



The *Nopp140* gene in *Drosophila melanogaster* displays length polymorphisms in its large repetitive second exon

Sonu Shrestha Baral¹ · Patrick J. DiMario¹

Received: 19 January 2019 / Accepted: 10 April 2019 / Published online: 20 April 2019
© The Author(s) 2019

Abstract

Nopp140, often called the nucleolar and Cajal body phosphoprotein (NOLC1), is an evolutionarily conserved chaperone for the transcription and processing of rRNA during ribosome subunit assembly. Metazoan Nopp140 contains an amino terminal LisH dimerization domain and a highly conserved carboxyl domain. A large central domain consists of alternating basic and acidic motifs of low sequence complexity. Orthologous versions of Nopp140 contain variable numbers of repeating basic–acidic units. While vertebrate *Nopp140* genes use multiple exons to encode the central domain, the *Nopp140* gene in *Drosophila* uses exclusively exon 2 to encode the central domain. Here, we define three overlapping repeat sequence patterns (P, P', and P'') within the central domain of *D. melanogaster* Nopp140. These repeat patterns are poorly conserved in other *Drosophila* species. We also describe a length polymorphism in exon 2 that pertains specifically to the P' pattern in *D. melanogaster*. The pattern displays either two or three 96 base pair repeats, respectively, referred to as *Nopp140-Short* and *Nopp140-Long*. Fly lines homozygous for one or the other allele, or heterozygous for both alleles, show no discernible phenotypes. PCR characterization of the long and short alleles shows a poorly defined, artifactual bias toward amplifying the long allele over the short allele. The significance of this polymorphism will be in discerning the largely unknown properties of Nopp140's large central domain in rDNA transcription and ribosome biogenesis.

Keywords Nopp140 · Repeat polymorphism · *Drosophila melanogaster*

Introduction

The nucleolar phosphoprotein of 140 kDa (Nopp140) is a ribosome assembly factor that locates within the dense fibrillar component of nucleoli and within nuclear Cajal bodies (CBs) (Meier and Blobel 1990, 1992). It was first described as a chaperone that shuttles between the nucleolus and cytoplasm (Meier and Blobel 1992), perhaps to facilitate the import of other ribosome assembly factors that have SV40 T

antigen-type nuclear localization signals (Meier and Blobel 1990). Subsequent work showed that Nopp140 associates with C/D box small nucleolar ribonucleoprotein (snoRNPs) that guide nucleotide-specific 2'-O-methylation of the pre-rRNA, and with H/ACA box snoRNPs that direct site-specific pseudouridylation of pre-rRNA and snRNAs (Yang et al. 2000). These interactions with Nopp140 may initiate within CBs as Nopp140 is thought to shuttle these snoRNPs from CBs to nucleoli (Isaac et al. 1998; Yang et al. 2000; Wang et al. 2002; Meier 2005; Lo et al. 2006; He and DiMario 2011). Nopp140 also interacts with RNA Pol I to regulate rRNA transcription, perhaps linking RNA Pol I transcription with pre-rRNA processing (Chen et al. 1999; Meier 2005).

While metazoan Nopp140 orthologues vary in size, they are conserved in overall domain organization (Meier and Blobel 1992; Cairns and McStay 1995; Meier 1996; Kelly et al. 2006; Lee et al. 2014). The amino terminus consists of a LisH dimerization domain (Mateja et al. 2006), followed by a large central domain with variable numbers of alternating basic and acidic peptide motifs. With one basic motif and one adjacent acidic motif forming a unit, human

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00438-019-01568-6>) contains supplementary material, which is available to authorized users.

✉ Patrick J. DiMario
pdimari@lsu.edu

Sonu Shrestha Baral
sshre22@lsu.edu

¹ Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

Nopp140 has 11 repeat units, *Xenopus laevis* Nopp180 has 17 repeat units (Cairns and McStay 1995), and *C. elegans* Nopp140 (*dao-5*) has 27 repeat units (Lee et al. 2014). The basic motifs are rich in lysine, alanine, and proline while the acidic motifs are rich in aspartate, glutamate, and serine. These serine residues are phosphorylated in vivo by casein kinase type II (CKII) enzymes thus contributing to the acidic properties of the motif (Meier 1996; Li et al. 1997). The precise function of this central domain remains largely unknown. The carboxyl domain is the most conserved region among the Nopp140 orthologues. While its function also remains poorly understood, a conserved but putative protein kinase A phosphorylation site within the carboxyl domain indicates that the carboxyl domain is involved in regulatory signaling events (Meier 1996).

Human Nopp140, also called nucleolar and Cajal body phosphoprotein (NOLC1; CCDS 65925.1), is encoded by *NOLC1* (ID: 9221) which maps to chromosome 10 (NC_000010.11). As an example of vertebrate *Nopp140* gene organization, *NOLC1* contains 13 exons that encode isoforms of 700 or 707 amino acids. Exon 1 encodes the LisH domain, and exon 2 encodes the remaining 18 residues of the amino terminal domain. Exons 3–11 encode the large central domain that contains the 11 basic–acidic repeat units; some of the individual basic or acidic motifs span exon junctions. Exons 12 and 13 in *NOLC1* encode the conserved carboxyl tail of 83 amino acids. No mutations are known to exist for human *NOLC1*, suggesting de novo mutations in *NOLC1* are embryonic lethal. On the other hand, *dao-5* in *C. elegans* is encoded by 5 exons. Exon 3 and the relatively large exon 4 encode the majority of the 27 basic–acidic repeat units, while the 3' end of exon 4 and exon 5 encode the well-conserved carboxyl domain. At least one mutation in *dao-5* has been described (Lee et al. 2014).

Treacle is another nucleolar ribosome biogenesis factor related to Nopp140 in structure and function, but thus far found only in vertebrates. Human treacle is encoded by the *TCOF1* gene (gene ID: 6949; chromosome 5, NC_000005.10); haplo-insufficiency mutations in *TCOF1* result in the Treacher Collins syndrome (TCS), a ribosomopathy in which select neural crest cells undergo p53-dependent apoptosis resulting in impaired craniofacial development (Jones et al. 2008; Sakai and Trainor 2009; Narla and Ebert 2011). Similar to *NOLC1*, the *TCOF1* gene consists of multiple exons. Exon 1 again encodes a LisH domain, and similar to mammalian Nopp140, treacle contains several basic–acidic repeats comprising a large central domain (Wise et al. 1997). As with human *NOLC1*, multiple exons in the *TCOF1* gene encode these repeats (Isaac et al. 2000).

