Wernig Zorc, Sara

Distinct pattern of NF-kappaB related transcription factor binding sites across long non coding RNAs

Abstract

Background

TRANSCRIPTION FACTORS

Transcription factors (TF) are proteins that bind to specific DNA sequences, called DNA binding motifs, which are typically 6-12bp long. These sequences are variable (degenerate consensus) and thus influence specificity and affinity of binding. TF are transcribed in the nucleus, translated in the cytoplasm and through nuclear localization signal on their own sequence, reenter the nucleus (*Davidson and Peter, 2015*). Upon reentering the nucleus each TF will find thousands of potential binding sites, albeit with different affinity. DNA binding motif alone is thus not sufficient for TFs to exert their function. The key in their combinatorial work with other TFs and/or other molecules, for example DNA polymerase II, histone marks etc.. TF also function cell type specific and context dependent. One such example are TF that help in neural differentiation by stabilizing the the molecules that are part of commitment to differentiate to neurons.

Many studies have been conducted on TF binding on promoter and enhancer regulatory regions, where they can aid in repressing or inducing transcription, normally in collaboration with other factors as mentioned above. Not much is known about TF binding on introns/exons and in particular introns/exons of non coding regions of the genome. There is an estimation that 98% of TF binding motifs are located in the introns or exons of non coding portion of the genome. There is a need for more data in order to understand why this is the case.

Several hypothesis exist regarding TF's function in these regions:

- Binding can have no effect

- Promoters can be downstream of TSSI

- An intron can harbor a regulatory element for another gene

- There can be alternative TSS within the introns

- TFs bound to an intron can regulate elongation or splicing

- TFs may regulate alternative splicing

LONG NON CODING RNAs

LONG NON CODING RNAs AND TRANSCRIPTION FACTORS IN DISEASE

To that extent I have developed a tool, called PREB, which can predict binding sites with high affinity for many lncRNAs and motifs at once. And outputs if there is a preference for certain exons, higher than by chance. Thus inferring TF binding on that region may have an important function.

I took NF-KappaB related group of TFs from JASPAR 2018 database and GenCode long non coding RNA (lncRNA) annotation for Homo sapiens hg38, version 27.

Methods

PREB tool was developed in python3.6, to address this question.

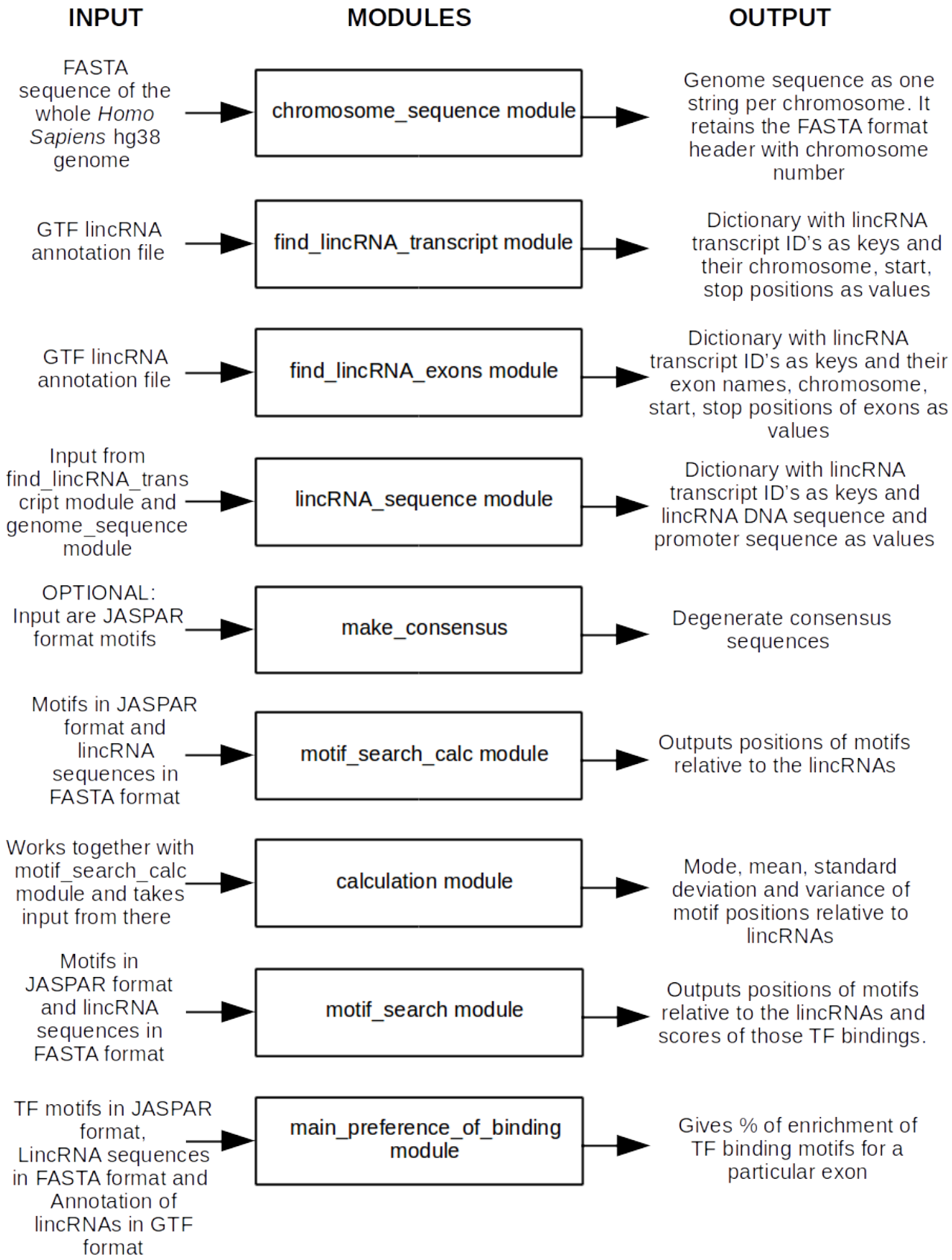

Figure 1: Diagram of PREB tools (PREferential Binding tool).

Results and discussion

References

GHOSH, S., SATI, S., SENGUPTA, S. and SCARIA, V. (2015). Distinct patterns of epigenetic marks and transcription factor binding sites across promoters of sense-intronic long noncoding RNAs. *Journal of Genetics*, 94(1), pp.17-25.

The Genome in Development

Eric H. Davidson, Isabelle S. Peter, in Genomic Control Process, 2015