Dear Sir or Madam:

We are pleased to submit the manuscript *"geneXtendeR: optimized functional annotation of ChIP-seq data"* for consideration as an Applications Note to *Bioinformatics*. This manuscript proposes an R/Bioconductor package for tailoring the annotation of ChIP-seq peaks to their nearest genes. Prior software in this area (ChIPpeakAnno, HOMER, and BEDTools) have focused exclusively on distance-minimizing algorithms between peaks and the TSS regions of their nearest genes. However, geneXtendeR significantly expands this definition to also focus on cis-regulatory and proximal promoter regions of nearest genes, allowing users to perform iterative gene-feature overlaps after adding to the gene-span a user-specified region upstream of the start of the gene model and a fixed (500 bp) region downstream of the gene, resulting in assigning to a gene the features that do not physically overlap with it but are sufficiently close. This allows users to iteratively align peaks to multiple gene transfer format (GTF) files to assess what global gene-spans optimize the genomewide alignment of peaks with their closest genes. This facilitates the process of deciphering which differentially enriched peaks are dysregulating which specific genes, which, in turn, aids experimental follow-up and validation in designing primers for a set of prospective genes during qPCR. Users have over a dozen functions at their disposal, ranging from peak hotspot analysis of statistically significant peaks, to gene ontology analysis, network analysis, and word cloud analysis of gene-GO pairs at multiple genomic intervals, to graphical explorations of peak length distributions, and even functions to transform individual peaks into merged peaks based on distance cutoffs of adjacent peaks. All in all, we tested geneXtendeR on 547 human transcription factor ChIP-seq ENCODE datasets and 214 human histone modification ChIP-seq ENCODE datasets, providing the analysis results as case studies and highlighting the key biological insights acquired in Figure 1.

We have also successfully applied geneXtendeR during the analysis of a histone modification ChIP-seq study investigating the neuroepigenetics of alcohol addiction (PMID: 27573876), where geneXtendeR was used to determine an optimal upstream extension cutoff for H3K9me1 enrichment (a commonly studied broad peak) in rat brain tissue based on geneXtendeR plots of both significant peaks and total peaks. This analysis helped us to identify, functionally annotate, and experimentally validate synaptotagmin 1 (Syt1) as a key mediator in alcohol addiction and dependence. The study has been cited 8 times since its publication last year.

To use geneXtendeR, please download the latest version from Bioconductor's development branch: https://bioconductor.org/packages/devel/geneXtendeR/. A detailed package vignette can be found here: https://bioconductor.org/packages/devel/bioc/vignettes/geneXtendeR/inst/doc/geneXtendeR.pdf.

With best regards,

--

Bohdan B. Khomtchouk, Ph.D.
Postdoctoral Research Scholar, Laboratory of Prof. Or Gozani M.D., Ph.D.
Department of Biology
Stanford University
Stanford, CA 94305