Nopp140 is the closest gene in *Drosophila* spp. to *TCOF1* in humans. In *Drosophila melanogaster*, *Nopp140* maps to the left arm of chromosome 3 proximal to the centromere in cytological region 78F4. Two protein isoforms

Fig. 1 The second exon of *Nopp140* gene contains multiple overlapping repeat patterns unique to the *Drosophila melanogaster* species. **a** Schematic representation of a region within the second exon (1650 bp in length) of *Nopp140* gene that contains three repeat patterns: repeat pattern P with five 48-base bp repeats, repeat pattern P' with three 96 bp repeats, and region P'' with three 96 bp repeats different from that of pattern P'. *Nopp140* genomic regions for two isoforms of Nopp140 protein, Nopp140-True and Nopp140-RGG, are shown (solid blocks are exons and lines are introns). **b** Complete protein sequence of *Drosophila* Nopp140-True and Nopp140-RGG are shown. The beginning and ending sequence of the common second exon are highlighted in blue. Within the second exon, alternating basic (containing lysine and arginine) and acidic (containing aspartate and glutamate) motifs are highlighted in green and yellow respectively. The N-terminus (first exon which is common to both protein isoforms) and C-terminus (which is unique to each protein isoform) are highlighted in grey. The entire repeat pattern P' region comprising three 96-bp repeats is marked by brackets. **c** Sequence alignment of *Drosophila melanogaster* *Nopp140* gene and Nopp140-like protein coding genes from seven other *Drosophila* species. Shown here is the alignment of *Nopp140* second exon containing the three repeat patterns P, P', and P''. The first nucleotide of each repeat region is indicated by an arrow, the nucleotide number is in parenthesis (numbering begins at the first nucleotide of the second exon), and the arrowhead indicates the end of all repeats

arise by alternative splicing. Nopp140-True contains 686 amino acids, while Nopp140-RGG contains 720 residues (Waggener and DiMario 2002). Both isoforms localize to nucleoli and CBs when expressed exogenously as GFP fusions in transgenic embryos, larvae, and adults (McCain et al. 2006). The mRNA encoding Nopp140-True consists of four exons: exon 1 encodes the LisH domain as in human NOLC1 and treacle. While the large central domains in human NOLC1 and treacle are encoded by several exons, the large central domain in *Drosophila* Nopp140-True consists of 16 repeating basic–acidic units, all encoded by exon 2. The carboxyl terminus of Nopp140-True is encoded by exons 3 and 4; this domain is 65% identical to the carboxyl terminus of human NOLC1 over a 94-amino acid stretch. Therefore, we consider Nopp140-True to be the canonical orthologue of vertebrate Nopp140 (Waggener and DiMario 2002). The second isoform in *Drosophila*, Nopp140-RGG, is encoded by the same two exons 1 and 2, and is thus identical to Nopp140-True in its first 583 amino acid residues. Alternative splicing, however, generates a distinctly different carboxyl domain rich in glycine and arginine residues that form repeating RGG tri-peptide motifs (see Fig. 1b). Similar Nopp140-like proteins with RGG carboxyl domains have been reported in *Trypanosoma brucei*, *Trypanosoma cruzi*, and in *Leishmania major* (Kelly et al. 2006). RGG motifs are common to many RNA-associated proteins (Raman et al. 2001).

The large central domains of Nopp140 and treacle display low amino acid complexity, and are intrinsically disordered, a common feature for RNA chaperones (Dyson and Wright 2002; Tompa and Csermely 2004). Here we describe several

repeating but overlapping amino acid sequence patterns within the large central domain of the Nopp140 isoforms in *Drosophila melanogaster*. We compare differences in these patterns within other *Drosophila* species, and we describe two *Nopp140* alleles in *Drosophila melanogaster*, *Nopp140-Long* and *Nopp140-Short*. The two alleles differ by exactly 96 bps that encode a 31-amino acid repeat unit within the central domain. Both alleles seem to be functional as there are no apparent adverse phenotypes in flies expressing one or the other allele. The significance of this polymorphism lies in discerning the properties and functional interactions of the large central domain of metazoan Nopp140 orthologues in rDNA transcription and pre-rRNA processing.

Materials and methods

Fly lines

Flies were obtained from the Bloomington *Drosophila* Stock Center (Indiana University) and maintained in the laboratory at room temperature on standard cornmeal medium. The following fly stocks were used: *w¹¹¹⁸* (Bloomington stock 3605), *Daughterless-GALA (Da-GALA)* (Bloomington stock 8641), *Oregon-R-C* (Bloomington stock 5), *Canton-S* (Bloomington stock 64349), *TM3/Er⁵⁰* (Bloomington stock 64349), *TM3/Sb* (Bloomington stock 4534), *CyO/Sp¹* (Bloomington stock 4199), and the *Nopp140* gene deletion line, *KO121* (He et al. 2015).

Genomic DNA extraction

Healthy well-fed flies were frozen at $-80\text{ }^{\circ}\text{C}$ for 10 min and subsequently homogenized in 100 mM Tris-HCl (pH 7.5), 100 mM EDTA, 100 mM NaCl, and 0.5% SDS. Following 30 min incubation at $70\text{ }^{\circ}\text{C}$, genomic DNA was precipitated in 1:2 ratio of 5 M KOAc: 6 M LiCl on ice for

10 min. The precipitated genomic DNA was purified using phenol–chloroform extraction, followed by ethanol precipitation. The DNA pellet was suspended in deionized water, and the DNA concentration was measured by a NanoDrop 1000 Spectrophotometer.

PCR amplification and band analysis

PCR reactions were performed in 25- μL total volume containing 50–100 ng (unless otherwise specified) of genomic DNA, 0.40 μM of each primer (Table 1), 0.160 mM of each dNTP, 0.30 mM of MgCl_2 , 0.5 \times Phusion GC Buffer, and 0.40 unit of Phusion high-fidelity DNA polymerase (M0530S, New England BioLabs). Amplification was performed in a BIO-RAD C1000 Thermal Cycler with the following thermal cycling conditions: 2 min initial denaturation at $95\text{ }^{\circ}\text{C}$, followed by 34 cycles of denaturation for 30 s at $95\text{ }^{\circ}\text{C}$, annealing for 30 s at varying temperatures as per the primer pairs, and elongation at $72\text{ }^{\circ}\text{C}$ for varying lengths of time depending upon amplicon sizes, followed by 5 min at $72\text{ }^{\circ}\text{C}$. Phusion polymerase was used for all genomic PCRs. For the PCRs in Fig. 3b, the 25- μL reaction volume contained 1.0 unit of *Taq* DNA Polymerase (MO267S, New England BioLabs), 1 \times ThermoPol Buffer, 0.40 μM of each primer, 0.160 mM of each dNTP, and 0.5 μL of gel extracted DNA fragments as templates. The PCR products were resolved on 1% agarose gels and imaged by ChemiDoc XRS+ system (Bio-Rad Laboratories). Image LabTM Software 6.0.1 (BioRad Laboratories) was used to quantify band intensities followed by Student's *t* test analysis on Microsoft Excel to obtain *p* values.

