Annotating Variants and Displaying Multiple Alignments on Your Assemblies Using the UCSC Genome Browser

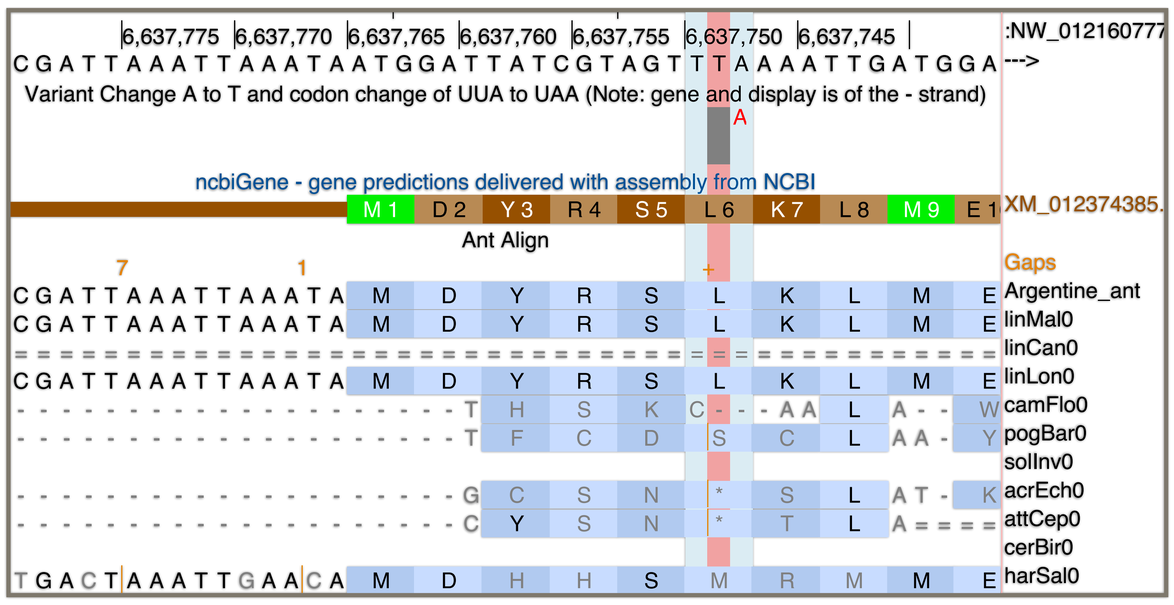
Brian T. Lee[1], Brian Raney[1], Matthew L. Speir[1], Galt P. Barber[1], Christopher M. Lee[1], Kate R. Rosenbloom[1], Angie S. Hinrichs[1], Jonathan Casper[1], Hiram Clawson[1], Ann S. Zweig[1], Max Haussler[1], Robert M. Kuhn[1], W. James Kent[1]

[1] Univ. of California Santa Cruz Genomics Institute Santa Cruz, CA, US

Creating Assembly Hubs

The UCSC Genome Browser (https://genome.ucsc.edu/) is a free, web-based tool that allows researchers to visualize and explore genomic data, including new sequences.

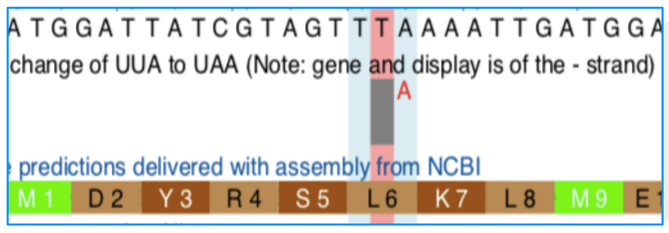
Using Assembly Hubs users can visualize any sequence and annotations upon it. An underlying binary indexed “2bit” file hosted remotely represents the ACTG of the new sequence. These assembly hubs can be further annotated by researchers with most of UCSC’s track formats including Gene Prediction Tracks (https://genome.ucsc.edu/goldenpath/help/bigGenePred.html), which can be used in turn to predict the effects of novel variants. Further additional remotely hosted binary indexed data (**"big"** files such as **bigPsl**, **bigChain**, **bigMaf**, ect.) can annotate alignments or other genomic data.



