DOI: 10.1534/genetics.110.121418

Defining Genetic Factors That Modulate Intergenerational CAG Repeat Instability in *Drosophila melanogaster*

Joonil Jung,[1] Marijn T. M. van Jaarsveld,[2] Shin-Yi Shieh, Kexiang Xu and Nancy M. Bonini[3]

Department of Biology and the Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018

Manuscript received July 27, 2010
Accepted for publication October 25, 2010

ABSTRACT

Trinucleotide repeat instability underlies >20 human hereditary disorders. These diseases include many neurological and neurodegenerative situations, such as those caused by pathogenic polyglutamine (polyQ) domains encoded by expanded CAG repeats. Although mechanisms of instability have been intensely studied, our knowledge remains limited in part due to the lack of unbiased genome-wide screens in multicellular eukaryotes. *Drosophila melanogaster* displays triplet repeat instability with features that recapitulate repeat instability seen in patients with disease. Here we report an enhanced fly model with substantial instability based on a noncoding 270 CAG (*UAS–CAG*$_{270}$) repeat construct under control of a germline-specific promoter. We find that expression of pathogenic polyQ protein modulates repeat instability of CAG_{270} in *trans*, indicating that pathogenic-length polyQ proteins may globally modulate repeat instability in the genome *in vivo*. We further performed an unbiased genetic screen for novel modifiers of instability. These studies indicate that different aspects of repeat instability are under independent genetic control, and identify CG15262, a protein with a NOT2/3/5 conserved domain, as a modifier of CAG repeat instability *in vivo*.

A large number of human hereditary disorders are caused by expansions of simple repeat sequences. Such repeats include CAG repeats as in the polyglutamine (polyQ) diseases (Huntington's disease, and spinocerebellar ataxia types 1, 2, and 3, for example), CTG repeats in myotonic dystrophy type I, and CGG repeats in fragile X syndrome (Gatchel and Zoghbi 2005). In the normal population, the respective repeat sequences are polymorphic in size and are generally stable upon transmission to the next generation. However, longer repeats including ones in the high normal range show instability, with disease occurring when the repeats expand beyond select thresholds. In addition, the hyperinstability seen with expanded repeats from patients shows a strong tendency to further expand in the germ cells, resulting in the transmission of longer repeats to the progeny. This underlies the phenomenon of anticipation, whereby the disease becomes progressively earlier in onset in successive generations due to expansion of the repeat. Since longer repeats generally also cause more severe disease, repeat instability with expansion bias is among the most devastating aspects facing the patients and their families.

Germline-specific processes such as meiotic recombination and epigenetic reprogramming have been proposed to play roles in inducing repeat instability (Jankowski and Nag 2002; Dion *et al.* 2008; Libby *et al.* 2008). In many repeat expansion diseases, germline instability shows gender-specific differences, suggesting that processes specific to male or female germ cell development may differentially affect repeat instability (Mirkin and Mirkin 2007; Wheeler *et al.* 2007). There is also experimental evidence that DNA replication *per se* may be involved (Mirkin and Mirkin 2007). During lagging-strand synthesis, unusual structures such as slipped strands may form, which may result in expansions or contractions in the next replication round. In addition, stalling of the replication fork could result in double-strand breaks (DSBs) or fork reversal, leading to repeat length alterations (Mirkin and Mirkin 2007). DNA repair has also been implicated in repeat instability (Pearson *et al.* 2005). In line with this notion, loss of several DNA repair genes has been shown to significantly modulate repeat instability in various experimental systems (Savouret *et al.* 2003; Pearson *et al.* 2005; Jung and Bonini 2007; Kovtun and McMurray 2008; McMurray 2010).

Repeat instability has been modeled in various organisms including bacteria, yeast, transgenic mice, and mammalian cell lines (Kovtun and McMurray 2008).

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.110.121418/DC1.

[1]*Present address:* Broad Institute, 7 Cambridge Center, Cambridge, MA 02142.

[2]*Present address:* Department of Medical Oncology, Josephine Nefkens Institute, Erasmus MC, 3015 GE Rotterdam, The Netherlands.

[3]*Corresponding author:* Department of Biology, 415 S. University Ave, 306 Leidy Laboratory, University of Pennsylvania, Howard Hughes Medical Institute, Philadelphia, PA 19104. E-mail: nbonini@sas.upenn.edu

We developed a Drosophila model of repeat instability by targeting the expression of a 78-CAG repeat-containing SCA3 transgene (SCA3trQ78) to germ cells. Germline instability is significantly enhanced by expression of the repeat-bearing gene, with the range of expansions and contractions remarkably reminiscent of that seen in human SCA3 patients (Jung and Bonini 2007; Lin and Wilson 2007, 2009). As with the human disease, an ~3:1 bias for repeat expansions over contractions is observed. Furthermore, genes involved in DNA repair (Mus201, an ortholog of human Rad/XPG) and the histone N-acetyltransferase (HAT) protein CBP (CREB-binding protein), modulate repeat instability in Drosophila (Jung and Bonini 2007; Lin and Wilson 2007, 2009).

Drosophila is a powerful model for the discovery of mechanisms of human disease, due to a large range of genetic tools, short generation time, and small genome with a high degree of evolutionary conservation with mammals (Adams and Sekelsky 2002; Bilen and Bonini 2005; Iijima and Iijima-Ando 2008). Here we have developed and utilized a second transgenic fly model for repeat disease with a high rate of instability, to define additional mechanisms. We show that an expanded polyQ repeat may affect the stability of a noncoding repeat in *trans*, suggesting that a common mechanism may regulate the instability of repeats in both protein-coding and noncoding sequences and that one expanded repeat can influence the stability of other expanded repeat sequences within the genome. A large scale genetic screen further identified a protein with a domain homologous to CNOT2 as a modifier of repeat instability *in vivo*.

MATERIALS AND METHODS

Fly lines and husbandry: Fly lines included *UAS–SCA3trQ46* and *UAS–SCA3trQ91* (Jung and Bonini 2007), *UAS–eGFP, nos–GAL4, Act88F–GAL4,* and standard balancer lines. The CAG repeats within the *UAS–CAG$_{270}$* line are located in the 3′-UTR of the DsRed reporter gene (Li *et al.* 2008). *UAS-CAG$_{270}$* flies were backcrossed to *w^{1118}* for five generations to reduce genetic variation due to background. The deficiency lines used were for chromosome 2L (Parks *et al.* 2004). Standard Drosophila culture medium and conditions were used.

Crosses for determining *trans* effects of repeats: Individual male flies bearing *UAS–eGFP, UAS–SCA3trQ46,* or *UAS–SCA3trQ91,* were crossed to recombinant flies expressing the noncoding CAG$_{270}$ in female germ cells (*w; nos–GAL4 UAS–CAG$_{270}$/TM6, Sb*). Individual female flies coexpressing CAG$_{270}$ repeats with either eGFP, SCA3trQ46, or SCA3trQ91 were crossed to *w; Act88F–GAL4* male flies. Flies expressing DsRed in the thorax, thus harboring *UAS–CAG$_{270}$* transgene, were sorted by fluorescence microscopy and the repeat length of the CAG$_{270}$ repeat was determined for 47 progeny flies of each cross, using the Genescan method (below); each experiment was repeated six to eight times. Two-way ANOVA was performed including the rate of no changes, to define the types of changes showing significant variation ($P < 0.0001$; Kruskall–Wallis nonparametric test). Dunn's multiple comparison post hoc test also indicates significance.

Crosses for the screen: Two crossing schemes were used for the screen (see Figure 3). Scheme 1 was used initially; however, the collection of flies required a fluorescence microscope and proved labor intensive, so scheme 2 was developed. In each scheme, in the final cross a total of 47 progeny flies were tested. The female bearing the parental repeat was also collected and analyzed, so that the repeat length of the parent for each individual cross was known. Of the 127 deficiency lines available for chromosome 2L, we were able to collect sufficient progeny from 109 crosses.

Preparing fly genomic DNA: To prepare DNA for Genescan analysis, the parent (1/cross) and offspring (47/cross) flies of the two initial crosses per deficiency line were transferred to a 96-well plate. The flies were homogenized by grinding with a 200-μl natural beveled tip (TipOne, USA Scientific), in 30 μl (males) or 40 μl (females) of buffer (10 mm Tris-HCl, 1 mm EDTA, 25 mm NaCl, pH 8.2), with 200 μg/ml of protease K (Sigma), and incubated on a thermal cycler as follows: 37° (35 min), 85° (2 min), 20° (10 min), and then held at 10°. Midway through the studies, to improve DNA isolation, another 10 μl of buffer with protease K was added and the plates incubated overnight at 4°, or another 10 μl of buffer with protease K was added and the plates incubated as above (37° (35 min), 85° (2 min), 20° (10 min), and then held at 10°). The plates were then incubated at 85° for 2 min to inactivate the protease K and stored at −20° until used for repeat amplification reactions.

