. *et al.* Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125,** 509–522 (2006).
24. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* **125,** 523–533 (2006).
25. Jackson, A. L. & Linsley, P. S. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20,** 521–524 (2004).
26. Lin, X. *et al.* siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* **33,** 4527–4535 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Perrimon and colleagues for their role in producing the dsRNA library, S. Yao for initial assistance in the Hh screen, and R. Gong for technical assistance. We also thank all members of the Beachy laboratory, R. Green and S. Dorner for helpful discussions, and G. Seydoux and J. Taipale for critical comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.B. (pbeachy@stanford.edu).