This image shows an assembly hub of Argentine Ant (GCA_000217595.1, limHum0) where the display is flipped to view the reverse strand that has been annotated with a gene track (showing a gene XM_012374385) above several Multiple Alignment Files for 10 other aligned ant species (using **bigMaf** format). A red highlight emphasizes a variant track (T>A) resulting in a change from coding for a Leucine to a Stop Codon highlighted in blue.

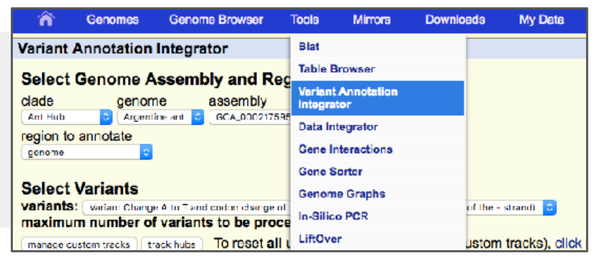
Annotating Variants in Assembly Hubs

The Variant Annotation Integrator (VAI) tool can be used with bigGenePred annotations to predict functional effects of variants on transcripts. For example, a variant might be located in the coding sequence of one transcript, but in the intron of an alternatively spliced transcript and VAI will predict the functional impact for each gene transcript.



1. This **bigGenePred** track defines exons and introns and displays codons with numbering. An additional variant track (either **VCF** or **pgSNP**) defines variants (in the image a variant changes a UUA to a UAA codon).

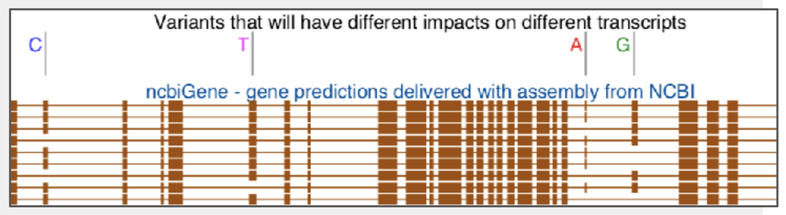
2. With a variant track and a gene track, the VAI tool can **annotate impacts** of up to 10,000 variants. Filters allow **screening out** synonymous, intronic and intergenic variants so only desired impacts are returned.



Feature	Feature type	Consequence	Position in cDNA	Position in CDS	Position in protein	Amino acid change
XM_012374385.1	Transcript	stop_gained	125	17	6	L/*

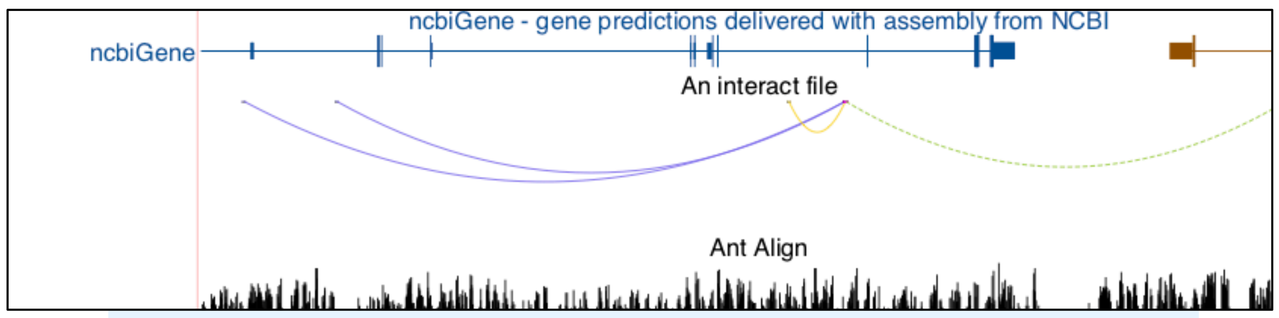
3. The VAI output predicts the variant’s impact. In this case a change from a **Leucine to a Stop Codon.** VAI output is in Ensembl VEP format.

4.VAI predicts results specific to each transcript in the region. In this image some variants fall into introns and so different predicted impacts will result.