DNA sequencing and sequence analysis

Following phenol–chloroform gel extraction and ethanol precipitation, the PCR products were sequenced with an

Table 1 Primer sequences and PCR conditions

Primer pairs	Annealing temperatures ($^{\circ}\text{C}$)	Elongation times	Sequences
I	61.0	1 min	Forward: 5' ATGACAGACCTGCTAAAGATAGCC Reverse: 5' AAACCTTAGCAGGAGCCAACCTTG
II	61.0	1 min	Forward: 5' GTCGCAAGTTGGCTCCTGCTAAGA Reverse: 5' AACATGCATAAGCGGACTGCGG
III	61.0	1 min	Forward: 5' CATCTAAAATTATCCGCAGTCCGC Reverse: 5' CTAGTCAAATTTATTGAATTGACTCCTACG
IV	57.0	1 min 30 s	Forward: 5' CGGCAAGGCTGTCAAAAAGGCA Reverse: 5' TTCTCATTGCCATTGGTAGC
V	68.0	25 s	Forward: 5' GTCGCAAGTTGGCTCCTGCTAAGA (same as II-forward primer) Reverse: 5' GGTGGAATTCGTCTTCGCTTGAAGACTTGGCCT
Intronic	57.0	1 min 45 s	Forward: 5' ATCTGCGTCCTCCTGATC Reverse: 5' AACATGCATAAGCGGACTGCGG (same as II-reverse primer)

ABI 3130XL Genetic Analyzer (Applied Biosystems) using BigDye Terminator Cycle Sequencing kit v.3.1. The forward and reverse primers used for PCR amplification (Table 1) were used as primers to sequence the PCR products in both forward and reverse directions. Sequences were analyzed and aligned using CLC Sequence Viewer (QIAGEN Bioinformatics).

Data availability

All fly stocks obtained through the Bloomington *Drosophila* Resource Center can be purchased from the center. The *Nopp140* gene deletion line, *KO121*, is available upon request. Sequences of all primers used in this study are provided in Table 1. Supplemental figures that include raw sequence reads are available in figshare.

Results

Repeat sequences within the *Nopp140* second exon

The large central domain of both *Nopp140* isoforms in *Drosophila melanogaster* consists of several alternating basic and acidic motifs. This domain is encoded by the second exon of the *Nopp140* gene (Flybase CG7421; FBgn0037137). We found a large segment within this second exon that can be displayed as three distinct but overlapping repeat patterns; we designate the patterns as P'', P', and P (Fig. 1a). The repeat patterns begin, respectively, at base pairs 874, 900, and 938 with these numbering designations starting at the first base pair of the second exon (Fig. 1a). Repeat pattern P' consists of three tandem repeats of 96 bp each, and repeat pattern P'' consists of three 96-bp tandem repeats and a fourth incomplete repeat of 26 bp. On the other hand, repeat pattern P has five tandem repeats of 48 bp followed by a sixth incomplete repeat of 10 bp.

These repeat patterns contribute to the alternating basic (green) and acidic (yellow) motif pattern within the central domain of both *Nopp140* isoforms in *Drosophila melanogaster* (Fig. 1b). For instance, the 96 bp repeat sequence in the P' pattern encodes basic residues (lysine) in its first half and acidic residues (glutamate and aspartate) along with serine residues in the second half: APAKATPAKAIP-AKKAASSDDSSSEEEAPKK. These serine residues are likely phosphorylated by casein kinase type II enzymes (Meier 1996) to contribute to the acidic properties of the region.

When comparing the second exon of *Nopp140* from *Drosophila melanogaster* with *Nopp140* genes from other *Drosophila* species (*D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. virilis*, and *D. grimshawi*), we found that only *D. melanogaster* carries the repeating

patterns P'', P' and P. The other species carry roughly one-third of the entire repeat segment (Fig. 1c). For instance, while *D. melanogaster* and *D. simulans* are closely related species, *D. simulans* lacks 195 bps encoding 65 amino acids, essentially the latter two P' repeats as compared to *D. melanogaster*. This suggests that a single P' repeat is functionally sufficient for *Nopp140* proteins in the other *Drosophila* species, and that the extra repeats may have duplicated independently in *D. melanogaster*.

Nopp140-Long and *Nopp140*-Short alleles in *Drosophila melanogaster*

Here we describe a polymorphism in the *Nopp140* gene of *Drosophila melanogaster* that involves the number of P' pattern repeats described in the previous section. The allele with two repeats is referred to as *Nopp140*-Short, while the allele with three repeats is referred to as *Nopp140*-Long. Both alleles appear to be functional: fly lines homozygous for one or the other allele show no discernible abnormality. We investigated five fly lines (*w¹¹¹⁸*, *Oregon-R-C*, *Canton-S*, *Daughterless-GAL4*, and *TM3/Er⁵⁰*). For each fly line, we used three PCR primer pairs I, II, and III (Table 1) to amplify the *Nopp140* gene region from genomic DNA extracted from forty adult flies. Together, the three primer pairs covered the entire *Nopp140* gene (Fig. 2a). While primer pairs I and III each produced a single PCR product of expected size (1046 bp and 1280 bp, respectively) in all five fly lines, primer pair II generated either one or two PCR products (881 bp and/or 977 bp) depending on the fly line, more specifically, the long or short alleles (Fig. 2a). The *w¹¹¹⁸* fly line produced only the short PCR product of 881 bp, indicating this stock is homozygous for *Nopp140*-Short. *Oregon-R-C*, *Canton-S*, and *Daughterless-GAL4* (*Da-GAL4*) fly lines produced the 977-bp PCR product. These lines are therefore homozygous for *Nopp140*-Long. Primer pair II amplified both PCR products of 881 bp and 977 bp from genomic DNA isolated from the *TM3/Er⁵⁰* fly stock. Thus, this stock is heterozygous for the *Nopp140*-Long and *Nopp140*-Short alleles (Fig. 2a). Subsequent work with a *Nopp140* gene deletion (He et al. 2015) balanced with *TM3* showed that this *TM3* balancer chromosome contains the *Nopp140*-Short allele (Supplemental Figure S1).