Repeat amplification: The CAG-repeat was amplified as follows. Primers used were (forward) 5′/56-FAM/AGG-TTC-CTT-CAC-AAA-GAT-CTT-C-3′ and (reverse) 5′-CAA-CCT-GTT-CCT-GTA-GCT-CG-3′ (IDT, Integrated DNA Technologies). The GC-Rich PCR kit from Roche Scientific was used, with dNTPs from Promega. A master mix with all ingredients except DNA was prepared and aliquoted to the PCR plates (9.3 μl of master mix per well), to which 0.7 μl of individual fly DNA was added. The PCR plate was kept on ice during this process. PCR was performed with the following program: 95° (3 min), then 14 cycles of 94° (30 sec), 63.7° (30 sec), touchdown 0.5° per cycle, 72° (90 sec), then 22 cycles of 94° (30 sec), 56.7° (30 sec), 72° (90 sec), then 72° (10 min), and then 4° (hold). One microliter of select samples were analyzed on a 1.2% agarose gel (with SYBR green 1:10,000); if a band was present indicating that the PCR reaction worked, 1–2 μl of each sample was then submitted to determine repeat size.

Determination of CAG repeat size: CAG repeat-containing amplicons were run on the ABI 3730 DNA sequencing machine (Jung and Bonini 2007). To better resolve long CAG repeats, the following modifications were made to the standard DNA fragment analysis procedure: CAG repeat amplicons were preincubated in formamide along with MapMarker 1000 size standard (Bioventures) for 2 hr prior to analysis and the run temperature was raised to 70°. The number of CAG repeats was determined using the Genescan program (Applied Biosystems) (Jung and Bonini 2007).

The CAG$_{270}$ amplicons gave rise to a group of ~16 peaks due to Taq polymerase stuttering during amplification. The highest peak position fluctuated among multiple repeat experiments of the same sample, while the position of the entire group was reproducible. Therefore, a reproducible criterion for repeat length was established on the basis of the range of the seventh, sixth, and fifth highest peaks; by including detailed analysis of this range of peaks, we established a reproducible repeat length method, which reduced the false positive rate (see supporting information, Figure S1). Only if these three ranges consistently shifted from the parent was a repeat change scored. In rare cases when the repeat size of parents could not be resolved, we used the most frequently detected repeat size from the progeny set as the standard on

the basis of the results that the majority of repeats transmitted from the parent are stable. With these improvements, we amplified and properly resolved CAG repeats from 80–95% of the flies from each cross.

Analysis of repeat length changes: The crosses were analyzed for (1) the overall repeat instability, (2) the number of contractions, +1 expansions, and large (greater than +1) expansions; and (3) the proportion of contractions, +1 expansions and large (greater than +1) expansions relative to the total number of events. The percentage of repeat changes for any one cross is defined as: number of flies with respective repeat changes divided by the total flies scored. For each of these criteria, an estimated 95% confidence interval was calculated on the basis of means of all crosses including the w^{1118} controls (average ±2*standard deviation). Deficiency lines with an average outside the 95% confidence interval (estimated *P*-value <0.05) for any of these analyses were considered lines with potential modification of repeat instability. In addition, lines that showed trends were selected for more detailed analysis. For the selected lines, two more crosses were analyzed to validate the phenotype. Two-tailed unpaired *T*-tests (Excel, Microsoft) and Dunnett's post hoc test (Prism) were used to assess significance. On the basis of this analysis, we defined three deficiency regions of greater interest: Exel6016, Exel6041, and Exel8034.

Defining genes within deficiencies: Available double-stranded (dsRNA) lines for selective knockdown of genes within deficiency Exel8034, and for select genes of special interest within deficiencies Exel6016 and Exel6041 [putative HATs (*CG9486* and *CG10414*), *tos*, and *CG10336*], were obtained from the Vienna Drosophila RNAi Center (Dietzl *et al.* 2007). These lines were crossed in a scheme similar to that of Figure 1, cross scheme 1. The RNAi lines were crossed to *w; UAS-Dicer2/CyO; nos-GAL4 UAS-CAG$_{270}$/Tb*. Ten individual females of genotype *w; UAS-Dicer2/+; Nos-GAL4 UAS-CAG$_{270}$/UAS-RNAi* or *w; UAS-Dicer2/UAS-RNAi; Nos-GAL4 UAS-CAG$_{270}$/+* were crossed to *w; Act88F-GAL4* and progeny flies bearing the repeat (expressing DsRed in the thorax) were collected and frozen at −20°. Genescan analysis was then performed. Initially flies from two crosses for each dsRNA line were examined and for dsRNA lines against three genes (CG9508, CG4161, and CG15262) that passed the initial threshold of significance (Dunnett's post hoc test (Prism) against the rest of the crosses) up to four more crosses were analyzed to confirm significance. To assess whether dsRNA lines had an effect on transgene expression level, three independent replicates of flies were raised at 25°. One-day female fly heads were collected and stored in Trizol reagent (Invitrogen) at −80°. Total RNA was extracted and purified using Trizol reagent. Genomic DNA was removed by TURBO DNA-free (Ambion), and cDNA was synthesized using High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Quantitative RT–PCR was performed using SYBR Green master mix (Applied Biosystems). All RNA samples were analyzed in triplicate and normalized relative to levels of ribosomal protein 49 (Rp49). Primer sets were: rp49 fw 5′-ACGTTGTGCACCAGGAACTT-3′; rp49 rv 5′-CCAGTCGG ATCGATATGCTAA-3′; DsRed fw 5′-CCTCCTCCGAGAACGT CATC-3′; DsRed rv 5′-CCCTCCATGCGCACCTT-3′.

RESULTS

CAG repeat instability is modulated by germline expression in *cis* and by expanded toxic polyQ proteins in *trans*: Previous data suggest that instability of non-coding trinucleotide repeats may be driven by some of the same mechanisms that are responsible for polyQ-encoding CAG repeat instability (Savouret *et al.* 2003; Jung and Bonini 2007). To further test this hypothesis, we developed a new model for instability with a CAG repeat in the 3′-UTR of the transgene, in contrast to a previous polyglutamine-encoding repeat model (Figure 1A). To perform analysis of the repeat instability, we examined the length of the repeats in progeny from flies that either express the transgene in the germline or do not express a GAL4 driver line; thus we are assessing intergenerational germline instability. Interestingly, we observed a substantial level of CAG repeat instability (~14% per transmission, assessing instability in ~50 progeny per transmission, repeating the experiments 6–8 times independently) even without a GAL4 driver line for germline expression of the transgene, indicating long trinucleotide repeats are problematic to cellular machinery and are prone to instability, possibly during replication. Of those changes, ~73% were +1 expansions. Upon expression of the repeats in the female germ cells using *nos–GAL4*, we observed a significantly increased total repeat instability (~35%, $P = 0.017$). Interestingly, the rate of +1 repeat expansions remained relatively unchanged (~10% without GAL4 *vs.* ~13% with nos–GAL4), while the rate of repeat contractions (~1% *vs.* ~9%, $P = 0.021$) and +2 or longer repeat expansions (~3% *vs.* ~13%, $P = 0.028$) were significantly increased upon transcription of the transgene with a driver GAL4 line (Figure 1B).

Previous studies indicated that partial loss of CBP, an important regulator of transcription and DNA repair (Kalkhoven 2004; Smolik and Jones 2007; Das *et al.* 2009), enhances repeat instability (Jung and Bonini 2007). Inhibition of CBP function by pathogenic polyQ protein could enhance instability in a self-amplifying loop, the consequence of compromised CBP function by polyQ proteins. The subsequent effect on repeat instability, however, might not be limited to polyQ-encoding CAG repeats, but may affect the stability of other repeats in *trans*. To test this, we utilized fly lines of different polyQ lengths (SCA3trQ46 *vs.* SCA3trQ91), which were derived from the same initial SCA3trQ78 transgenic line (Jung and Bonini 2007) and thus have identical genomic position effects. While the toxicity of the highly expanded SCA3trQ91 protein was evident when expressed in the eye, the moderately expanded SCAtrQ46 protein conferred little toxicity (Figure 2A). In addition, we used an eGFP-expressing line as an additional control. These lines were independently combined with the *UAS–CAG$_{270}$* line and the consequences of coexpression of expanded polyQ proteins on the instability of noncoding CAG$_{270}$ repeats in *trans* were determined.

Coexpression of the CAG$_{270}$ repeats with eGFP in the female germline resulted in ~28% total repeat changes in the offspring, of which contractions alone contributed ~13%, +1 expansions ~10% and larger expansions ~5% (Figure 2B), which was statistically similar to

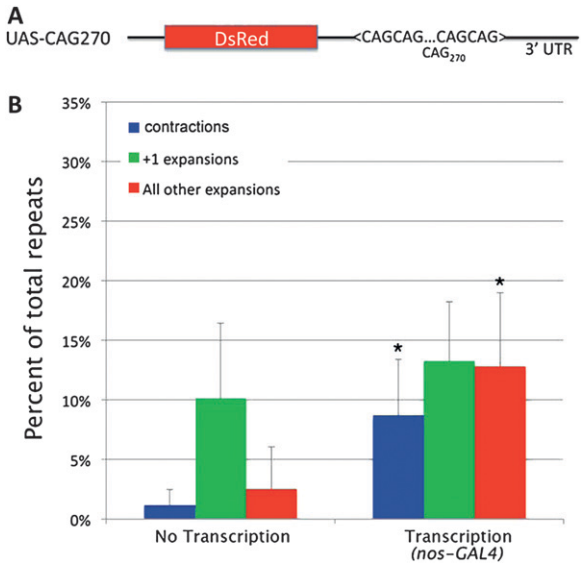

FIGURE 1.—Expression of the transgene in the female germline significantly increases noncoding CAG repeat instability. (A) Schematic of *UAS–CAG*$_{270}$ construct used in the study. The ~270 CAG repeats are located in the 3′-UTR of the DsRed gene. (B) Expression of the transgene by a germ cell-specific GAL4 driver (*nos–GAL4*) significantly enhances repeat instability. Increase in total instability with expression, $P = 0.02$; repeat contractions, $P = 0.02$; expansions greater than +1, $P = 0.03$ (unpaired *t*-tests, two tailed). Error bars, SD.

what was seen in Figure 1B. Coexpression of the CAG270 repeat with a protein bearing a short polyQ repeat (SCA3trQ46) gave rise to similar results in both total repeat instability and types of repeat changes. Coexpression with a longer and more toxic polyQ repeat (SCA3trQ91) did not alter the total rate of noncoding CAG repeat instability significantly (~28% *vs.* ~29%). However, the rate of +1 repeat expansions dramatically increased to ~20% ($P = 0.03$), while the repeat contraction rate dropped from ~13% to less than ~4% ($P = 0.02$). Analysis of the relative abundance, or repeat change bias, indicated that the likelihood of observing a +1 repeat expansion per every incidence of repeat size change increased from ~36% (eGFP coexpression) to ~68% (SCA3trQ91 coexpression; $P = 0.005$). The rate of larger expansions remained the same. These data indicate that the long polyQ protein enhanced instability of the noncoding CAG$_{270}$ repeat in *trans*, in a manner reminiscent of the effect of partial loss of CBP activity on instability of polyQ-coding CAG repeats (see JUNG and BONINI 2007).

A deficiency screen for modifiers of instability: The above data indicated that an expanded repeat within the genome can have an effect on the instability of other repeats. In the course of these studies, we noted that the relatively high rate of repeat instability observed in a single generation with the noncoding CAG$_{270}$ repeat construct (25–35%) made this fly model suitable for a genome-wide screen for modifiers of repeat instability. To perform a screen, we used a deficiency chromosome

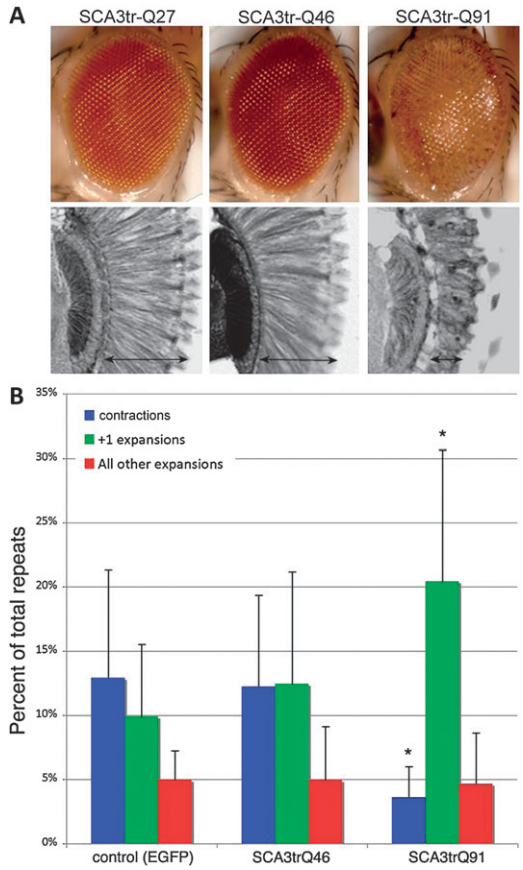

FIGURE 2.—A long polyQ repeat in one gene enhances repeat instability of another gene in *trans*. (A) Difference in toxicity between SCA3trQ46 and SCA3trQ91. One-day flies expressing SCA3trQ27, Q46, or Q91 in the eye (*w; gmr-GAL4/+;UAS-SCA3trQ27/Q46/Q91/+*) were embedded in paraffin and sectioned to determine the degree of retinal degeneration (reverse grayscale autofluorescent images). Flies expressing SCA3trQ46 showed minimal degeneration and were similar to flies expressing normal length SCA3trQ27. Flies expressing SCA3trQ91 had extensive degeneration. SCA3trQ46 and SCA3trQ91 were derived from the same SCA3trQ78 line; SCA3trQ27 is an independent insertion. (B) The length of noncoding CAG repeats in progeny were determined from flies also expressing a control transgene (eGFP), a shorter SCA3 polyQ repeat (SCA3trQ46), or a longer SCA3 polyQ repeat (SCA3trQ91). Flies coexpressing Q91 showed enhanced germline contractions (blue) and +1 class expansions (green) of the *UAS–CAG*$_{270}$ gene (*$P = 0.03$, unpaired *t*-test, two tailed compared to control or SCA3trQ46). Red bars, expansions greater than +1. Errors bar, SD.

collection originally generated by Exelixis (PARKS *et al.* 2004), and focused on chromosome arm 2L. Such a deficiency set allows large genomic regions to be assessed for effects upon 50% reduction in level, with a limited number of crosses; despite the greater degree of instability seen with the CAG$_{270}$ transgene, the approach was still considerably time consuming, requiring individual crosses and assessment of repeat lengths in a 96-well plate of flies. Thus, a screen using deficiencies was a practical option to define novel modifiers and mechanisms of instability. The Exelixis deficiency set has

the added advantage of a homogeneous genetic background, greatly reducing variability in the data due to disparate fly backgrounds. Each deficiency chromosome contains on average an ~140-kb deletion, and so we followed up with the use of dsRNA lines to target individual genes within deficiency regions of interest.

The noncoding CAG_{270} repeat is located in the 3′-untranslated region of a DsRed coding sequence (Li *et al.* 2008). We initially selected for CAG repeat-carrying flies by simply scoring for DsRed expression under a fluorescence dissecting microscope (Figure 3A). We later developed an approach for identifying CAG repeat-carrying flies by selecting against visible markers (Figure 3B). We used both approaches to scan chromosome arm 2L with similar results. Five crosses for each deficiency line were set up by mating individual female flies with three male flies, and two crosses were analyzed for each deficiency line initially.

All parameters of repeat instability were scored; the rate of total repeat instability, as well as the number of contractions and expansions, the latter of which were divided into +1 expansions and larger expansions (Figure 4). The overall rate of repeat instability was 15–25%, with varying proportions of contractions and expansions. Most repeat size changes involved a small number of repeats, with +1 repeat expansions observed most frequently. Repeat changes of ±10 repeats were found in crosses of 43 deficiency lines, with 8 of them displaying two or more such events. In particular, crosses with the Exel7078 had at least one progeny with long repeat changes from all four crosses examined. The largest repeat expansion was +43 repeats, while the biggest contraction involved a 156-repeat shrinkage. Overall, long repeat contractions were three times more common than long repeat expansions.

To identify potential modifiers of repeat instability, we calculated a 95% confidence interval on the basis of controls and selected lines with averages outside this interval for further focus. On the basis of this criterion, we did not find any lines with a lower or higher rate of total repeat instability. However, examination of individual parameters of repeat instability identified 11 lines with a change in an absolute or relative number of contractions: Exel8034, Exel8041, Exel7078, Exel9043, and Exel6021; +1 expansions: Exel6041 and Exel8005; and greater than +1 expansions: Exel8012, Exel6016, Exel7023, and Exel6027. Six other lines (Exel6005, Exel6256, Exel7006, Exel7015, Exel7022, and Exel8033) showed positive trends, indicating that analysis of additional crosses was warranted (Figure 4, see Figure S2 and Table S1). For these potentially interesting deficiencies, additional crosses were analyzed. Statistically significant differences from the averages of all crosses for the following deficiency lines were found: Exel8034, where +1 expansions were reduced; Exel6041, where +1 expansions were enhanced; and Exel6016, where large expansions were increased (Figure 5).

