Comment

https://doi.org/10.1038/s41588-022-01194-w

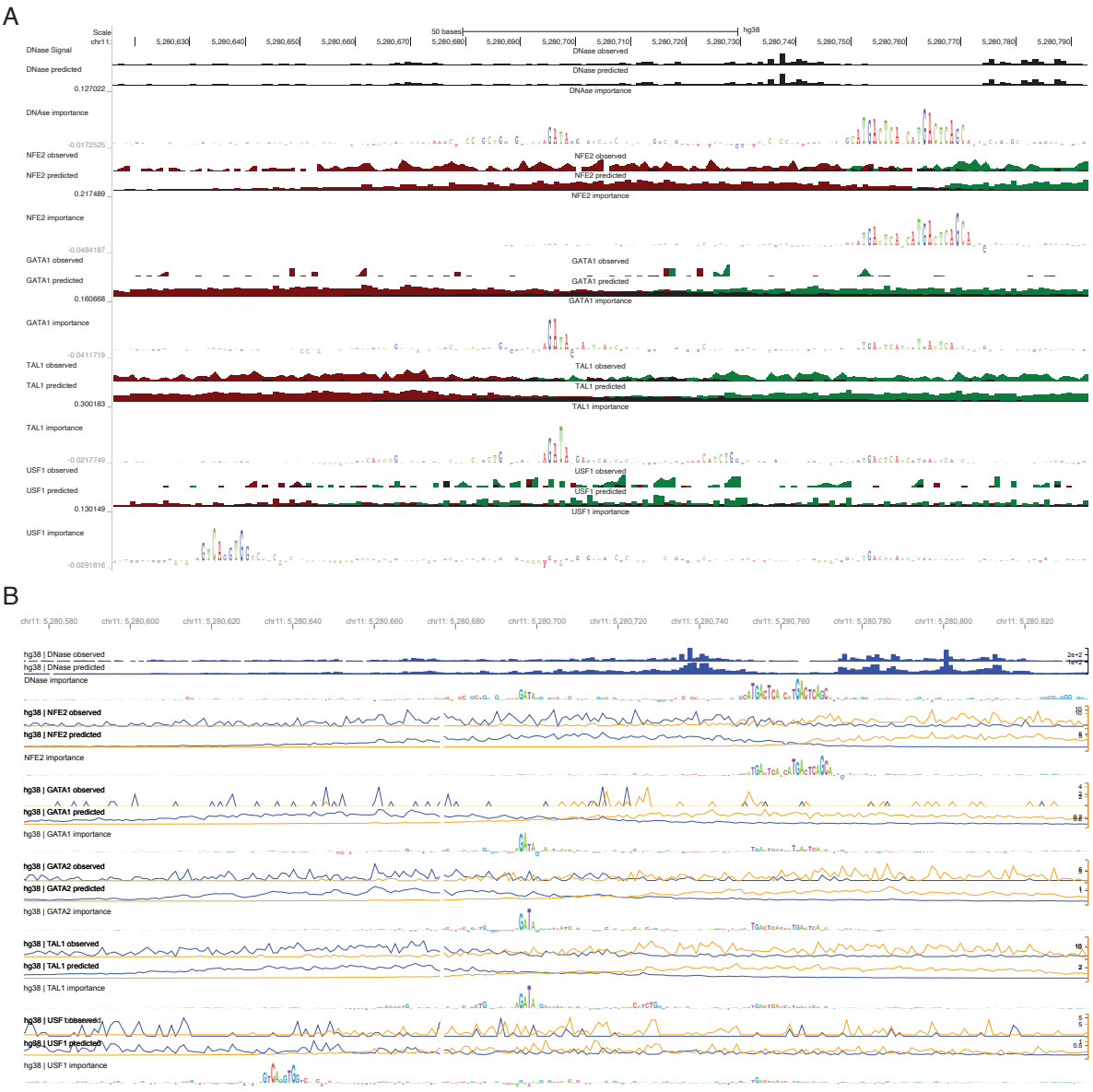
The dynseq browser track shows context-specific features at nucleotide resolution

