

JBrowse

PAG 2015

Scott Cain
GMOD Coordinator
scott@scottcain.net

What is GMOD?

- A set of interoperable open-source **software** components for visualizing, annotating, and managing biological data.
- An active **community** of developers and users asking diverse questions, and facing common challenges, with their biological data.

Who uses GMOD?



Plus hundreds more organizations.

What are Genome Browsers good for?

- Visualizing dense data from a multitude of sources (genes from a GFF file, RNASeq data from a BAM file, variation data from a VCF file)
- Homology and gene expression support for gene models

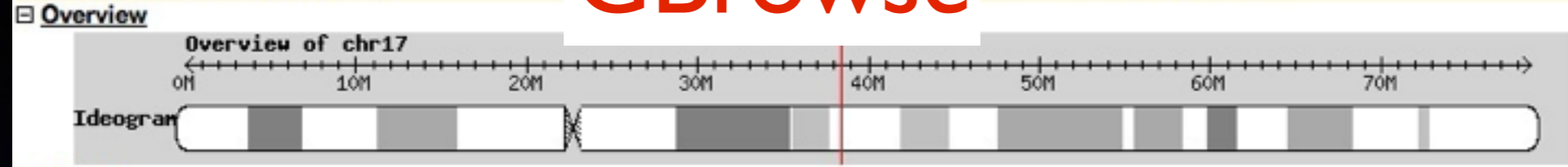
GBrowse

Landmark or Region: chr17:38350227..384002; Search

Reports & Analysis: Download GFF File Configure... Go

Data Source: Welcome to eqtl.uchicago.edu

kbp + >>> Flip



Details

Entrez genes

NM_025267
AARSD1: alanyl-tRNA synthetase domain containing 1

NM_173079
RUNDC1: RUN domain containing 1

Degner, Pai, Pique-Regi et al. (2012): $-\log_{10}(P)$, LCLs, 70 Nigerian HAPMAP ids, DNase sensitivity QTLs (dsQTLs) by

Schadt et al. (2007): $-\log_{10}(P)$, liver, 427 ids, European descent

Myers et al. (2007): $-\log_{10}(P)$, cortex from control brain, 279 ids, European descent

Stranger et al. (2007): $-\log_{10}(P)$, LCLs, 210 HAPMAP ids, 4 single populations.

Veyrieras et al. (2008): $-\log_{10}(P)$, LCLs, 210 HAPMAP ids, multi-population.

Veyrieras et al. (2008): posterior probability, LCLs, 210 HAPMAP ids, multi-population.

Pickrell et al. (2010): $-\log_{10}(P)$, LCLs, 69 Nigerian HAPMAP ids, RNA-Seq for eQTLs.

Innocenti et al. (2011): $\log_{10}(\text{Bayes Factor})$, Liver, 266 ids, RNA-Chip for eQTLs.

Pickrell et al. (2010): $-\log_{10}(P)$, LCLs, 69 Nigerian HAPMAP ids, RNA-Seq for splicing QTLs.

Montgomery et al. (2010): $-\log_{10}(P)$, LCLs, 60 European HAPMAP ids, RNA-Seq for transcript QTLs.

Zeller et al. (2010): $-\log_{10}(P)$, Monocytes, 1,490 ids recruited in Germany.

Montgomery et al. (2010): $-\log_{10}(P)$, LCLs, 60 European HAPMAP ids, RNA-Seq for exon QTLs.

Dimas et al. (2009): $-\log_{10}(P)$, Fibroblasts, 75 Europeans

Dimas et al. (2009): $-\log_{10}(P)$, LCLs, 75 Europeans

Dimas et al. (2009): $-\log_{10}(P)$, T-cells, 75 Europeans

ers

Track height: Drag/Select:

32.40 Mb .80 Mb

Ensembl

32.40 Mb 32.60 Mb 32.80 Mb

processed transcript

RNA gene

Forward strand

32.38Mb 32.40Mb

IFIT1P1-001 > processed pseudogene

BRCA2-002 > nonsense mediated decay

BRCA2-004 > processed transcript

BRCA2-005 > nonsense mediated decay

BRCA2-006 > retained intron

AL445212.9 >

AL137247.14 >

< N4BP2L1-001 protein coding

- Many “specialty” browsers (eg, Biodalliance, Savant)

Why Install Your Own?