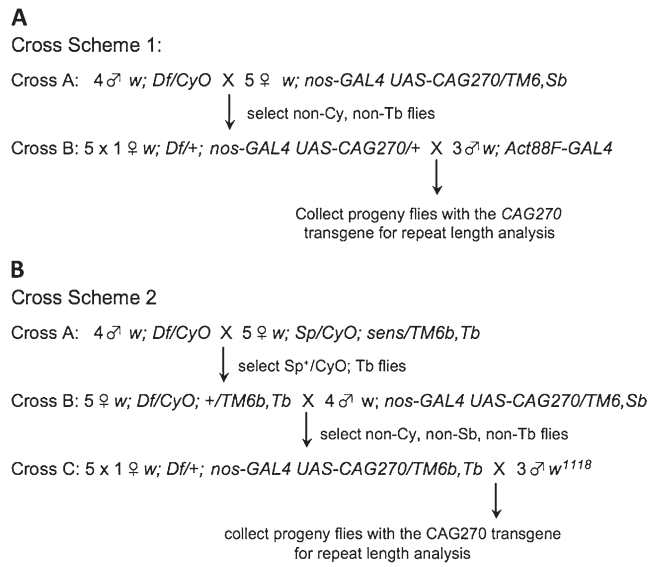

Figure 3.—Fly cross schemes for genetic deficiency screen. (A) In cross scheme 1, progeny flies bearing CAG repeats were selected by fluorescence by the screening for flies expressing DsRed from the CAG repeat transgene in thoracic muscle with the *Act88F–GAL4* driver. About 50 crosses were performed in this manner. (B) In cross scheme 2, a balancer on the third chromosome was introduced into the deficiency lines, to allow selection of progeny of cross C by selecting against the *TM6b, Tb* balancer. About 70 crosses were performed in this manner.

Detailed analysis of select deficiency lines: To identify specific genes responsible for the modifier effect within the deficiencies of interest, we employed dsRNA lines targeting genes deleted within these deficiencies. Fourteen genes in 175 kb are deleted in Exel8034. Four genes have a predicted role in the regulation of gene transcription, 2 others in protein translation or protein degradation, and the functions of the remaining 8 are unknown (Table 1). Exel6041 has a 160-kb deletion removing 27 genes, and Exel6016 has a 158-kb deletion removing 18 genes. The functions of the genes in both of these deficiencies range from ion transport, cytoskeletal organization, and protein modifications to regulation of transcription and DNA repair (Table 1). In the two latter regions, of special interest were *CG9486* and *CG10414*, which might function as HATs. Like the HAT CBP, a known modifier of repeat instability (Jung and Bonini 2007), they may influence the stability of repeats. In addition, genes *CG10336* (Tweedie *et al.* 2009) and *CG10387* (*tosca*) (Digilio *et al.* 1996) have been implicated in DNA repair, a process thought to contribute to repeat instability.

The deficiency screen tested 50% loss of gene expression. On the basis of the assumption that a greater loss of gene expression might result in a stronger effect, we coexpressed Dcr-2 with the dsRNA constructs to enhance the loss of function (see materials and methods; (Dietzl *et al.* 2007)). We used essentially the same cross scheme as scheme 1 in Figure 3A by crossing individual

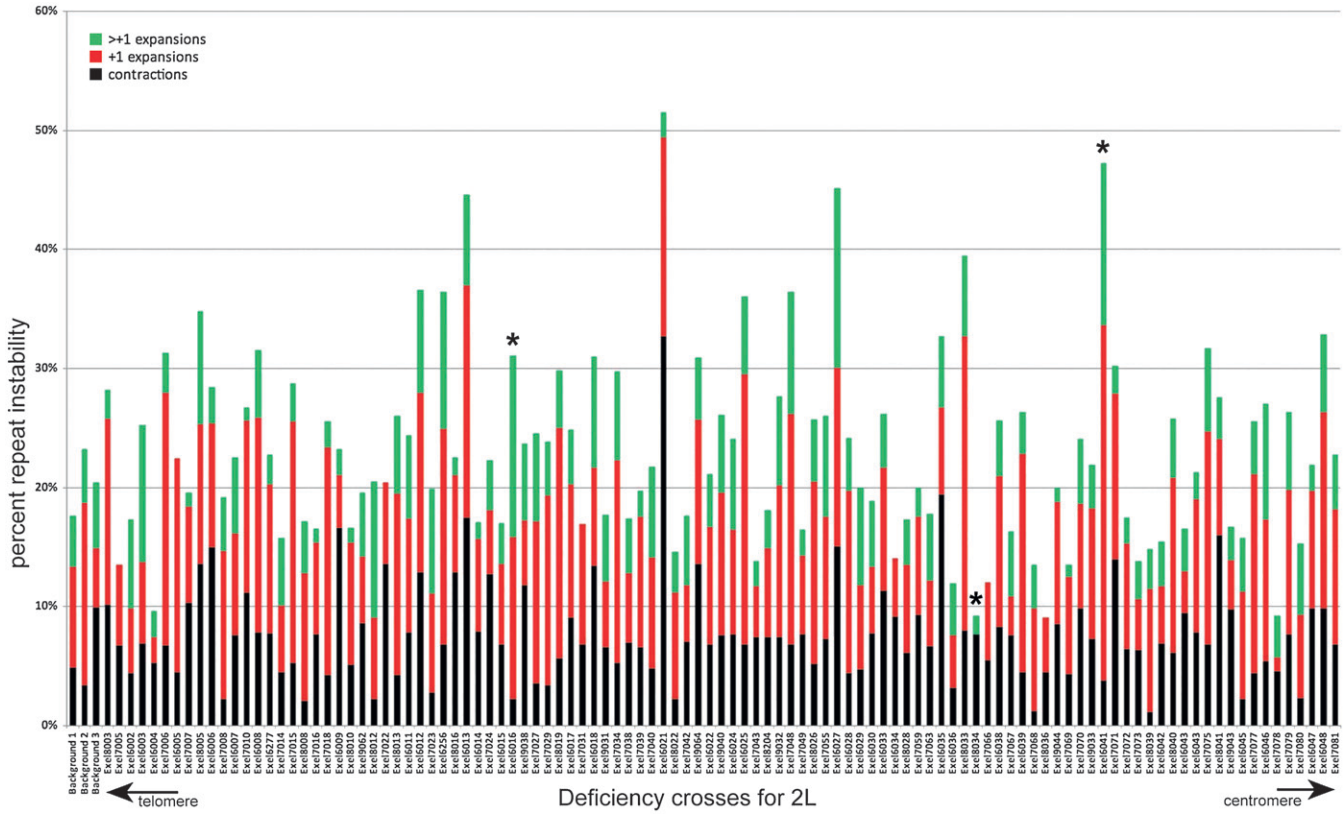

FIGURE 4.—Consolidated summary of entire set of deficiency lines screened for repeat changes. There were three potentially different genetic backgrounds, which did not differ significantly in their repeat instability and are aligned on the left. Asterisks highlight the crosses with deficiency lines (Exel6016, Exel8034, and Exel6041) with statistically significant changes after examination of 5+ individual crosses and that were the subject of further analysis.

female flies expressing Dcr-2, CAG_{270}, and a dsRNA construct in the germline with 3–5 male *Act88F–GAL4* flies. Progeny flies positive for the repeat were selected and the repeat size was compared with the parental repeat. Samples with significant changes ($P < 0.05$) after examining the progeny from two crosses were further tested with progeny of up to four more crosses.

We tested 13 dsRNA lines targeting genes within the deficiency region of line Exel8034, as well as several dsRNA lines to genes of interest with the other deficiencies (see Table S2). Loss of several genes (*CG15260*, *CG11861*, and *CG10336*) showed trends, with selective loss of *CG15262* within Exel8034 causing a significant ($P < 0.05$) effect and selective decrease in the rate of +2 or longer repeat expansions ($P = 0.02$; Figure 6). Knockdown of CG15262 had no effect on the expression of the transgene (see Figure S3), indicating that reduction of the gene is affecting features of instability and not the level of transcription. The *CG15262* gene has a domain with homology to mammalian CNOT2, which is part of the CCR4–NOT complex. Studies in yeast suggest that this complex is an important regulator of gene expression, and functions include negative regulation of transcription, stimulation of transcription elongation, and promotion of mRNA degradation (COLLART 2003). In addition, several studies suggest that the complex might have a role in DNA repair. Thus, the activity of a counterpart protein may play a role in repeat instability to promote repeat expansions.

DISCUSSION

We present a Drosophila model of trinucleotide repeat instability using a long noncoding CAG repeat. An advantage of this system is the enhanced instability observed after a single generation compared to a previous model (JUNG and BONINI 2007). We provide support for the hypothesis that toxic polyQ protein can enhance repeat instability, showing that expression of SCA3trQ91 enhances the rate and proportion of +1 repeat expansions of CAG_{270}. This indicates that a toxic polyQ protein can affect the stability of long repeats elsewhere within the genome in a *trans*-acting manner. We further present data on an unbiased genetic screen for modifiers of CAG repeat instability, which revealed a fly gene containing a CNOT2 domain as a modifier of expansions.