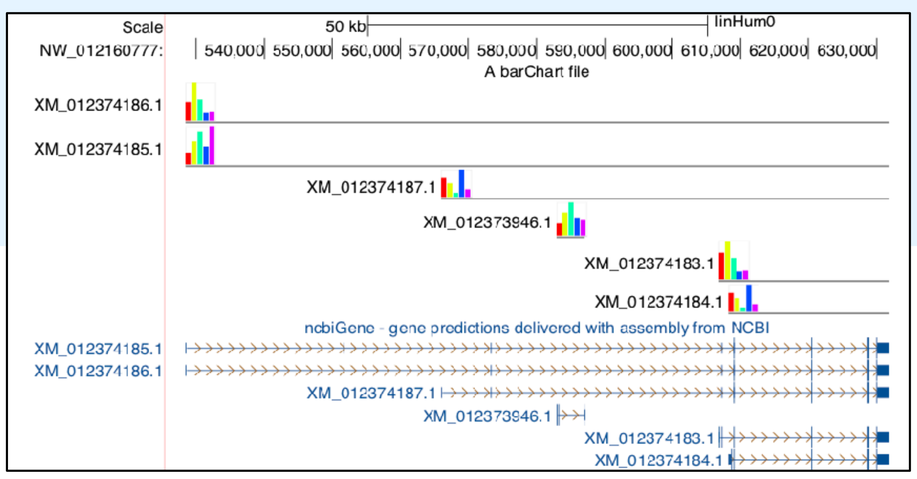
Displaying Multiple Alignment and other Custom Data Types and Views in Assembly Hubs

bigMaf: displays Multiple Alignment Files visualizing other aligned sequences (https://genome.ucsc.edu/goldenpath/help/bigMaf.html).

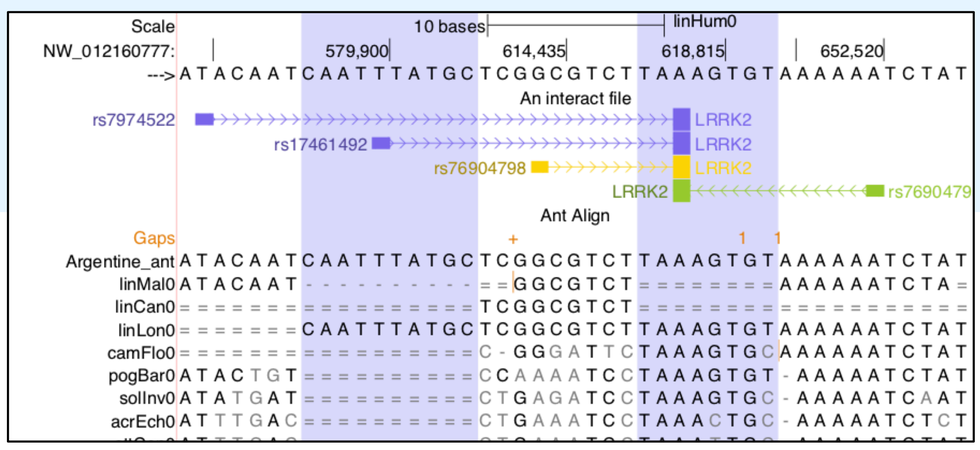


bigInteract: displays pairwise interactions as arcs or half-rectangles connecting two genomic regions on the same chromosome. For directional interactions such as SNP/gene, the interactions in the reverse direction is dashed (https://genome.ucsc.edu/goldenpath/help/intearact.html).

bigBarChart: displays a graph of category-specific values over genomic regions (https://genome.ucsc.edu/goldenpath/help/barChart.html).



Multi-Region View: displays a genome sliced into user-defined regions, even across different chromosomes. End-points of SNP/gene interactions are shown (http://genome.ucsc.edu/goldenPath/help/multiRegionHelp.html).



Future work

- Allow for **PCR** searches on Assembly Hubs. The future addition of PCR gfServers would allow verifying primers on novel genomes.
- Allow bigGenePred in assembly hubs to define the **exon-only** and **gene-only** modes of **multi-region.**

More information

Assembly Hub help:
http://genomewiki.ucsc.edu/index.php/Assembly_Hubs
Contact us: https://genome.ucsc.edu/contacts.html

Download this poster:
http://genomewiki.ucsc.edu/index.php/PENDINGLINK

Follow us:
@GenomeBrowser /ucscGenomeBrowser UCSC Genome Browser

References

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 2018 Nov 8; https://doi.org/10.1093/nar/gky1095

Funding and acknowledgements

Funding provided by National Human Genome Research Institute [U41HG002371].

We would like to thank our loyal users, our collaborators and data contributors, and our tireless systems administrators for their support.

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute