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

Sonu Shrestha Baral and Patrick J. DiMario✉

Authors' Affiliation: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803 USA.

Email addresses: Sonu Shrestha Baral – sshre22@lsu.edu, Patrick DiMario – pdimari@lsu.edu

✉ To whom correspondence should be addressed (pdimari@lsu.edu)

Key words: Nopp140, exon, repeat polymorphism, *Drosophila melanogaster*

Conflict of interest:

The authors report no conflict of interest.

Supplementary Information

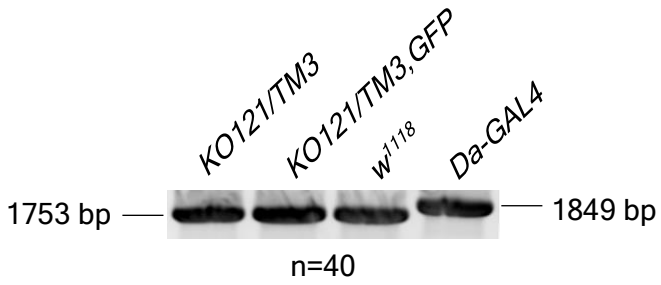


Fig. S1 The *TM3* balancer chromosomes (*TM3 and TM3,GFP)* present in the *Nopp140* gene deletion line, *KO121* (He et al., 2015), contains the *Nopp140-Short* allele similar to the w^{1118} fly line. Intronic primers flanking the entire second exon of *Nopp140* amplified short PCR product (1753 bp) for both *KO121* and w^{1118} genomic DNA, and long PCR product (1849 bp) for *Da-GAL4* genomic DNA (*Da-GAL4* fly line carries the *Nopp140-Long* allele). 'n' refers to the number of flies used to extract genomic DNA.

Primer pair I

Da-GAL4

5’ATGACAGACCTGCTAAAGATAGCCGATGCAATCGTCCTGGAGTACTTGCAGTCCAAGGACAA
GAACCTAGCCAAGGTTTTCCAGCAGAAGACGAAGGCGGTAAGTTTCAGTGCAGAGCGGCAA
GTGCAATATGTGCGGTTTGGATGCACCCGAAACATGTGGTATCTGCGTCCTCCTGATCTAGA
ATCCGAGGGCCTTGCTGTCCGCCGATGGCTCATCGAGCTGTCCACCTGCCACTTGCTGCAA
GTGCCACGTGCTTCGTCCAGATAACCAAGAATTTCTGCCTTTTTCTAACTTGCAGGCCAGTG
TCGCCAAAAGCAGCCCAAAGTTGAGTGAAATCCTTCAGTTCTACCAGACCAAAAGCCCCAAA
AAGATCCCAGCAATCAAGGCGACAGCCGGAGACTCCAGCGAAGACAGCGACTCAGACTCTG
AGTCTGATGCGGCGCCTAAGAAGCCAGCGGCAGCTCCTGCACTAACAAACGGCAAGGCTGT
CAAAAAGGCAGCTTCTTCAACCAGCGAGGACAGCGATTCAGAGGAGGAGAAGAAGCCAGCA
GCAAAGGCCACGCCGGCCAAGGCTGTCGGCAAGAAGGCAAAATCCTCCAGCGAGGACAGT
TCCTCGGAGGAAGAAGCACCCAAAAAAGCTGCTCCAGTGAAGGCACCTCCGGCCAAGGCG
GCTCCCGCGAAGAAGGTTGAGTCCAGCAGCGAGGATAGCAGTTCCGAGGAAGAACCGGCA
AAGCCCGCCGTCAAGGCCACAACGACTAAGGTTGCTCCCGCCAAGAAGGCCGACTCCAGC
AGCGAAGAAAGCAGCTCCGATGAGGAGACTAAGCCAGCTGCCAAGCCTGTAGCCAAGGCT
GCACCGGCAAAGAAAGCAGCATCCTCCAGCGAGGAAAGCGACTCCGATGACGAGCCAGCC
GCCAAGAAGCCTGCCGTCCAGCCTGCCGCAAAGCCAGCTCCTAAGGCAGCCGCTTCTAGTT
CCGAAGACAGCAGCTCAGAGGAGGAGGTCAAACCCGCTGCTAAATCTGCTGCCAAGTTGGC
TCCTGCTAAGA