Instability of noncoding repeats and *trans* interactions due to the polyQ protein: Many human diseases are caused by the expansion of small repeat sequences within genes, including myotonic dystrophy, fragile X syndrome, as well as the class of diseases known as the

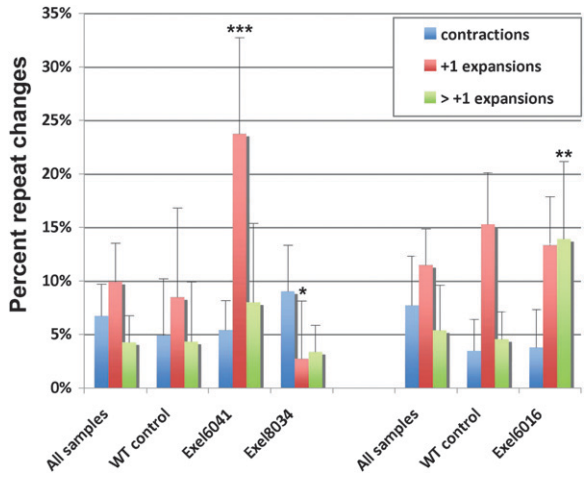

Figure 5.—Three deficiency regions with changes in repeat instability. The percentage of offspring exhibiting contractions, +1 expansions and greater than or equal to +2 expansions of the maternally inherited CAG_{270} repeats. The deficiencies are grouped separately, comparing with the average of all crosses and the wild-type (WT) control for the same genetic background. Asterisks indicate percentages significantly different from the average of all samples ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (unpaired *t*-test, two tailed) compared to average of all crosses for that background). Error bars, SD.

polyQ diseases, like Huntington's disease (Gatchel and Zoghbi 2005). Our studies have focused on instability of CAG repeat sequences. Unlike previous studies with a shorter repeat size (SCA3trQ78), the current studies employed ~270 CAG repeats (*UAS–CAG*$_{270}$). With the longer repeat, we observed a significant level of intergenerational repeat instability even in the absence of added gene expression (see Figure 1). This basal level instability showed primarily +1 repeat expansions, consistent with the "replication slippage" model (Kovtun and McMurray 2008). When the CAG repeat was then expressed in germ cells, there was a significant increase in the overall rate of repeat instability, similar to previous findings (Jung and Bonini 2007; Lin and Wilson 2007). Interestingly, however, when each subtype of repeat change was examined, no significant changes in +1 expansions were observed; rather, the frequency of contractions and +2 or longer expansions increased. This suggests that expression of the transgene may not significantly affect replication slippage in cells, but instead it may enhance instability through alternate processes. These data are in agreement with previous studies suggesting that transcription-coupled nucleotide excision repair (TC-NER) may enhance repeat instability. It is also possible that the TC-NER process can physically or functionally interact with replication machinery and thereby lead to repeat instability (Lin *et al.* 2009).

Previous findings indicate that inhibition of CBP/HAT activities by expanded polyQ protein enhances repeat instability (Jung and Bonini 2007). Given this, we hypothesized that the effect of highly expanded polyQ protein on DNA repair or replication proteins should influence repeat instability of *UAS–CAG*$_{270}$ in *trans*. Indeed, this study showed a significant increase in +1 repeat expansions when the highly toxic SCA3trQ91 was coexpressed (see Figure 2). These data further support the hypothesis that polyQ protein toxicity enhances repeat instability (Jung and Bonini 2007). Surprisingly, however, we did not observe an outright increase in the overall rate of total repeat instability, as seen with genetic reduction of CBP activity (Jung and Bonini 2007). Rather expanded polyQ protein expression may favor a process that drives +1 repeat expansions of noncoding CAG repeats at the expense of repeat contractions. Although a repeat change of +1 may seem small in the context of the fly with its short lifespan, this may translate into a greater change in the context of humans. Humans show both germline instability as well as somatic instability (McMurray 2010); instability may become worse with age or accumulate. We have occasionally (rate of ~0.1%) seen evidence of somatic instability in flies, when a progeny fly is assessed for repeat length and two significant lengths are apparent, despite the fact that the fly should bear only a single repeat. As the peaks are typically of equal size, this may reflect early somatic mosaicism.

In addition to histones (Downs 2008; Van Attikum and Gasser 2009), HAT proteins such as CBP and p300 are known to directly acetylate several DNA repair proteins or function as transcriptional coactivators (Pao *et al.* 2000; Tini *et al.* 2002; Bhakat *et al.* 2006; Smolik and Jones 2007; Stauffer *et al.* 2007). Interestingly, flap structure-specific endonuclease 1 (FEN1) haplo-insufficiency in a Huntington's disease (HD) model of CAG_{120} repeats resulted in a similar pattern of intergenerational repeat size changes: a reduction in deletions and an increase in small repeat expansions (Spiro and McMurray 2003). Acetylation of FEN1 by p300 leads to reduced endonuclease activity in human cells (Hasan *et al.* 2001). FEN1 acetylation may modulate its activity (Friedrich-Heineken *et al.* 2005), which may in turn affect the function in DNA processivity, and thereby impact repeat instability. While the exact mechanisms by which polyQ protein enhances repeat instability remain to be fully elucidated, that several DNA repair and replication related proteins implicated in instability are regulated either by CBP or related HAT proteins makes this an interesting area of future research (Spiro and McMurray 2003; Yang and Freudenreich 2007; Entezam and Usdin 2008; Kovtun and McMurray 2008).

A screen for modifiers of repeat instability: Although candidate genes or pathway-based approaches to probe mechanisms of repeat instability have been productive, unbiased genetic screens are instrumental in advancing our understanding of many processes including human disease (Adams and Sekelsky 2002;

TABLE 1

Overview of the genes within each deficiency region

Genomic region	FlyBase ID	Name	Potential function
Exel8034	*CG3758*	*escargot (esg)*	Regulation of transcription
	CG4158	*worniu (wor)*	Regulation of transcription
	CG4161		Unknown
	CG11861	*guftagu (gft)*	Ubiquitin protease ligase
	CG15258		Unknown
	CG15259	*no hitter (nht)*	Regulation of transcription
	CG15260		Unknown
	CG15261	*UK114*	Protein folding; regulation of translation
	CG15262		Regulation of transcription
	CG15263		Unknown
	CG18482		Unknown
	CG31733	*ms(2)35Ci*	Unknown
	CG31828		Unknown
	CG31829		Unknown
Exel6041	*CG10336*		replication-fork protection (DNA repair)
	CG10338		Unknown
	CG10341		Unknown
	CG10343		Unknown
	CG10346	*Grip71*	Mitotic spindle organization
	CG10348		Regulation of translation
	CG10369	*Inwardly Rectifying Potassium Channel 3 (Irk3)*	Potassium ion transport
	CG10372	*Fas Associated Factor (Faf)*	Unknown
	CG10373		Unknown
	CG10376		Protein phosphatase
	CG10383		Unknown
	CG10385	*male-specific lethal 1 (msl-1)*	X chromosome dosage compensation
	CG10387	*tosca (tos)*	Exodeoxyribonuclease (DNA repair)
	CG10391	*Cyp310a1*	Wnt signaling
	CG10393	*absent MD neurons and olfactory sensilla (amos)*	Regulation of transcription
	CG10413		Sodium/potassium chloride transporter
	CG10414		N-acetyltransferase
	CG15160		Unknown
	CG15161		Cilium assembly
	CG15162	*Misexpression suppressor of Ras 3 (MESR3)*	Unknown
	CG31751		Unknown
	CG31789		Unknown
	CG31789		Unknown
	CG31790		Unknown
	CG31791	*Male-specific transcript 36b (Mst36Fb)*	Phosphoenolpyruvate phosphotransferase
	CG31801	*Male-specific transcript 36a (Mst36Fa)*	Unknown
	CG42490		Unknown
Exel6016	*CG9486*		N-acetyltransferase
	CG9491	*Gef 26*	Guanine Nucleotide Exchange factor
	CG9493	*Pez*	Protein phosphatase
	CG9497		Unknown
	CG9498		Unknown
	CG9499	*pickpocket 7 (ppk7)*	Sodium channel
	CG9500		Signal transduction; transport
	CG9501	*pickpocket 14 (ppk14)*	Sodium channel
	CG9505		Metalloendopeptidase
	CG9506	*slow as molasses (slam)*	Migration, cellular organization
	CG9507		Metalloendopeptidase
	CG11567	*Cytochrome P450 reductase (Cpr)*	NADPH reductase
	CG13983		Unknown
	CG13984		Unknown
	CG33531	*Discoidin domain receptor (Ddr)*	Transmembrane receptor
	CG42368[a]		Unknown
	CG42369[a]		Unknown
	CG42370[a]		Unknown

[a]The gene formerly known as *CG9508* is now split up in coding regions *CG42368, CG42369* and *CG42370.*

Bilen and Bonini 2005; Iijima and Iijima-Ando 2008; Lessing and Bonini 2009). Whereas novel modifiers of repeat instability have been identified using yeast (Bhattacharyya *et al.* 2002), no such screen has been reported in a multicellular model organism. By combining a set of deficiency lines (Parks *et al.* 2004) with dsRNA stocks (Dietzl *et al.* 2007), we screened the effect of 50% loss of ~2000 genes in a time- and cost-effective manner. We note that although the screen was possible, it was still labor intensive. Future efforts directed toward the genetic engineering of constructs that would allow more efficient screens would be of benefit. Importantly, however, these and earlier studies establish that Drosophila does display many complex aspects of repeat instability making it an appropriate and important model system for investigating this process.

