Promoter targeting preferences of the *D. melanogaster* P-element
Raquel S. Linheiro and Casey M. Bergman (Faculty of Life Sciences, University of Manchester, Manchester, UK)

Introduction

The *Drosophila* melanogaster P-element is among the best-studied eukaryotic transposons, with many aspects of its excision and integration process characterized at the molecular level. One of the most striking features of the targeting preferences of the P-element is a strong tendency to insert into the proximal promoter regions of protein coding genes. Despite longstanding speculation about the role of open chromatin during transcription, the genomic determinants of P-element promoter targeting remain a mystery. Here we integrate data from large-scale transposon insertion collections with sequence and chromatin properties of promoter regions to decode the genomic factors associated with P-element promoter targeting in *D. melanogaster*.

Materials and methods

- P-element insertion sites and mRNA start site annotations were obtained from release 5.14 of the *D. melanogaster* genome annotation [1].
- Promoter motif predictions were generated using PATSER with PWMs from the JASPAR database and Ohler et al. [2].
- Genome-wide location data for nucleosomes, Polycomb and Trithorax proteins, RNA polymerase and general transcription factors were obtained from the literature (see Table 1).
- Statistical modelling of the association between genomic factors and the presence or absence of P-elements in a 1 Kb window around the TSS was performed using the R statistical computing language.

Table 1. Type, source and reference for the genome-wide protein binding data used in this study.

<table>
<thead>
<tr>
<th>Data</th>
<th>Cells</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleosomes</td>
<td>Embryos (0 to 12 hours)</td>
<td>[3]</td>
</tr>
<tr>
<td>Polycomb/Trithorax</td>
<td>Embryos (4 to 12 hours)</td>
<td>[4]</td>
</tr>
<tr>
<td>RNA Polymerase</td>
<td>Tollmimic embryos (2 to 4 hours)</td>
<td>[5]</td>
</tr>
<tr>
<td>General transcription factors</td>
<td>S2 cells (20 to 24 hours)</td>
<td>[6]</td>
</tr>
</tbody>
</table>

Results

Analysis of the fine-scale distribution of P-elements around the transcription start site (TSS) of protein-coding genes revealed that (1) P-elements insert randomly with respect to the orientation of transcription, (ii) tend to insert upstream of the TSS and (iii) target a 1 Kb window around the TSS with a peak from -190 to +100 around the TSS (Figure 1). We also find that the insertion sites of P-elements in promoters are affected by avoidance of nucleosomes (Figures 1 & 2). We conclude that nucleosome positioning is not the primary cause of P-element promoter targeting since other DNA transposons also exhibit nucleosome avoidance (Figure 2). Lastly, we find that that distribution of P-element insertion sites in promoter regions is affected by RNA polymerase status of the TSS (Figures 1 & 4).

Given that the polymerase status of a TSS has been shown to correlate with specific core promoter motifs [7], we investigated whether the composition of core promoter motifs differed in promoters with and without P-elements. Analysis of the base and motif composition of promoters with and without P-elements revealed that P-elements can insert into a wide variety of core promoter architectures, but prefer to insert in TATA-less promoters (Figures 3 & 4).

Figure 1. P-element insertions are influenced by nucleosome positioning. (Top) Fine scale pattern of P-element insertion shows promoter targeting over a 1 Kb window with a peak of insertion from -190 to +100 around the TSS. (Middle) When RNA polymerase is present, P-elements tend to insert further downstream, with a greater shift downstream in promoters with actively transcribing polymerase relative to those with a paused polymerase. (Bottom) Shifts in the location of P-element insertion in the -190 to +100 window correlate with nucleosome occupancy, such that when the nucleosome coverage in this region decreases the number of insertions goes up.

Figure 2. Nucleosome avoidance is a general feature of DNA transposons. Shown is the ratio of observed insertions in nucleosomes relative to expected based on nucleosome coverage. Both piggyBac and Minos show even stronger nucleosome avoidance than the P-element both genome-wide and specifically in promoter regions. These results indicate that nucleosome avoidance may be a more general feature of DNA-based transposons.

Figure 3. P-elements prefer TATA-less promoters. (Top) Base composition of TSSs that have a P-element insertion in a ±1 Kb window. (Center) Base composition of TSSs with no P-element in a ±1 Kb window, showing with a significant increase in AT content in the TATA box region. (Bottom) Difference in base composition between promoters with and without P-elements is shown in black, with the null chi-square difference profile based on randomization shown in grey. Significant differences in base composition are observed in the locations of all three major core promoter motifs (TATA, INR and DPE).

Figure 4. Association of P-element insertions with genomic features. Model coefficients of univariate effects in a simplified GLM are shown as histograms and statistical significance is indicated by shading. In addition, a significant negative interaction was observed between the Recruiter PcG proteins and trxG proteins (P-value<0.00001).

Acknowledgements

We thank Matthew Ronshausen and members of the Bergman Lab for discussion. This work was funded by the Foundation for Science and Technology (FCT) Portugal.

References