Primer pair II

w^{1118}

5’AGGGGGCTTCCTCTAGTGATGATAGTTCCTCTGAGACGAGGCACCCAAAAAGGCAGCAACT
CTCGCAAAGCCCATTTCTAAGGCTGCTCCCACCAAAAAGGCCGACAGTTCCACTGAGGATA
GTTCTTCGGAAGACGATGCTCCTAAGAAGGTAGCTCCAGCAAAGGCTACGCCAGCTAAGGC
AATTCCTGCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAGGAAGAGGCGCCCAA
GAAGGCAGCTCCAGCAAAGGCCACGCCAGCTAAGGCACCTCCTGCCAAGAAGGCAGCTTC
CAGCGATGACAGCTCCTCGGAAGAAGAGGCGCCCAAGAAGGCTGCCGCCCCAGCAAAGGC
GACACCGGCCAAAAAGGCCAAGTCTTCAAGCGAAGACAGCGACTCCTATGAGGAGGAAGCT
CCCAAGAAACCAGCCGCGAAGGCCGTTGCAAAGGCTGCTTCTTCGGAAGATAGCGATAGCT
CAGAAGACGAAAAGCCAGCAAAGGCTGCTCCCAAGGCTCTGGCCAAGTCTGCAAAGGCTGC
CTCCTCAGACAGTGACGATTCCAGCGATGAAGAAACGCCGGCAGTGAAGCCAGCTGTCAAG
AAGACTGCTGCTCCTGCGAAGAAGGCTGACAGCAGCAGCGACGAGAGTGACTCTGGTGAA
GAGTCCGGCGAGGTCAAGCCCAACTCAGCTACCAATGGCAATGAGAAGACCGCTCAGAAGC
GCAAGTTTAGTGGTGGCGACCAGGACGAGGCCACTCCCAACAAGAAGTACAACAACTTCTG
TCAAATCGGGAGAGCAACAGGTAAGGCGGTGTAACTTATCTATGTACATCTAAAATTATCCG
CAGTCCGCTTATGCATGTT

Primer pair II

Da-GAL4

5’AGGGGGCTTCCTCTAGTGATGATAGTTCCTCTGAGACGAGGCACCCAAAAAGGCAGCAACT
CTCGCAAAGCCCATTTCTAAGGCTGCTCCCACCAAAAAGGCCGACAGTTCCACTGAGGATA
GTTCTTCGGAAGACGATGCTCCTAAGAAGGTAGCTCCAGCAAAGGCTACGCCAGCTAAGGC
AATTCCTGCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAGGAAGAGGCGCCCAA
GAAGGCAGCTCCAGCAAAGGCCACGCCAGCTAAGGCACCTCCTGCCAAGAAGGCAGCTTC
CAGCGATGACAGCTCCTCGGAAGAAGAGGCGCCCAAGAAGGCAGCTCCAGCAAAGGCCAC
GCCAGCTAAGGCAACTCCTGCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAAGAA
GAGGCGCCCAAGAAGGCTGCCGCCCCAGCAAAGGCGACACCGGCCAAAAAGGCCAAGTCT
TCAAGCGAAGACAGCGACTCCGATGAGGAGGAAGCTCCCAAGAAACCAGCCGCGAAGGCC
GTTGCAAAGGCTGCTTCTTCGGAAGATAGCGATAGCTCAGAAGACGAAAAGCCAGCAAAGG
CTGCTCCCAAGGCTCTGGCCAAGTCTGCAAAGGCTGCCTCCTCAGACAGTGACGATTCCAG
CGATGAAGAAACGCCGGCAGTGAAGCCAGCTGTCAAGAAGACTGCTGCTCCTGCGAAGAAG
GCTGACAGCAGCAGCGACGAGAGTGACTCTGGTGAAGAGTCCGGCGAGGTCAAGCCCAAC
TCAGCTACCAATGGCAATGAGAAGACCGCTCAGAAGCGCAAGTTTAGTGGTGGCGACCAGG
ACGAGGCCACTCCCAACAAGAAGTACAACAACTTCGTCAAATCGGGAGAGCAACAGGTAAG
GCGGCGGTGTAACTTATCTATGTACATCTAAAATTATCCGCAGTCCGCTTATGCATGTT