Of the 109 deficiency lines tested, three lines, Exel6016, Exel6041, and Exel8034 had consistent effects on repeat instability. Because different aspects of repeat instability could be under independent genetic control, we analyzed repeat contractions, +1 repeat expansions, and +2 or greater repeat expansions. Interestingly, the three modifier lines all showed different effects, reinforcing the hypothesis that multiple pathways or genes with different specificity may be involved in repeat instability. A similar complexity in repeat instability has been suggested in mammalian systems (Dragileva *et al.* 2009).

Among 109 deficiency lines tested, Exel7078 was the only one that showed repeat change events of ±10 repeats (−31, −22, −17, +13) from all crosses examined. In patients, long noncoding trinucleotide repeats often result in massive repeat size changes in intergenerational transmissions, contrary to polyQ coding repeats (Brouwer *et al.* 2009). The molecular mechanisms responsible for such large repeat changes, which can sometimes add or delete more than hundreds of repeats in a single generation, is poorly understood. Very long expansions of polyQ-coding CAG repeats may also occur in the germline, but due to the expected extreme toxicity of such polyQ proteins, these repeats may not be successfully transmitted to the next generation. Under such an hypothesis, the polyQ protein toxicity may be dictating the upper limit of expandability, not the lack of specific molecular mechanisms necessary for long repeat expansion. On the other hand, it is possible that very long repeat changes seen with noncoding trinucleotide repeats may be due to specific molecular mechanisms yet to be identified.

While we observed a 2:1 expansion bias on average from the deficiency crosses, when only the ±10 repeat expansions or contractions are assessed, we saw the opposite trend with large repeat contractions being more prevalent (3:1). It is not clear whether this is due to the limitation of the technique (PCR and capillary DNA sequencing machine-based repeat size determination) or a genuine biological phenomenon. Polyacrylamide gel electrophoresis (PAGE)-based systems lack the ability to resolve small differences in repeat size, but have been successfully used to demonstrate long repeat expansions involving up to hundreds of repeat expansions (Kennedy and Shelbourne 2000). It will be important to combine a PAGE-based method with a multigeneration cross scheme to address this issue.

A protein with homology to CNOT2 modulates repeat instability: The dsRNA-based system, used to pinpoint a specific gene responsible for the effect of the Exel8034 deficiency on instability, did not identify a gene that precisely recapitulated the effect of the deletion: the deletion caused a reduction in +1 expansions, whereas knockdown of the *CG15262* gene caused reduction in expansions greater than +1. There are a number of possible explanations for this discrepancy. For example, the effect of the deficiency may reflect the collective results of loss of multiple genes from the region. Alternatively, the effect of the gene on repeat instability may be sensitive to gene dosage, such that loss of one copy of the gene *vs.* more severe knockdown by RNAi could produce different outcomes (Dietzl *et al.* 2007).

The dsRNA-based secondary screen of the genes in Exel8034 led to the identification of *CG15262* as a modifier of repeat instability. Loss of this gene results in a significantly lower proportion of +2 and greater expansions, whereas the number of contractions and +1 expansions were relatively unaffected. Interestingly, CG15262 has homology to mammalian CNOT2, a component of the CCR4–Not complex, which has not been previously implicated in repeat instability. In yeast, the complex consists of Ccr4, Caf1, Caf40, Caf130, and Not1-5, and the components are conserved in fruit flies and mammals (Collart 2003; Temme *et al.* 2004). Originally, the complex was identified as a transcriptional regulator in yeast (Liu *et al.* 1998), but subsequently, the complex has been implicated in diverse processes such as transcriptional elongation (Denis *et al.* 2001), mRNA degradation, protein degradation (Collart 2003), and apoptosis (Shi and Nelson 2005).

The CCR4–Not complex may also be involved in the response to DNA repair in yeast (Westmoreland *et al.* 2004; Mulder *et al.* 2005; Deshpande *et al.* 2009). Induction of DNA damage by hydroxyurea (HU) and methylmethane sulfate (MMS) stimulates transcription of the ribonucleotide reductase (RNR) complex in a CCR4–NOT-dependent fashion (Mulder *et al.* 2005; Woolstencroft *et al.* 2006). The RNR enzymes stimulate the production of dNTPs, needed for DNA replication and repair. Further, a recent report shows that CCR4–NOT mutants have an impaired TC-NER (Gaillard *et al.* 2009). If the Drosophila *CG15262* gene acts in a similar way, defects in TC-NER might explain the

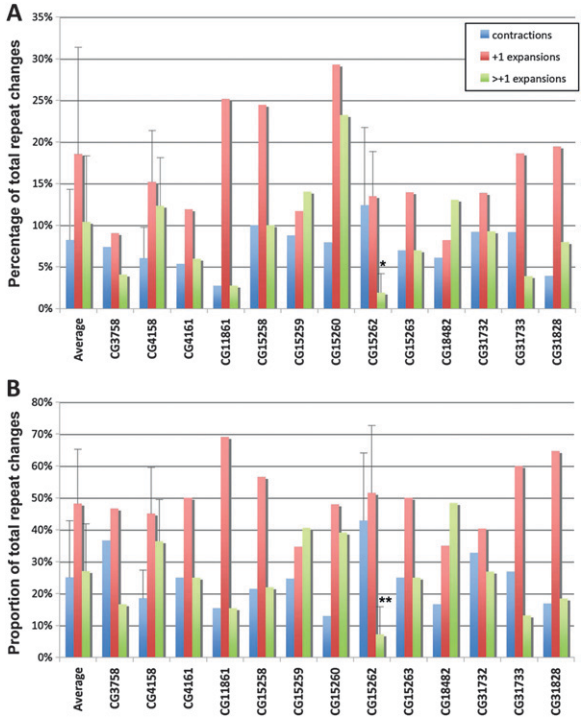

FIGURE 6.—Knockdown of *CG15262* modifies repeat instability. Data from RNAi crosses, targeted to genes within the deficiency region of Exel8034. *CG31732*, just outside the gene region of Exel8034, was included as a random gene control. (A) Comparison of the rates of repeat instability in the three categories revealed that repeat expansions greater than +1 were significantly rarer when *CG15262* was targeted with a dsRNA construct. (B) Relative abundance of repeat expansions greater than +1 was significantly lower when *CG15262* was knocked down. Asterisks indicate percentages significantly different from the average of all deficiency lines ($*P < 0.05$, $**P < 0.01$, unpaired *t*-tests, two tailed).

phenotype of the *CG15262* knockdown. Interestingly, a deletion in Mus201 (ortholog of Rad2/XPG), an endonuclease crucial for NER and/or TCR, drastically reduced the number of repeat expansions in the female germ cells as well (JUNG and BONINI 2007). Although further research is needed, our data indicate that *CG15262* and the CCR4–NOT complex may be a novel component of repeat instability in parent-to-offspring long CAG repeat transmissions.

Owing to its medical significance and biological novelty, mechanisms of trinucleotide repeat instability have been intensely studied using systems ranging from bacteria to patients. Despite this, our understanding of the process remains limited. As in so many other processes, Drosophila is likely to provide an important tool to further our understanding in the mechanisms of repeat instability, whether in more targeted candidate-gene–based approaches or in unbiased genetic screens.

We thank Derek Lessing for critical comments and the Bloomington Drosophila Stock Center for fly lines. This work was supported in part by grants from the National Institutes of Health (5R01-NS043578 and P01-AG009215). N.M.B. is an investigator of the Howard Hughes Medical Institute.

LITERATURE CITED

ADAMS, M. D., and J. J. SEKELSKY, 2002 From sequence to phenotype: reverse genetics in *Drosophila melanogaster*. Nat. Rev. Genet. **3:** 189–198.

BHAKAT, K. K., S. K. MOKKAPATI, I. BOLDOGH, T. K. HAZRA and S. MITRA, 2006 Acetylation of human 8-oxoguanine-DNA glycosylase by p300 and its role in 8-oxoguanine repair in vivo. Mol. Cell. Biol. **26:** 1654–1665.

BHATTACHARYYA, S., M. L. ROLFSMEIER, M. J. DIXON, K. WAGONER and R. S. LAHUE, 2002 Identification of RTG2 as a modifier gene for CTG*CAG repeat instability in *Saccharomyces cerevisiae*. Genetics **162:** 579–589.

BILEN, J., and N. M. BONINI, 2005 Drosophila as a model for human neurodegenerative disease. Annu. Rev. Genet. **39:** 153–171.

