

1 Introdução

Tal propriedade nos fornece uma interessante ferramenta: as cópias de Pererê-3 observadas no genoma podem ser distinguidas umas das outras, visto que possuem uma região característica a cada uma delas. Desta forma, novas questões podem ser levantadas, e abre-se um novo campo de exploração.

É possível, por exemplo, que se avalie os níveis de transcrição de cada cópia separadamente, além de que, se cada réplica do transposon carrega o mesmo promotor, as diferenças destes níveis devem se originar de fatores externos à sequência do transposon em si, a exemplo da compactação da cromatina ou influência de regiões codificantes próximas. Visto isso, os diferentes níveis de transcrição em conjunto poderiam oferecer um panorama da atividade do genoma ao longo das bases, e ¹.

2 Métodos

Nota-se que a Pode-se, por exemplo, questionar-se acerca do nível de transcrição de cada cópia no genoma, e de sua dependência com a proximidade dos genes vizinhos. Criou-se, então, quatro parâmetros a serem comparados e caracterizados entre as diferentes cópias do transposon Pererê-3, descritos a seguir.

2.1 Contagem de reads

Nas questões relacionadas aos níveis de transcrição, faz-se uso das bibliotecas de RNA-Seq (...) que compreendem diferentes estágios de vida do *S. mansoni* e conta-se, nas regiões de interesse do genoma, quantos reads dessas bibliotecas podem ser alinhados. Divide-se, então, o valor final da contagem pelo comprimento da região em que se procura o alinhamento dos reads, pois assume-se que cada par de base da região tem igual probabilidade de alinhar-se com cada read e, portanto, a contagem final depende linearmente do comprimento do intervalo de DNA que se analisa. A divisão é feita, então, a fim de eliminar esse fator para fins comparativos, e ao valor resultante dá-se o nome de coeficiente de transcrição.²

2.2 Comprimento-mãe

Com curiosidade sobre os efeitos da integridade de cada cópia, chama-se de comprimento-mãe o comprimento do alinhamento BLAST que gerou cada sequência cauda. Toma-se cuidado para considerar apenas os alinhamentos que contém a extremidade 3' da cópia de referência, de forma a ser provável a presença da sequência cauda na região pós-3' no genoma. Assume-se que este valor esteja fortemente correlacionado com a integridade da cópia, e a princípio não são consideradas as mutações pontuais,

que também são componente importante ao se pensar no grau de integridade de uma sequência.

2.3 Correlação com genes vizinhos

Espera-se que grande parcela das cópias encontradas no genoma sejam expressas também por efeito de readthrough, ou transcrição passiva, em que as cópias do transposon são transcritas por se encontrarem em regiões já codificantes, na fase de leitura aberta (ORF) de um gene, por exemplo, e não fazendo uso da maquinaria própria de transcrição. Um efeito semelhante ocorre se o transposon esta inserido a jusante do gene, de forma a poder ser expresso por ocasionais falhas do terminador genico. Supõe-se que esses casos resultem em coeficientes de expressão independentes da integridade da cópia de Pererê-3, uma vez que as proteínas nela codificadas são dispensáveis para a presença da cópia no transcriptoma.

Avalia-se, portanto, a correlação entre os coeficientes de transcrição de cada cópia e o coeficiente de seu gene mais próximo ou do gene em que se insere, para os casos em que a cópia se encontra na região expressa de um gene. A esta correlação será referida daqui em diante como correlação com o gene vizinho, e espera-se que seja maior no casos em que a transcrição da sequência head é passiva.

2.4 Distância ao gene vizinho

Ainda sob a questão da influência dos genes mais próximos no coeficiente de transcrição das sequências head, outro fator a se considerar é a distância em si entre a head e o gene vizinho, em pares de base, pois espera-se que a correlação entre eles se intensifique quanto menor for tal distância. Avaliando os casos em que a head está à jusante ou montante do gene separadamente, espera-se também determinar se há atividade do promotor gênico na transcrição da cópia do transposon, isto é, verificar a ocorrência de transcrição por readthrough ou vazamento. Separando as copias na mesma fita das em fitas diferentes, espera-se também observar efeito da compactação da cromatina na região.