Primer pair III

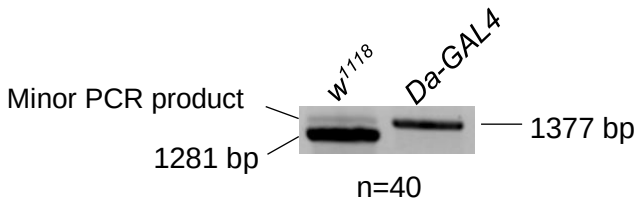
Da-GAL4

5’CATCTAAAATTATCCGCAGTCCGCTTATGCATGTTTTCCTTCAGACATCAATTTTGAATGTCGGATAAA
AATAATTCCTGAGTTAAAGGATTCCCTTAAAAGTAGCTTTTTAAAGCCCTATTGAAATTTGTAG
CCGAACTTTAAACAGCACATAATTTTGGCTCCTTGTCCGATTCCTGTGCAGACAGTTTTACAT
AACTTCATTGCCATTCGCAGAAGAATGACTTCACCTCCACACCTAACAACACCTTCAGCCGA
AACCATAATATGAACAATAGCGGAGGGGGAAGTGGGCGACGGTCGCCCTTCCGAAGGGTA
CGCACCGAGGACGTGGTGGTGGACTCCCGAGTCCAGGACATGTCCTTCGAGGCGAAGGTA
AGTGCCGCAATGGATTTGAGTCGCGTGGGTGTGTGTGTTTGCGACGGCTGTTGAGGATTAG
GCGAGGGGGATCTTTCTACGAGGCTGAAAGGGGCAGACACACAAATTAGCAAGTTGCGCCT
CGAAGAATCTCAAACCACTTGTGGCACTTTCCCAATACATTTTGAACTTATATATCAAATCAAA
TCGTTGCTAATTTGTGGTTTGCCCCTTTCGGCGTTTCTACCATTTCAGGAAAACGACTTTAAG
AAGCACAACAACGGACGGGGAGGCCGAGGAGGCTCAGGCTTCTCCGGACGACCGGATCGC
AGCACTTGGGAGACCAACAAGTTCAACGGCGAGGGCGGCGGTGATGGAGGCGGCTTCAAG
AAGATTGGCGATCGCAAGAGCTTTGGAGGCTTTGATAACAACCAGCGCGGAGGACGTGGTG
GTGGCGGTCGTGGAGGCGGCGGAGGCTTCGGTGGTCGTGGCGGTGGCGGCCGTGGCGG
CGGTGGTGGTTTCGGAGGACGTGGTGGCGGCGGACGCGGAGGCGGCGGCTTTGGAGGAC
GCGGGGGCCGCGGTGGAGGCGGACGTGGAGGCGGATTCGGCAACAAATCCTTCGACTCG
TCGGCGCCAAAGCAAAACAAGAAGATCACCTTCGACAATTAGAAATAGTTAGTTAGCGCCAC
CCACCCGCAAGCACACACACACACACTAGATGCAGTTATATTATCGGATTAACCCAATTTTAT
TATTTGCTCACCACGATAGAAAAACGCTGCCGGATCCTGGGGCGAGCGAGCCAACAAGGAT
CTGAAGCACACGCGCGGCAAGTCCTTCAAACACGAGAAGACCAAGAAGAAGCGCGGCAGC
TACCGAGGTGGCCAAATTGACGTAGGAGTCAATTCAATAAAATTT

Fig. S2 Representative sequences for PCR products amplified with primer pair I from *Da-GAL4* (identical to all other fly lines tested: w^{1118}, *Oregon-R-C, Canton-S,* $TM3/ET^{50}$), primer pair II from w^{1118} (contains repeat pattern P’ with repeats 1 and 2) and *Da-GAL4* (contains repeat pattern P’ with repeats 1, 2 and 3), and primer pair III from *Da-GAL4* (similar to all other fly lines tested; mentioned above). The *Nopp140* genomic sequences around the primer annealing regions that were not included in the sequencing output are also provided (shaded grey). Any primer sequences that are not part of *Nopp140* gene are excluded.

Fig. S3 Sequence chromatograms of minor PCR products amplified from a 1371 bp long template DNA containing three 96 bp repeats 1, 2, and 3 of repeat pattern P' (from Fig 4B). **a** Approximately 400 bp minor PCR product (marked by an asterisk in Fig 4B) contained two 96 bp repeats, either repeat 1 and 2, or repeat 1 and 3. Sequence chromatogram starting from the first base pair of repeat 1 and ending a few bases downstream of the repeat pattern P' is provided. The unique codons that are used to identify the individual repeats are underlined. Most of the sequencing reads indicated presence of repeat 1 (ATT) and 2 (CCT) in the ~400 bp minor PCR product, while others revealed presence of repeat 1 and 3 (ACT) as shown in the boxed inlet by the two overlapping peaks for the first base of the unique codon (green for base 'A' and blue for base 'C'). **b** Approximately 300 bp minor PCR product (marked by an asterisk in Fig 4B) contained a single 96 bp repeat which was either repeat 1 or 3 as indicated by two overlapping peaks for the second base of the unique codon (blue for base 'C' and red for base 'T'). Following the first 96 bp repeat, the downstream sequence reads contained either sequence for the region downstream of the repeat patter P' or a second 96 bp repeat. The latter was detected likely due to gaussian distribution of the ~400 bp minor PCR fragment.