Sequencing the various PCR products obtained from the five fly lines (Fig. 2a) revealed that the *Nopp140*-Short allele consists of two P' repeats (1 and 2), whereas the *Nopp140*-Long allele contains three P' repeats (1, 2, and 3), all of which are exactly 96 bp in length. Each repeat encodes a 31-amino acid peptide sequence, with functional codons starting with the second nt within the repeat thus leaving 2 nts at the end of the repeat (see Fig. 2b).

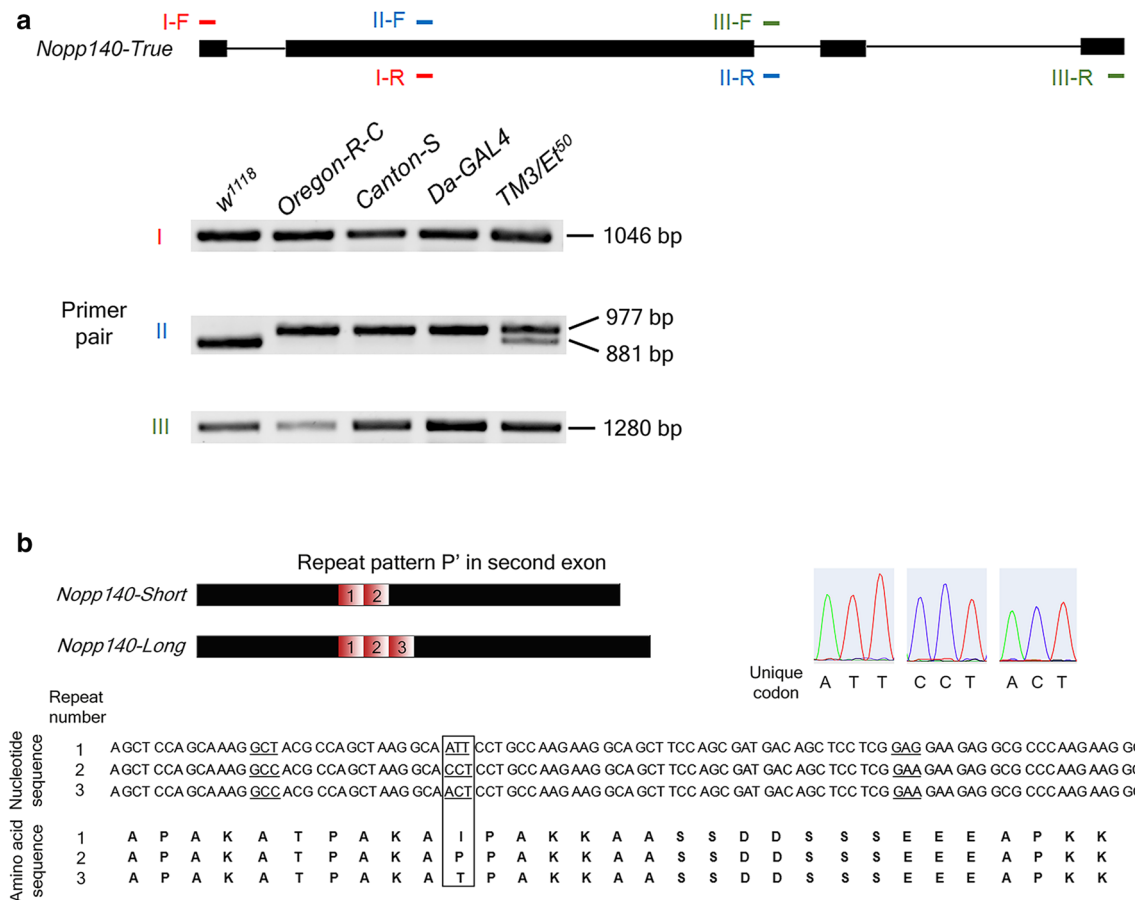


Fig. 2 Two *Nopp140* alleles, *Nopp140-Short* and *Nopp140-Long*, found in *Drosophila melanogaster*. **a** PCR amplification of *Nopp140* genomic DNA region (from start to stop codons for *Nopp140-True* isoform) in five *Drosophila* fly lines ($n=40$; 20 adult male flies and 20 adult female flies): *w¹¹¹⁸*, *Oregon-R-C*, *Canton-S*, *Daughterless-GAL4* (*Da-GAL4*), and *TM3/Er⁵⁰*, using three primer pairs, I, II and III (F forward, R reverse). **b** *Nopp140-Short* and *Nopp140-Long*

alleles have two and three repeats respectively, corresponding to repeat pattern P', with the repeats numbered 1, 2 and 3 in the 5'–3' direction on the coding strand. The nucleotide sequences of the three repeats of pattern P' and the corresponding amino acid sequences for each repeat are shown. The codons at which the three repeats differ are underlined, and within the box are the codons that encode for variable amino acids that are unique to each repeat

The nucleotide sequence of the three repeats is identical except at three codons (codons 5, 11, and 25; underlined in Fig. 2b). Codon 5 encodes alanine in all three repeats using either GCT or GCC. Codon 25 encodes glutamate in all three repeats using either GAG or GAA. Interestingly, codon 11 encodes amino acids that are unique to each repeat; specifically, repeat 1 uses ATT encoding isoleucine, repeat 2 uses CCT encoding proline, and repeat 3 uses ACT encoding threonine (boxed enclosure in Fig. 2b; the sequencing chromatograms for these unique codons are provided). Thus, we can distinguish among the three repeats of the P' pattern using these particular codons. While sequencing the PCR primer pair II products, we simultaneously sequenced the PCR products obtained with primer pairs I and III; the sequences for these PCR

products were identical among all five fly lines examined (Supplemental Figure S2).