3 Resultados

3.1 As cópias são de fato diferentes?

Visando constatar a unicidade das cópias de transposon pelo genoma, efetuou-se alinhamentos BLAST das sequências head entre elas mesmas, e verificou-se que aproximadamente 87,6% das sequências não foram alinhadas com nenhuma outra e por volta de 96,0% alinharam no máximo uma vez, mostrando que, de fato,

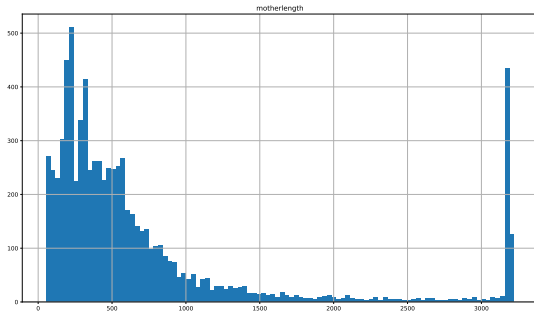
¹Isso é realmente algo para se centrar o projeto?

²Já existe?

não há coincidências gerais claras entre as regiões UTR 3' do Pererê-3, e a temática de análise proposta mantém-se válida.

3.2 A integridade afeta os níveis de transcrição?

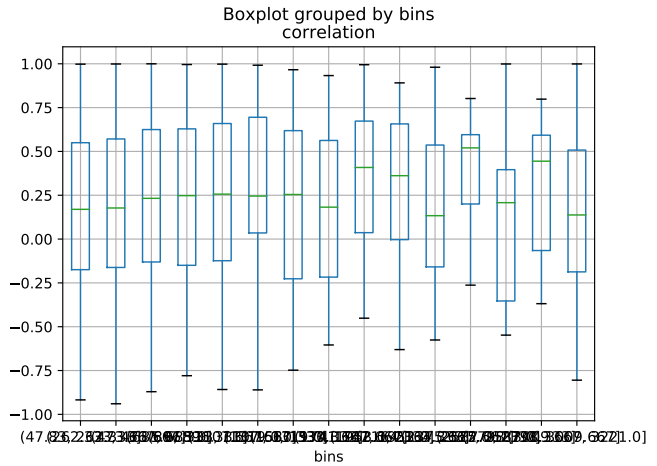
A fim de investigar a influência da completude das cópias em sua capacidade de produzir novos elementos, inicialmente observou-se a distribuição de comprimentos-mãe encontrados.



Já nesse ponto, há uma clara predominância das cópias menores, principalmente com menos de 600 pares de base de comprimento, representadas mais à esquerda na figura 3.2. Contudo, é interessante observar também um aumento significativo da quantidade de cópias íntegras ou quase íntegras, na extrema direita da figura 3.2. Esse aumento sugere fortemente que a integridade das cópias é um fator determinante para sua disseminação (ou ao menos sobrevivência) e corrobora a hipótese de que há grande efeito de readthrough: já previa-se que cópias truncadas, de comprimento menor, teriam menor efeito no produto final de um gene ³, visto que menos aminoácidos seriam adicionados, e por-

tanto seriam mais prováveis de se instalar em um gene mantendo-o ativo e funcional, de forma que a cópia poderia tirar proveito da maior conservatividade das regiões gênicas e predominar no genoma ⁴.

Em seguida, passou-se a analisar mais diretamente a influência do comprimento-mãe na transcrição: elaborou-se gráfico de dispersão entre o coeficiente de transcrição de cada head e seu comprimento mãe, e o resultado é apresentado na figura 3.2.



Embora a quantidade de dados seja muito maior nas extremidades direita e esquerda do gráfico, como já apontava a figura 3.2, não há diferença significativa de nível de transcrição entre as duas regiões, o que pode sugerir que, embora incompletas, as cópias de menor comprimento-mãe ainda sim façam uso da maquinaria própria ⁵.

Se aprofundando neste efeito, procura-se então observar separadamente as cópias sobrepostas ou não a algum gene, bem como à jusante ou montante de algum, a fim de esclarecer efeitos de readthrough ou vazamento.

[placeholder]

³Só quando em fase, não? investigar melhor. E se for em regiões intrônicas?

⁴Elas se instalam degradadas ou se degradam depois de se instalar? Em tese, degradá-las favorece o gene. Escrever sobre hipótese da taxa de degradação diferente ao longo do tempo?

⁵Mesmo as com 60 bp? Parece improvável. Lembre-se de que a parte descartada é justamente o início em 5'. Não deveria estar aí nosso promotor?