- You have data no one else has
- You want to be able to share it with your group, community, the world (potentially with “less savvy” users)
- You want to have control over how it looks

Why JBrowse?

- (Fairly) Easy install
- Good user experience (getting close to a browser-desktop hybrid)
- Good community support (mailing lists, tutorials, software updates)
- Nice integration with iPlant

Installation

- Only requires:
 - Web server (apache, lighttpd, nginx, etc)
 - Perl/make/standard unix-y tools

JBrowse Attributes

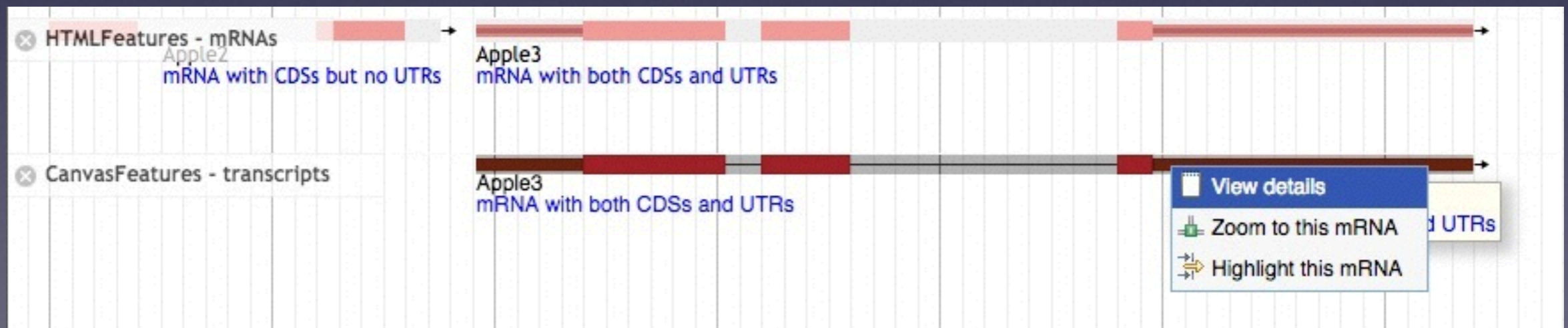
- Do everything possible on the client side, in JavaScript.
- Fast, smooth navigation.
- Supports GFF3, BED, Bio::DB::*, Chado, WIG, BAM, BigWig, VCF, and UCSC import (intron/exon structure, name lookups, quantitative plots).
- Is stably funded by NHGRI.
- Is open source, of course.
- Did I mention it's fast?

The JBrowse Project

- free and open source (license: LGPL / Artistic)
- a GMOD project
 - <http://gmod.org>
- developed using git, hosted on GitHub
 - <http://github.com/GMOD/jbrowse>
- Pls most involved: Ian Holmes, Lincoln Stein, Suzi Lewis

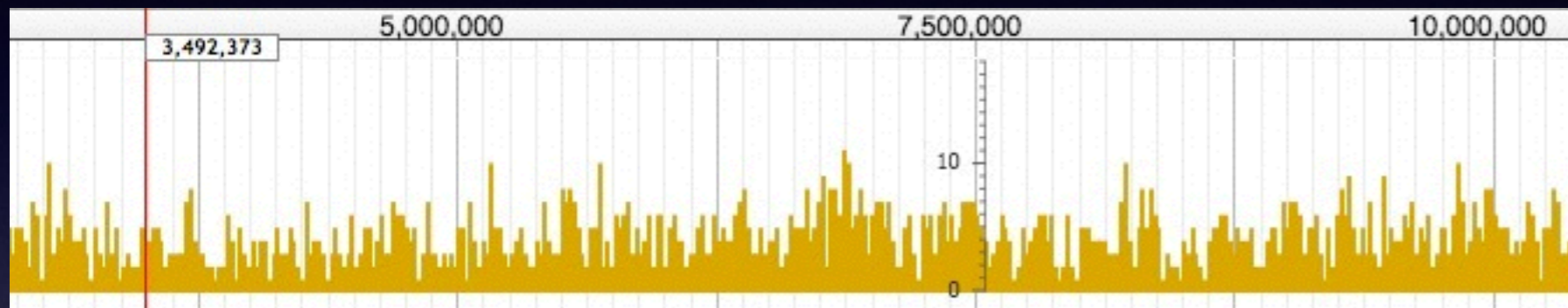
Feature Tracks

- HTMLFeatures: Rectangles (<div>s) with various fills and heights to represent the feature spans
- CanvasFeatures: Much prettier, more configurable glyphs
- Super-configurable left clicking and right-click menus.