Preferential amplification of *Nopp140-Long* allele in genomic PCRs

In our PCR analyses of fly lines heterozygous for the *Nopp140-Long* and *Nopp140-Short* alleles, such as in *TM3/Er⁵⁰* or *TM3/Sb*, we consistently observed the long PCR product in greater abundance compared to the short product. We used primer pairs II and IV to amplify short and long products from the *TM3/Er⁵⁰* fly line (see Figs. 2a, 3a, respectively), and primer pair IV to amplify short and long products from the *TM3/Sb* fly line (Fig. 3a). We had similar observations using genomic DNA extracted from the F1 progeny (*Da-GAL4*+, $n=40$) using primer pair IV

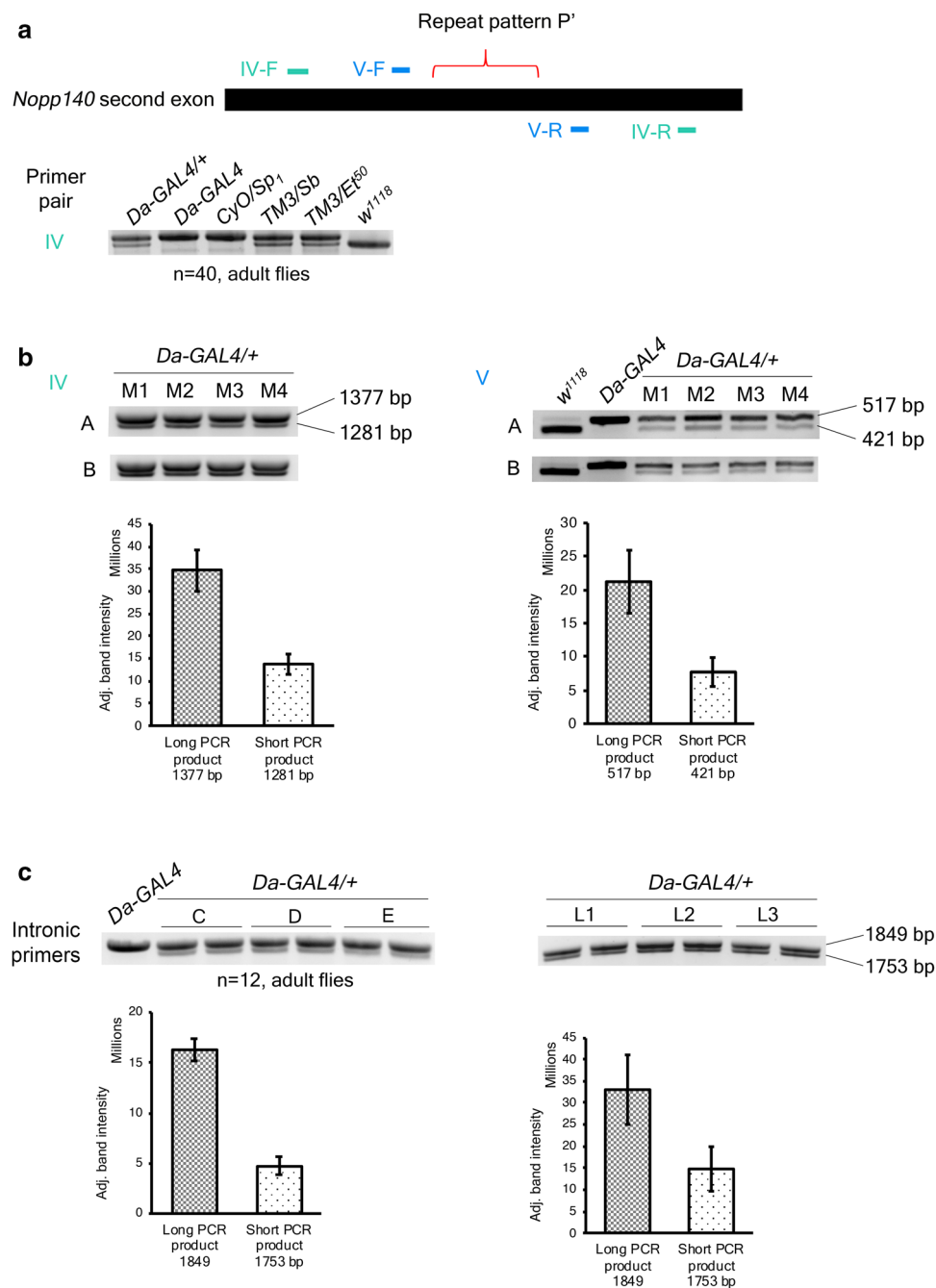


Fig. 3 *Nopp140*-Long allele is preferentially amplified in genomic PCRs. **a** PCR-amplified *Nopp140* second exon containing repeat pattern P' in different *Drosophila melanogaster* fly lines ($n=40$; 20 adult male flies and 20 adult female flies) using two different primer pairs, IV and V (F forward, R reverse). Primer annealing sites on the *Nopp140* second exon are indicated in the schematic diagram. **b** Primer pairs IV and V were used to amplify the repeat pattern P' from F1 heterozygotes (*Da-GAL4/+*) obtained from reciprocal crosses, (indicated as gels A and B) between *Da-GAL4* and *w¹¹¹⁸* flies. Two PCR products of 96 bp difference were amplified indicating presence of both long and short *Nopp140* alleles. M1–M4 are PCR samples in which genomic DNA extracted from single adult male flies was used as template. Quantifications of the band intensities of the long and short PCR products for the single male genomic PCRs are included. An unpaired one-tailed

t test with unequal variance was performed; *p* values were $2.1\text{E}-07$ and $4.6\text{E}-06$ for PCRs with primer pairs IV and V, respectively). **c** PCR-amplified *Nopp140* second exon containing repeat pattern P' in F1 heterozygotes *Da-GAL4/+* ($n=12$ adult male flies) obtained from three separate crosses (indicated as gel lanes C, D, and E) between *Da-GAL4* and *w¹¹¹⁸* flies, and from individual third instar larvae (indicated as gel lanes L1–L3) obtained from reciprocal cross B, using an intronic primer pair that spans the entire *Nopp140* second exon. Quantification of the band intensities of the long and short PCR products from the adult male genomic PCR and the single third instar larva genomic PCR are included. An unpaired one-tailed *t* test with unequal variance was performed; *p* values were $1.5\text{E}-09$ and $2.4\text{E}-4$ for PCRs using genomic DNA of twelve adult flies and single third instar larvae respectively

(Fig. 3a). These F1 progeny were obtained from a cross between the *Da-GAL4* stock (homozygous for *Nopp140-Long*) and the *w¹¹¹⁸* stock (homozygous for *Nopp140-Short*). The *Da-GAL4* transgene resides on the third chromosome along with *Nopp140-Long*. Since the *Da-GAL4* transgene is marked with *mini-white⁺*, we can track this particular chromosome in subsequent crosses. We obtained the same unequal product abundance between long and short products when we performed single fly PCR analyses. We used primer pairs IV and V to amplify genomic DNA extracted from single *Da-Gal4/+* F1 adult male flies (Fig. 3b). The long PCR product was always in greater abundance compared to the short product.

Since we were dealing with repeat sequences, we considered the possibility that the overabundance of the long product was a PCR artifact. Therefore, we amplified the repeat region with primer pairs that were set further away from the repeats and within the two introns flanking the second exon; increasing distance of the primer annealing sites relative to the repetitive sequence is known to improve amplification of major PCR products and significantly reduce undesired PCR artifacts (Hommelsheim et al. 2014). With these intronic primers, we expected equal amplification of the long and short PCR products from heterozygous flies. However, the long PCR product was still more abundant than the short PCR product when using genomic DNA extracted from both male adult flies ($n = 12$) from three separate crosses, C, D, and E, and from a set of three single third instar larva, L1–L3 (Fig. 3c).

Based on these results, we considered the possibility of an in vivo gene conversion event occurring in the heterozygous flies whereby an insertion of a 96 bp repeat into the *Nopp140-Short* alleles would increase the abundance of the *Nopp140-Long* alleles. This could explain the overabundance of the long PCR product in the heterozygous genomic PCRs. In order to test for this possibility, we amplified the second exon of *Nopp140* from a mixture of genomic DNAs containing equal amounts (50 ng each) from *w¹¹¹⁸* flies (homozygous for *Nopp140-Short* allele) and *Da-GAL4* flies (homozygous for *Nopp140-Long* allele) using the same intronic primers. This in vitro mixing should closely mimic the genomic DNA sample from heterozygous flies, and eliminate the possibility of any gene conversion events. Once again however, the long PCR product was more abundant than the short PCR product (Fig. 4a) indicating that this is simply a PCR artifact, but one that is consistent and predictable.

In summary, we report a PCR phenomenon in which the longer genomic fragment with repeats 1, 2 and 3 was preferentially amplified compared to the shorter genomic fragment with two repeats 1 and 2, despite equal amounts of the *Nopp140-Short* and the *Nopp140-Long* alleles present as templates. We never observed the opposite result

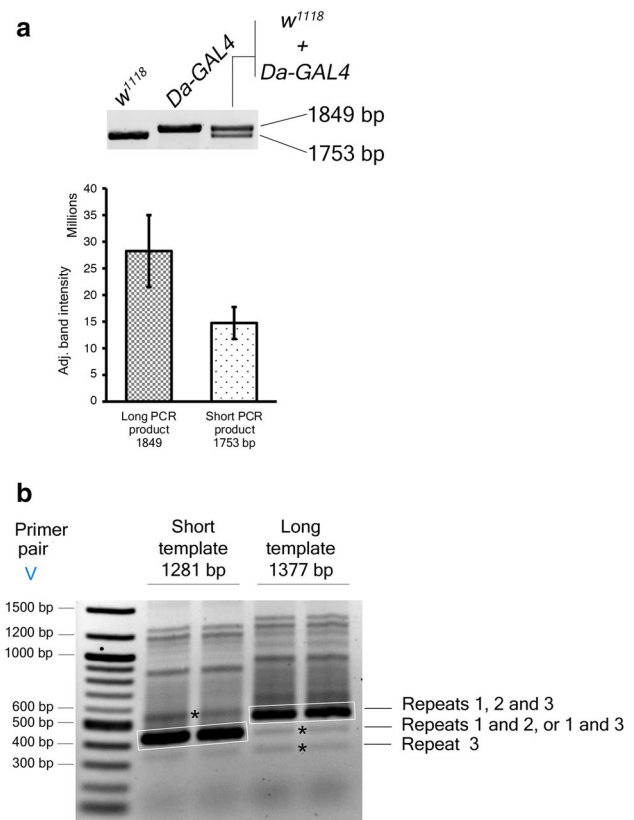


Fig. 4 PCR of repetitive DNA sequences generates artifacts. **a** Genomic PCRs using an intronic primer pair and one of the three templates: 50 ng of *w¹¹¹⁸* genomic DNA (contains only the *Nopp140-Short* allele), 50 ng of *Da-GAL4* genomic DNA (contains only the *Nopp140-Long* allele), or a mixture of 50 ng each of *w¹¹¹⁸* and *Da-GAL4* genomic DNA (contains both *Nopp140-Short* and *Nopp140-Long* alleles in equal quantities). Two PCR products of 96 bp difference, 1849 bp and 1753 bp, were amplified among which the longer one was over amplified in PCR with the mixed genomic DNA sample (*Da-GAL4 + w¹¹¹⁸*). An unpaired one-tailed *t* test with unequal variance was performed for the mixed genomic PCR sample ($n = 5$ reactions); a *p* value of 0.0040 was obtained. **b** PCR using primer pair V amplified the repeat pattern P' from either a 1281-bp short template DNA containing two 96-bp repeats or a 1371-bp-long template DNA containing three 96-bp repeats. The bands enclosed in white boxes are the major PCR products of expected sizes, and the undesired fragments that were also sequenced and analyzed in this study are indicated by asterisks

where the short product was more abundant than the long product.

PCR amplification of undesired fragments with repetitive DNA sequences

Studies have shown that amplification of DNA with repetitive sequences can produce an array of undesired fragments with variable number of repeats (Hommelsheim et al. 2014). Indeed, upon amplification of the repeat pattern P' with primer pair V, PCR products

of variable sizes were amplified producing a laddering effect (Fig. 4b). There were undesired PCR products both longer and shorter than the major PCR product (boxed in Fig. 4b). Surprisingly, when using a short DNA fragment of 1281 bp (the short template) that contained two P' repeats (1 and 2) as template, sequencing analysis confirmed that one of the higher molecular weight PCR fragments (marked by an asterisk in Fig. 4b) contained three P' repeats: 1, 2, and 3 (in this order 5'–3', each distinguishable by the unique codons described earlier; similar to results shown in Supplemental Figure S4). Likewise, when using the long template (1377 bp) that had three P' repeats (1, 2, and 3), one of the lower molecular weight PCR products contained two P' repeats 1 and 2, another had two P' repeats 1 and 3, and yet another had either repeat 1 or 3 (marked by asterisks in Fig. 4b; see Supplemental Figure S3). The results again suggest that the undesired PCR products with variable P' pattern repeats arise as amplification artifacts.

Discussion

Vertebrate *NOLCI* and *TCOF1* genes use multiple exons to encode alternating basic and acidic motifs that repeat to comprise large central domains in their protein products. The *Drosophila Nopp140* gene, however, uses a single large exon to encode a similar large central domain with alternating basic and acidic repeating motifs. We described three repeat peptide sequence patterns within this central domain. The repeats within a single pattern occur in tandem, but the three different patterns overlap with each other (Fig. 1a). One of the patterns, P', consists of 96 bp tandem repeats in *Drosophila melanogaster*, but other closely related *Drosophila* species did not carry the sequence as tandem repeats, suggesting that the repeat region in the *Drosophila melanogaster Nopp140* gene may have originated relatively recently in evolutionary time.

Here we identified length polymorphisms affecting the number of P' repeats. Various fly lines carry either two or three 96 bp P' tandem repeats. Hence, we describe the two *Nopp140* alleles as either *Nopp140-Short* (two repeats) or *Nopp140-Long* (three repeats). The individual repeats can be identified by single nucleotide polymorphisms that exist in codon 11 (see Fig. 2b); the *Short* allele contains repeats 1 (ATT) and 2 (CCT), and the *Long* allele contains repeats 1 (ATT), 2 (CCT), and 3 (ACT). We described *D. melanogaster* fly lines that are homozygous for either the long allele or the short allele. Balanced fly lines can be heterozygous for both alleles.

Possible PCR artifacts

We consistently observed the second exon of *Nopp140-Long* allele was over produced by PCR compared to that of the *Nopp140-Short* allele. The two alleles should be present in equal numbers in the genomic DNA extracted from heterozygous individuals, and hence PCR amplification should theoretically yield equal amounts of product from the two alleles. Initially we considered an in vivo gene conversion event in which the *Nopp140-Short* allele is preferentially converted to the *Nopp140-Long* allele as a possible explanation because we found that *w¹¹¹⁸* genomic PCRs amplified a minor PCR product of the size expected for a long PCR product on certain occasions (Supplemental Figure S4). While we did not carry unbalanced heterozygotes (*Da-GAL4/w¹¹¹⁸*) past two generations to test the stability of the long versus short alleles, we did test different larval tissues to see if amplification of the long allele was restricted to polyploid (midgut) versus diploid (brain) tissues, but we saw no difference; the long allele was always preferentially amplified regardless of the source of genomic DNA.

In a simple mixing experiment, we combined genomic DNA homozygous for *Nopp140-Long* with equal amounts of genomic DNA homozygous for *Nopp140-Short*, and observed the same preferential amplification of the long allele. This in vitro mixing experiment refutes an in vivo gene conversion event. Confusing the issue however, *w¹¹¹⁸* genomic DNA is homozygous for *Nopp140-Short* allele containing only tandem repeats 1 and 2, but we identified minor amplicons from *w¹¹¹⁸* genomic DNA with repeat pattern P' containing tandem repeats 1, 2 and 3 (Supplemental Figure S4). Likewise, we consistently amplified (and sequenced) a 521 bp PCR product with repeat pattern P' that contained tandem repeats 1, 2 and 3 (similar to the results shown in Supplemental Figure S4), although the 1281 bp template with repeat pattern P' contained only tandem repeats 1 and 2.

We have yet to explain this preferential PCR amplification of the long product versus the short product, but we may have an explanation for the amplification of short PCR products with repeats 1 and 2, 1 and 3, or a single repeat 1 (Fig. 3b) from the 1371 bp template DNA that contained all three tandem repeats. Studies by others have shown that PCR amplification of repetitive DNA regions can generate undesired fragments of variable sizes containing haphazard combinations of repeats, similar to our observation shown in Fig. 3b. PCR products with variable repeats could potentially arise from an initial production of incomplete, single-stranded fragments that then act as mega-primers in subsequent PCR cycles. These mega-primers would then misalign with the existing templates at the repetitive sites, thereby resulting in an array of PCR fragments of sizes other than the major product (Hommelsheim et al. 2014).

Implications for long versus short *Nopp140* alleles

Earlier, we showed that partial depletion of Nopp140 in *Drosophila* by siRNA expression resulted in *Minute*-like phenotypes such as deformed wings, legs, and tergites (Cui and DiMario 2007). These phenotypes were similar to those observed with the various haplo-insufficiency *Minute* mutations in genes encoding ribosomal proteins, and reminiscent of the phenotypes associated with the TCS (Sæbøe-Larssen et al. 1998; Cui and DiMario 2007). Flies homozygous for either the short allele or the long allele, or flies heterozygous for the two alleles are viable and do not exhibit any of the discernible phenotypes associated with the partial depletion of the Nopp140 isoforms.

While the P' repeat polymorphism does not seem to have a significant impact on the core functions of Nopp140, the protein's central domain is generally considered to be an unstructured region. Recent studies have revealed the important role of inherently disordered proteins (IDPs) in promoting phase separation to form subcellular membrane-less compartments or assemblages (Toretsky and Wright 2014), especially within nucleoli and CBs (Feric et al. 2016). Nopp140 resides within the dense fibrillar components of nucleoli and in CBs when present (Isaac et al. 1998); both nucleoli and CBs are membrane-less compartments within nuclei. The cDNA clones we have used in the past express the long versions of the two *Drosophila* Nopp140 isoforms; we have not worked with cDNAs encoding the short versions. Future work should determine how the large central domain of Nopp140 with variable numbers of basic–acidic repeat units may influence molecular interactions during rDNA transcription, pre-rRNA processing, snoRNP dynamics, and perhaps nucleolar and CB phase-separations and assemblages.

Funding This study was funded by the National Science Foundation (Grant no. MCB1712975).