In the format provided by the authors and unedited

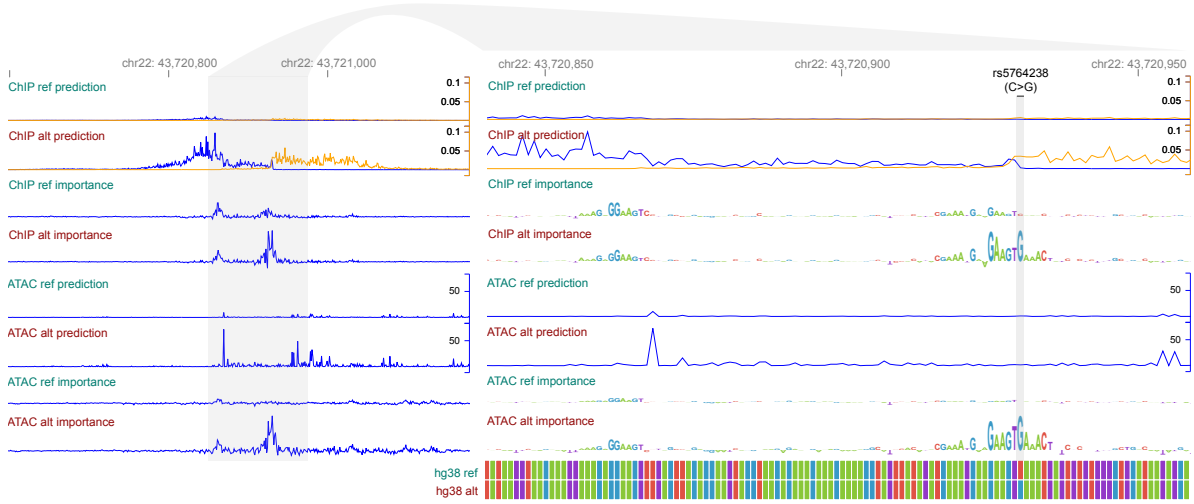
Supplementary Note

A salient application for dynseq tracks is to explore the sequence features impacted by functional genetic variants. As a case study, we trained BPNet models on ChIP-seq profiles of the SPI1 transcription factor and ATAC-seq profiles in the GM12878 cell line[1,2]. We used these models to predict the allelic effect of non-coding variants on ATAC-seq and SPI1 ChIP-seq profiles and derive nucleotide-resolution DeepLIFT importance scores for the sequences containing the reference and alternate alleles of each variant. NC_000022.11:g.43720930C>G (rs5764238:C>G) is a single nucleotide variant that has been previously shown to have a significant allelic effect on SPI1 binding in a binding quantitative trait locus (bQTL) study[3]. We used the Resgen HiGlass browser to visualize the predicted SPI1 ChIP-seq and ATAC-seq profiles as BigWig tracks and the DeepLIFT importance scores as dynseq tracks for the pair of sequences containing the reference and alternate allele of this variant (**Supplementary Figure 2**). The BigWig tracks show that the alternate (G) allele is predicted to enhance SPI1 binding and chromatin accessibility, in agreement with the bQTL study. The dynseq tracks show that the G allele creates a strong SPI1 binding motif with high importance scores from the SPI1 ChIP-seq and ATAC-seq models, thereby revealing this motif as the primary driver of enhanced signal. Thus, the dynseq tracks add an additional level of interpretability to the study of genetic variants by focusing attention on visually recognizable sequence features in a context-specific manner.

1. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
3. Tehranchi, A. K. *et al.* Pooled ChIP-seq links variation in transcription factor binding to complex disease risk. *Cell* **165**, 730–741 (2016).



Supplementary Figure 1. UCSC Genome Browser and Resgen session for cis-regulatory element dissection Same tracks as in **Figure 1** in **(A)** UCSC Genome Browser and **(B)** Resgen around hg38 chr11:5280600-5280800. For ChIP-seq observed and predicted tracks, the red track denotes the plus (+) strand and the green track denotes the minus (-) strand in UCSC **(A)** and blue and orange respectively in Resgen **(B)**.



Supplementary Figure 2. Resgen/HiGlass browser session for model-guided, non-coding variant interpretation Zoomed out (left, hg38 chr22:43720597-43721177) and close-up (right, chr22:43720840-43720958) views of predicted base-resolution coverage profiles (as BigWig tracks) and nucleotide importance scores (as dynseq tracks) from BPNet models of SPI1 ChIP-seq and ATAC-seq data for a genomic sequence containing the reference (C) and alternate (G) allele of the rs5764238 genetic variant, previously identified as a SPI1 binding QTL. For ChIP-seq predicted tracks, blue tracks denote the plus (+) strand coverage and orange tracks denote the minus (-) strand coverage. The G allele is predicted to increase SPI1 ChIP-seq and ATAC-seq signal by creating a strong SPI1 motif, as seen in the dynseq tracks.

	WashU	UCSC	HiGlass/Resgen
Input format	BigWig	BigWig	BigWig
Auto or fixed y-axis	✅	✅	✅
Reverts to BigWig display when zoomed out	✅	✅	✅
Support for files on remote servers	✅	✅	✅
Vector export	✅	✅	✅
Customizable background color	✅	❌	❌
Customizable font and nucleotide colors	❌	❌	✅
Custom fasta file per dynseq track	❌	❌	✅

Supplementary Table 1. Supported functionalities of current implementations of the dynseq track