BROUWER, J. R., R. WILLEMSEN and B. A. OOSTRA, 2009 Microsatellite repeat instability and neurological disease. Bioessays **31:** 71–83.

COLLART, M. A., 2003 Global control of gene expression in yeast by the Ccr4-Not complex. Gene **313:** 1–16.

DAS, C., M. S. LUCIA, K. C. HANSEN and J. K. TYLER, 2009 CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature **459:** 113–117.

DENIS, C. L., Y. C. CHIANG, Y. CUI and J. CHEN, 2001 Genetic evidence supports a role for the yeast CCR4-NOT complex in transcriptional elongation. Genetics **158:** 627–634.

DESHPANDE, G. P., J. HAYLES, K. L. HOE, D. U. KIM, H. O. PARK *et al.*, 2009 Screening a genome-wide S. pombe deletion library identifies novel genes and pathways involved in genome stability maintenance. DNA Repair (Amst.) **8:** 672–679.

DIETZL, G., D. CHEN, F. SCHNORRER, K. C. SU, Y. BARINOVA *et al.*, 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature **448:** 151–156.

DIGILIO, F. A., A. PANNUTI, J. C. LUCCHESI, M. FURIA and L. C. POLITO, 1996 *Tosca:* a Drosophila gene encoding a nuclease specifically expressed in the female germline. Dev. Biol. **178:** 90–100.

DION, V., Y. LIN, L. HUBERT, JR., R. A. WATERLAND and J. H. WILSON, 2008 Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. Hum. Mol. Genet. **17:** 1306–1317.

DOWNS, J. A., 2008 Histone H3 K56 acetylation, chromatin assembly, and the DNA damage checkpoint. DNA Repair (Amst.) **7:** 2020–2024.

DRAGILEVA, E., A. HENDRICKS, A. TEED, T. GILLIS, E. T. LOPEZ *et al.*, 2009 Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol. Dis. **33:** 37–47.

ENTEZAM, A., and K. USDIN, 2008 ATR protects the genome against CGG.CCG-repeat expansion in Fragile X premutation mice. Nucleic Acids Res. **36:** 1050–1056.

FRIEDRICH-HEINEKEN, E., M. TOUEILLE, B. TANNLER, C. BURKI, E. FERRARI *et al.*, 2005 The two DNA clamps Rad9/Rad1/Hus1 complex and proliferating cell nuclear antigen differentially regulate flap endonuclease 1 activity. J. Mol. Biol. **353:** 980–989.

GAILLARD, H., C. TOUS, J. BOTET, C. GONZALEZ-AGUILERA, M. J. QUINTERO *et al.*, 2009 Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. PLoS Genet. **5:** e1000364.

GATCHEL, J. R., and H. Y. ZOGHBI, 2005 Diseases of unstable repeat expansion: mechanisms and common principles. Nat. Rev. Genet. **6:** 743–755.

HASAN, S., M. STUCKI, P. O. HASSA, R. IMHOF, P. GEHRIG *et al.*, 2001 Regulation of human flap endonuclease-1 activity by acetylation through the transcriptional coactivator p300. Mol. Cell **7:** 1221–1231.

IIJIMA, K., and K. IIJIMA-ANDO, 2008 Drosophila models of Alzheimer's amyloidosis: the challenge of dissecting the complex mechanisms of toxicity of amyloid-beta 42. J. Alzheimers Dis. **15:** 523–540.

JANKOWSKI, C., and D. K. NAG, 2002 Most meiotic CAG repeat tract-length alterations in yeast are SPO11 dependent. Mol. Genet. Genomics **267:** 64–70.

JUNG, J., and N. BONINI, 2007 CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. Science **315:** 1857–1859.

KALKHOVEN, E., 2004 CBP and p300: HATs for different occasions. Biochem. Pharmacol. **68:** 1145–1155.

Kennedy, L., and P. F. Shelbourne, 2000 Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? Hum. Mol. Genet. **9:** 2539–2544.

Kovtun, I. V., and C. T. McMurray, 2008 Features of trinucleotide repeat instability *in vivo*. Cell Res. **18:** 198–213.

Lessing, D., and N. M. Bonini, 2009 Maintaining the brain: insight into human neurodegeneration from *Drosophila melanogaster* mutants. Nat. Rev. Genet. **10:** 359–370.

Li, L. B., Z. Yu, X. Teng and N. M. Bonini, 2008 RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature **453:** 1107–1111.

Libby, R. T., K. A. Hagerman, V. V. Pineda, R. Lau, D. H. Cho *et al.*, 2008 CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet. **4:** e1000257.

Lin, Y., and J. H. Wilson, 2007 Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Mol. Cell. Biol. **27:** 6209–6217.

Lin, Y., and J. H. Wilson, 2009 Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. DNA Repair (Amst.) **8:** 878–885.

Lin, Y., L. Hubert, Jr. and J. H. Wilson, 2009 Transcription destabilizes triplet repeats. Mol. Carcinog. **48:** 350–361.

Liu, H. Y., V. Badarinarayana, D. C. Audino, J. Rappsilber, M. Mann *et al.*, 1998 The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. EMBO J. **17:** 1096–1106.

McMurray, C. T., 2010 Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. **11:** 786–799.

Mirkin, E. V., and S. M. Mirkin, 2007 Replication fork stalling at natural impediments. Microbiol. Mol. Biol. Rev. **71:** 13–35.

Mulder, K. W., G. S. Winkler and H. T. Timmers, 2005 DNA damage and replication stress induced transcription of RNR genes is dependent on the Ccr4-Not complex. Nucleic Acids Res. **33:** 6384–6392.

Pao, G. M., R. Janknecht, H. Ruffner, T. Hunter and I. M. Verma, 2000 CBP/p300 interact with and function as transcriptional coactivators of BRCA1. Proc. Natl. Acad. Sci. USA **97:** 1020–1025.

Parks, A. L., K. R. Cook, M. Belvin, N. A. Dompe, R. Fawcett *et al.*, 2004 Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster genome*. Nat. Genet. **36:** 288–292.

Pearson, C. E., K. Nichol Edamura and J. D. Cleary, 2005 Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet. **6:** 729–742.

Savouret, C., E. Brisson, J. Essers, R. Kanaar, A. Pastink *et al.*, 2003 CTG repeat instability and size variation timing in DNA repair-deficient mice. EMBO J. **22:** 2264–2273.

Shi, J., and M. A. Nelson, 2005 The cyclin-dependent kinase 11 interacts with NOT2. Biochem. Biophys. Res. Commun. **334:** 1310–1316.

Smolik, S., and K. Jones, 2007 Drosophila dCBP is involved in establishing the DNA replication checkpoint. Mol. Cell. Biol. **27:** 135–146.

Spiro, C., and C. T. McMurray, 2003 Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. Mol. Cell. Biol. **23:** 6063–6074.

Stauffer, D., B. Chang, J. Huang, A. Dunn and M. Thayer, 2007 p300/CREB-binding protein interacts with ATR and is required for the DNA replication checkpoint. J. Biol. Chem. **282:** 9678–9687.

Temme, C., S. Zaessinger, S. Meyer, M. Simonelig and E. Wahle, 2004 A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in Drosophila. EMBO J. **23:** 2862–2871.

Tini, M., A. Benecke, S. J. Um, J. Torchia, R. M. Evans *et al.*, 2002 Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol. Cell **9:** 265–277.

Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. McQuilton *et al.*, 2009 FlyBase: enhancing Drosophila gene ontology annotations. Nucleic Acids Res. **37:** D555–D559.

van Attikum, H., and S. M. Gasser, 2009 Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol. **19:** 207–217.

Westmoreland, T. J., J. R. Marks, J. A. Olson, Jr., E. M. Thompson, M. A. Resnick *et al.*, 2004 Cell cycle progression in G1 and S phases is CCR4 dependent following ionizing radiation or replication stress in Saccharomyces cerevisiae. Eukaryot. Cell **3:** 430–446.

Wheeler, V. C., F. Persichetti, S. M. McNeil, J. S. Mysore, S. S. Mysore *et al.*, 2007 Factors associated with HD CAG repeat instability in Huntington disease. J. Med. Genet. **44:** 695–701.

Woolstencroft, R. N., T. H. Beilharz, M. A. Cook, T. Preiss, D. Durocher *et al.*, 2006 Ccr4 contributes to tolerance of replication stress through control of CRT1 mRNA poly(A) tail length. J. Cell Sci. **119:** 5178–5192.

Yang, J., and C. H. Freudenreich, 2007 Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. Gene **393:** 110–115.