Compliance with ethical standards

Conflict of interest Sonu Shrestha Baral declares that she has no conflict of interest. Patrick J. DiMario declares that he has no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Cairns C, McStay B (1995) Identification and cDNA cloning of a *Xenopus* nucleolar phosphoprotein, xNopp180, that is the homologue of the rat nucleolar protein Nopp140. *J Cell Sci* 108:3339–3347
- Chen H-K, Pai C-Y, Huang J-Y, Yeh N-H (1999) Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol Cell Biol* 19:8536–8546. <https://doi.org/10.1128/MCB.19.12.8536>
- Cui Z, DiMario PJ (2007) RNAi knockdown of Nopp140 induces *Minute*-like phenotypes in *Drosophila*. *Mol Biol Cell* 18:2179–2191. <https://doi.org/10.1091/mbc.e07-01-0074>
- Dyson HJ, Wright PR (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:54–60. [https://doi.org/10.1016/S0959-440X\(02\)00289-0](https://doi.org/10.1016/S0959-440X(02)00289-0)
- Feric M, Vaidya N, Harmon TS et al (2016) Coexisting liquid phases underlie nucleolar sub-compartments. *Cell* 165(7):1686–1697. <https://doi.org/10.1016/j.cell.2016.04.047>
- He F, DiMario PJ (2011). Structure and function of Nopp140 and treacle. In: Olson MOJ (ed) *The nucleolus, protein reviews*, 15th edn. Springer, Berlin, pp 253–278 https://doi.org/10.1007/978-1-4614-0514-6_11
- He F, James A, Raje H et al (2015) Deletion of *Drosophila* Nopp140 induces subcellular ribosomopathies. *Chromosoma* 124:191–208. <https://doi.org/10.1007/s00412-014-0490-9>
- Hommelsheim CM, Frantzeskakis L, Huang M, Ülker B (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci Rep* 4:5052. <https://doi.org/10.1038/srep05052>
- Isaac C, Yang Y, Meier UT (1998) Nopp140 functions as a molecular link between the nucleolus and the coiled bodies. *J Cell Biol* 142:319–329. <https://doi.org/10.1083/jcb.142.2.319>
- Isaac C, Marsh KL, Paznekas WA, Dixon J, Dixon MJ, Jabs EW, Meier UT (2000) Characterization of the nucleolar gene product, treacle, in Treacher Collins syndrome. *Mol Biol Cell* 11:3061–3071. <https://doi.org/10.1091/mbc.11.9.3061>
- Jones NC, Lynn ML, Gaudenz K, Sakai D, Aoto K, Rey J-P et al (2008) Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nat Med* 14:125–133. <https://doi.org/10.1038/nm1725>
- Kelly S, Singleton W, Wickstead B, Ersfeld K, Gull K (2006) Characterization and differential nuclear localization of Nopp140 and a novel Nopp140-like protein in Trypanosomes. *Eukaryot Cell* 5:876–879. <https://doi.org/10.1128/ec.5.5.876-879.2006>
- Lee C-C, Tsai Y-T, Kao C-W, Lee L-W, Lai H-J, Ma T-H et al (2014) Mutation of a Nopp140 gene *dao-5* alters rDNA transcription and increases germ cell apoptosis in *C. elegans*. *Cell Death Dis* 5:e1158. <https://doi.org/10.1038/cddis.2014.114>
- Li D, Meier UT, Dobrowolska G, Krebs EG (1997) Specific interaction between casein kinase 2 and the nucleolar protein Nopp140. *J Biol Chem* 272:3773–3779. <https://doi.org/10.1074/jbc.272.6.3773>
- Lo SJ, Lee C-C, Lai H-J (2006) The nucleolus: reviewing oldies to have new understandings. *Cell Res* 16:530–538. <https://doi.org/10.1038/sj.cr.7310070>
- Mateja A, Cierpicki T, Paduch M, Derewenda ZS, Otlewski J (2006) The dimerization mechanism of Lis1 and its implication for protein containing the LisH motif. *J Mol Biol* 357:621–631. <https://doi.org/10.1016/j.jmb.2006.01.002>
- McCain J, Danzy L, Hamdi A, Dellafosse O, DiMario P (2006) Tracking nucleolar dynamics with GFP-Nopp140 during *Drosophila* oogenesis and embryogenesis. *Cell Tissue Res* 323:105–115. <https://doi.org/10.1007/s00441-005-0044-9>

- Meier UT (1996) Comparison of the rat nucleolar protein nopp140 with its yeast homolog SRP40. Differential phosphorylation in vertebrates and yeast. *J Biol Chem* 271:19376–19384
- Meier UT (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma* 114:1–14. <https://doi.org/10.1007/s00412-005-0333-9>
- Meier UT, Blobel G (1990) A nuclear localization signal binding protein in the nucleus. *J Cell Biol* 111:2235–2245. <https://doi.org/10.1083/jcb.111.6.2235>
- Meier UT, Blobel G (1992) Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* 70:127–138. [https://doi.org/10.1016/0092-8674\(92\)90539-O](https://doi.org/10.1016/0092-8674(92)90539-O)
- Narla A, Ebert BL (2011) Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115:3196–3205. <https://doi.org/10.1182/blood-2009-10-178129>
- Raman B, Guarnaccia C, Nadassy K, Zakhariev S, Pintar A, Zanuttin F et al (2001) N^{ω} -arginine dimethylation modulates the interaction between a Gly/Arg peptide from human nucleolin and nucleic acids. *Nucleic Acids Res* 29:3377–3384. <https://doi.org/10.1093/nar/29.16.3377>
- Sæbøe-Larssen S, Lyamouri M, Merriam J, Oksvold MP, Lamberts-son A (1998) Ribosomal protein insufficiency and the *Minute* syndrome in *Drosophila*: a dose–response relationship. *Genetics* 148:1215–1224
- Sakai D, Trainor PA (2009) Treacher Collins syndrome: unmasking the role of *Tcofl1*/treacle. *Int J Biochem Cell Bio* 41:1229–1232. <https://doi.org/10.1016/j.biocel.2008.10.026>
- Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperone. *FASEB J* 18(11):1169–1175. <https://doi.org/10.1096/fj.04-1584rev>
- Toretsky JA, Wright PE (2014) Assemblages: functional units formed by cellular phase separation. *J Cell Biol* 206(5):579–588. <https://doi.org/10.1083/jcb.201404124>
- Waggener JM, DiMario PJ (2002) Two splice variants of Nopp140 in *Drosophila melanogaster*. *Mol Biol Cell* 13:362–381. <https://doi.org/10.1091/mbc.01-04-0162>
- Wang C, Query CC, Meier UT (2002) Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140. *Mol Cell Biol* 22:8457–8466. <https://doi.org/10.1128/MCB.22.24.8457-8466.2002>
- Wise CA, Chiang LC, Paznekas WA, Sharma M, Musy MM, Ashley JA et al (1997) *Tcof1* gene encodes a putative nucleolar phosphoprotein that exhibits mutations in Treacher Collins Syndrome throughout its coding region. *Proc Natl Acad Sci USA* 94:3110–3115. <https://doi.org/10.1073/pnas.94.7.3110>
- Yang Y, Isaac C, Wang C, Dragon F, Pogacic V, Meier UT (2000) Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol Biol Cell* 11:567–577. <https://doi.org/10.1091/mbc.11.2.567>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.