a



b

w^{1118} Short PCR product

5’GGCCTCTTGTGATGATAGTTCCTCTGAGACGAGGCACCCAAAAAGGCAGCAACTCT
CGCAAAGCCCATTTCTAAGGCTGCTCCCACCAAAAAGGCCGACAGTTCCACTGAGG
ATAGTTCTTCGGAAGACGATGCTCCTAAGAAGGTAGCTCCAGCAAAGGCTACGCCA
GCTAAGGCAATTCCTGCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAGGA
AGAGGCGCCCAAGAAGGCAGCTCCAGCAAAGGCCACGCCAGCTAAGGCACCTCCT
GCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAAGAAGAGGCGCCCAAGA
AGGC**[**TGCCGCCCCAGCGAAGGCGACACCGGCCAAAAAGGCCTTGTCTTCACGCGA
AGACGGATTCCACC**]** end of sequence

w^{1118} Long PCR product

5’GGCCTCTTGTGATGATAGTTCCTCTGAGACGAGGCACCCAAAAAGGCAGCAACTCT
CGCAAAGCCCATTTCTAAGGCTGCTCCCACCAAAAAGGCCGACAGTTCCACTGAGG
ATAGTTCTTCGGAAGACGATGCTCCTAAGAAGGTAGCTCCAGCAAAGGCTACGCCA
GCTAAGGCAATTCCTGCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAGGA
AGAGGCGCCCAAGAAGGCAGCTCCAGCAAAGGCCACGCCAGCTAAGGCACCTCCT
GCCAAGAAGGCAGCTTCCAGCGATGACAGCTCCTCGGAAGAAGAGGCGCCCAAGA
AGGC**[**AGCTCCAGCAAAGGCCACGCCAGCTAAGGCAACTCCTGCCAAGAAGGCAGC
TTCCAGCGATGACA**]**GCTCCTCGGAAGAAGAGGCGCCCAAGAAGGCTGCCGCCCCA
GCAAAGGCGACACCGGCCAAAAAGGCCAAGTCTTCAAGCGAAGACGAATTCCACCA

c

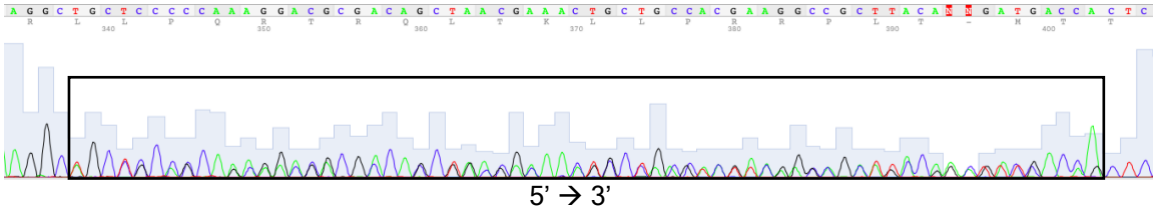


Fig. S4 *w*^{1118} genomic PCRs amplified a minor PCR product that contained repeats 1, 2, and 3 of the repeat pattern P’. **a** Primer pair IV was used to amplify a section of second exon of *Nopp140* containing the repeat pattern P’. Expected sizes were 1377 bp for long PCR product amplified from the *Nopp140-Long* allele present in *Da-GAL4* fly line and 1281 bp for short PCR product amplified from the *Nopp140-Short* allele present in *w*^{1118} fly line. However, *w*^{1118} genomic PCRs produced a minor product of the size expected for a long amplicon produced from *Da-GAL4* genomic PCRs. **b** Sequencing analysis revealed that the minor PCR product in *w*^{1118} genomic PCR contained two distinct DNA molecules. One was short PCR product with repeat pattern P’ that contained tandem repeats 1 (highlighted in green) and 2 (highlighted in yellow) detected due to gaussian distribution of DNA molecules during gel electrophoresis. The other was a long PCR product with repeat pattern P’ that contained tandem repeats 1, 2 and 3 (highlighted in pink) although *w*^{1118} genomic DNA is homozygous for *Nopp140-Short* allele. Codons unique to each repeat are underlined. **c** Sequences marked by bold square brackets in **b** were manually extracted from the boxed region of sequence chromatogram pertaining to *w*^{1118} minor PCR product.