Communicating editor: J. Sekelsky

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.110.121418/DC1

Defining Genetic Factors That Modulate Intergenerational CAG Repeat Instability in *Drosophila melanogaster*

Joonil Jung, Marijn T. M. van Jaarsveld, Shin-Yi Shieh, Kexiang Xu and Nancy M. Bonini

DOI: 10.1534/genetics.110.121418

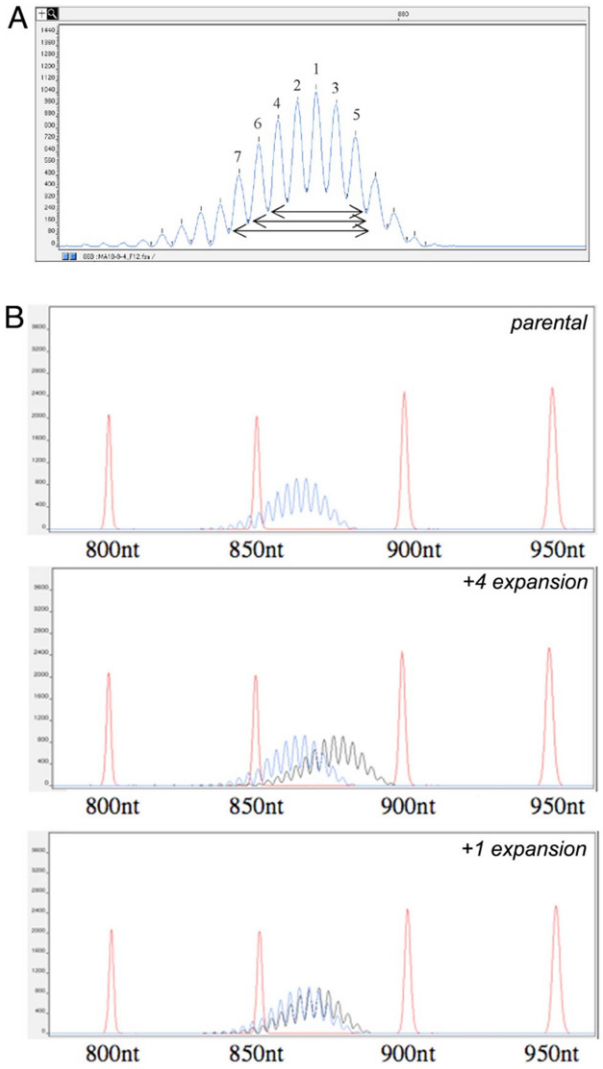

FIGURE S1.—Genescan analysis of the CAG_{270} repeat. (A) Genescan™ analysis of a typical CAG_{270} repeat. The CAG-repeat gave rise to a pattern of peaks, each differing in size by ~3 nucleotides (1 repeat), a likely result of Taq polymerase stuttering in the amplification reaction. A change in repeat length resulted in a shift of the entire pattern to a size position distinct from the parental. The pattern of the 7 (bottom arrow), 6 (middle arrow) and 5 (top arrow) peaks were carefully assessed in order to determine whether a fly had undergone a repeat change from the parental. (B) Example of repeat expansions in progeny flies, compared to the parental (top). Large repeat changes are obvious (middle panel, +4 expansion), and smaller changes are also readily and reliably detected (bottom panel, +1 expansion).

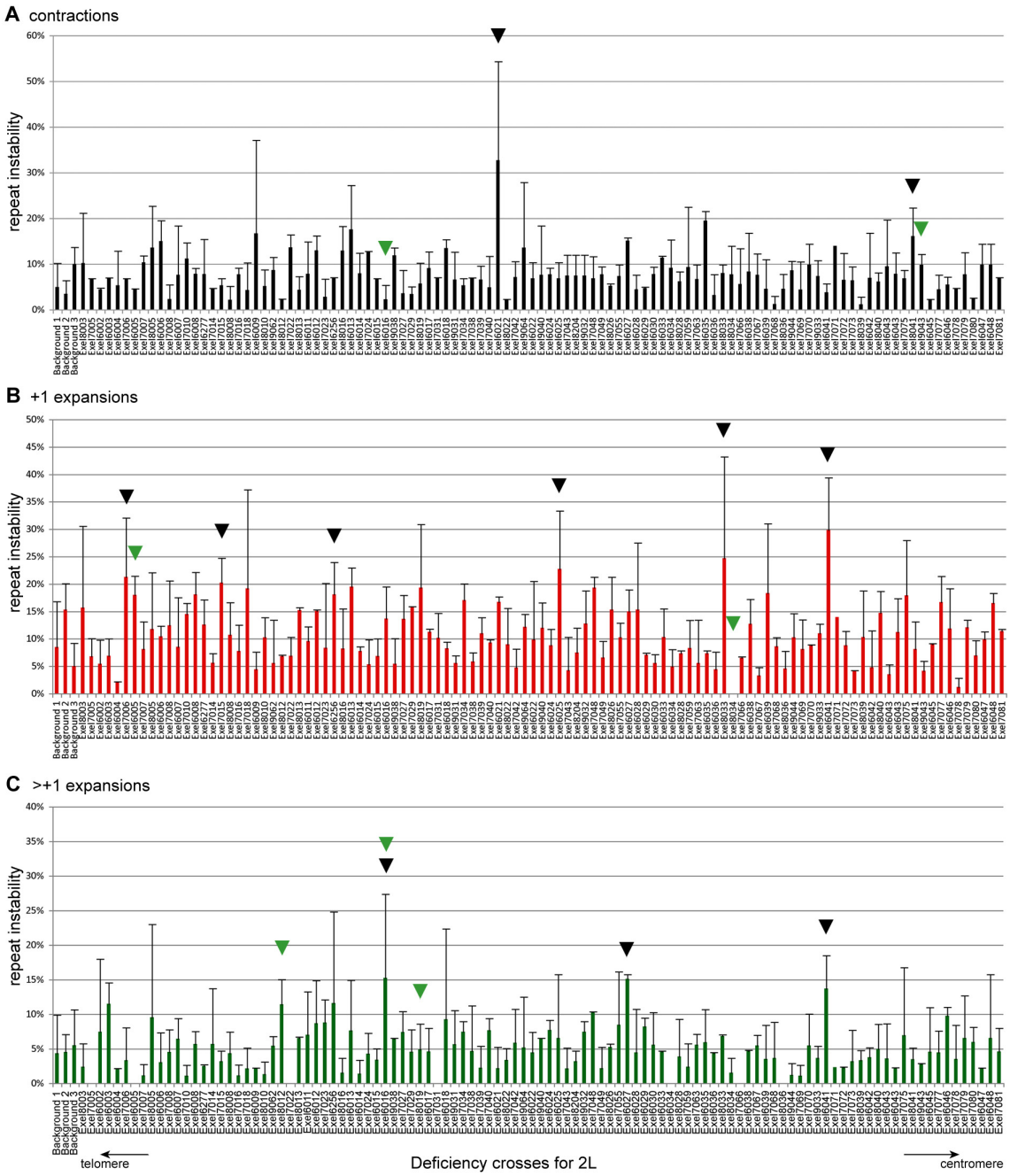

FIGURE S2.—Initial screen data, showing breakdown of expansions. Screen for modifiers of repeat instability during maternal transmission of a long non-coding CAG_{270} repeat. (A) contractions (B) +1 expansions (C) >+1 expansions. Depicted in to the left are controls (w^{1118}) in each specific background used within the screen. Arrowheads point to the lines that were selected for further analysis in three categories of repeat instability (contractions, +1 repeat expansions, and expansions >+1). Black arrowheads indicate lines further examined for possible changes in the rate of instability, while green arrowheads indicate lines for examined for possible changes in the relative proportion of each category of repeat changes.

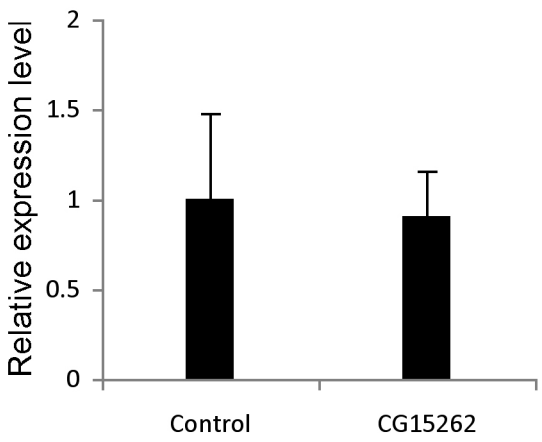

FIGURE S3.—Knockdown of *CG15262* has no effect on expression level of the transgene from the GAL4/UAS system. We confirmed that knock-down of *CG15262* had no effect on expression level of the transgene. A DsRed control transgene was expressed with *gmr-GAL4*, together with *UAS-dcr2* and *UAS-CG15262-RNAi* (*CG15262*), or with *UAS-dcr2* alone (control), and the level of transcription of the DsRed transgene was assessed by quantitative realtime polymerase chain reaction, relative to rp49 control. No significant difference in expression level was seen.

TABLE S1

Complete overview of final deficiency data

TABLE S2

Results of dsRNA crosses to genes within deficiency regions

Tables S1 and S2 are available for download as Excel files at http://www.genetics.org/cgi/content/full/genetics.110.121418/DC1.