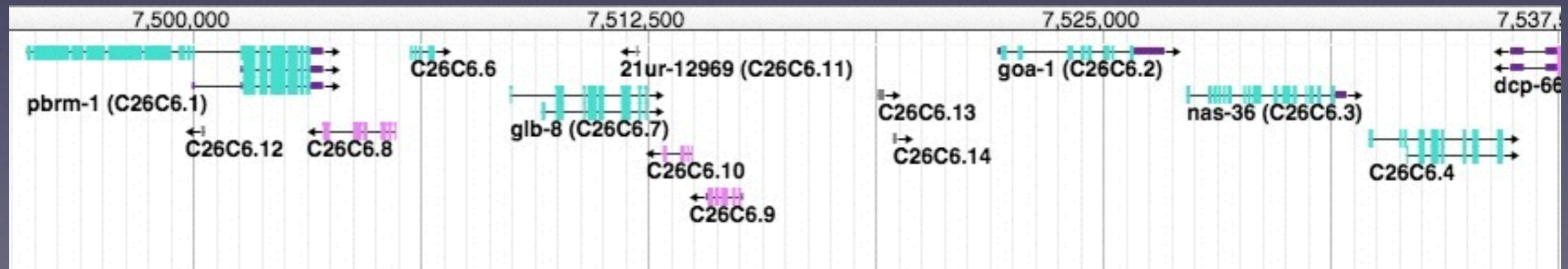


Feature Density Plots

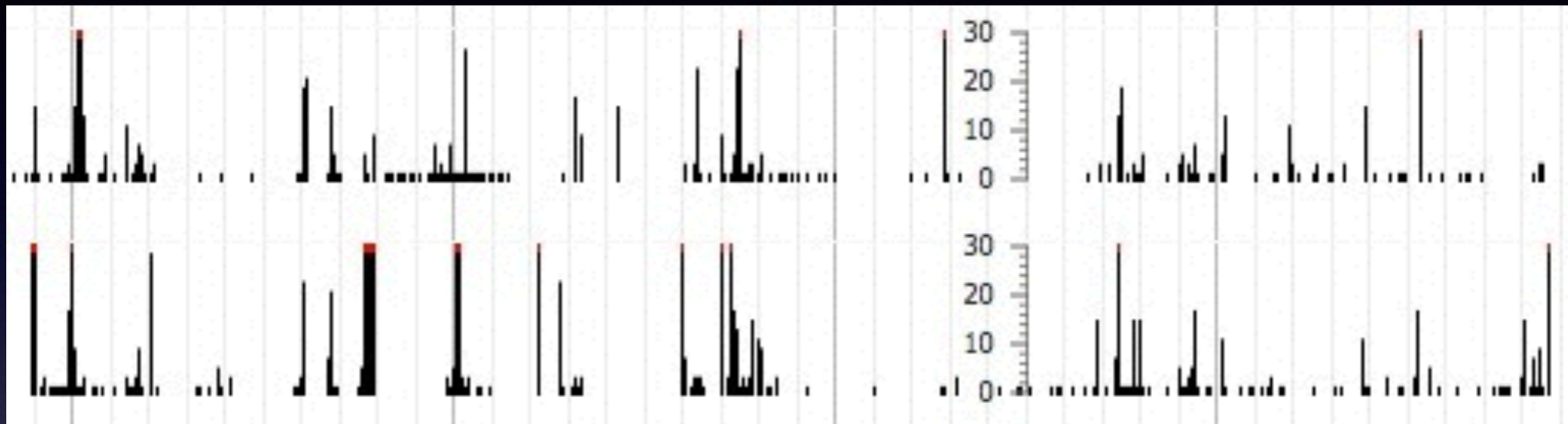
Zoomed out



Zoomed in

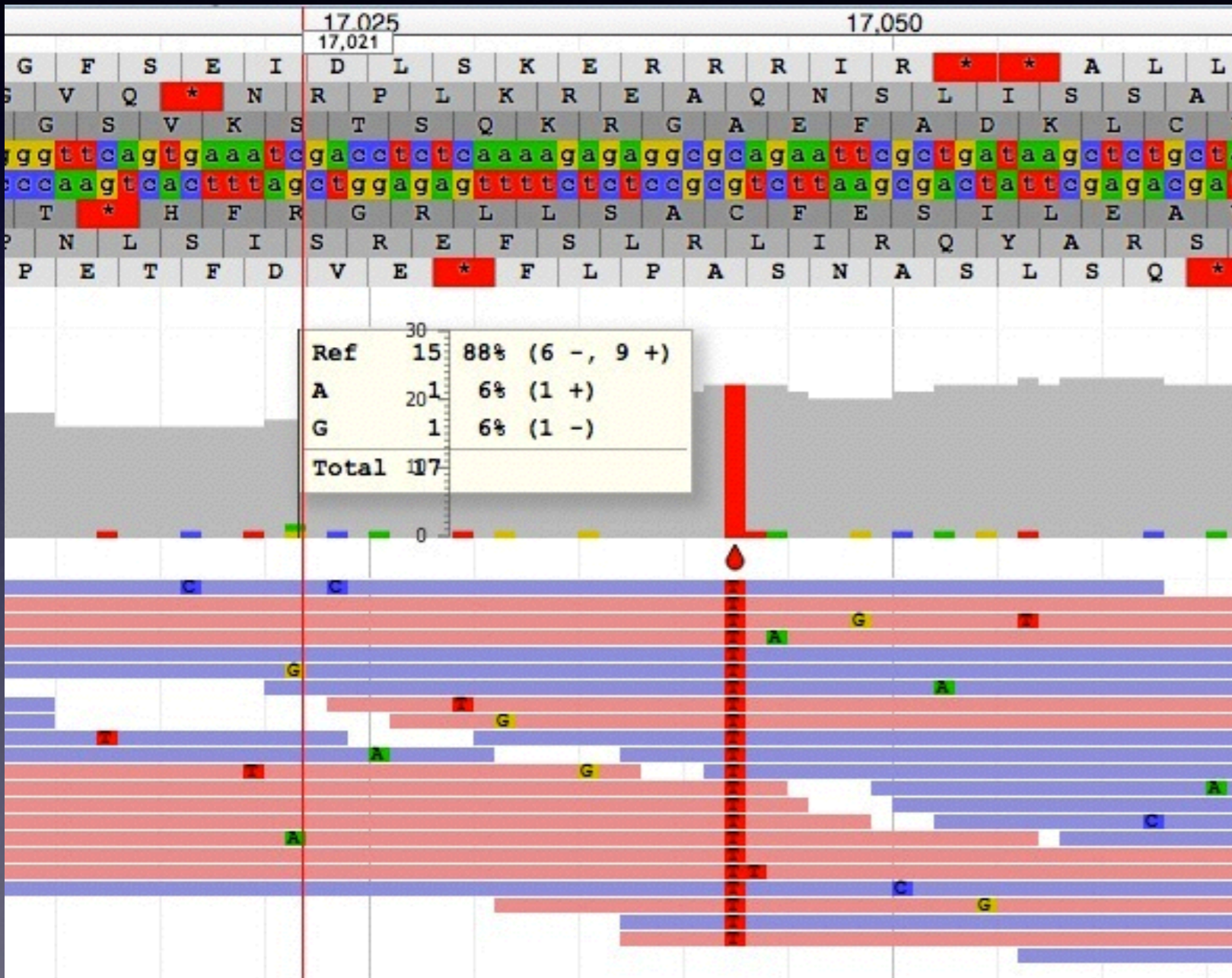


Wiggle/BigWig Tracks



- Reads small chunks directly from BigWig file.
- Needs only a not-super-old (< 5 yrs) browser, except for Internet Explorer.
- IE requires version 10.

BAM Alignment Tracks



- Reads small chunks directly from BAM file.
- Coverage and mismatches.

VCF Tracks

The screenshot displays a VCF track viewer interface. On the left, a 'Primary Data' panel shows key variant information: Type (SNV), Score (119), Description (SNV C -> T), Position (ctgA:17042..17042), and Length (1 bp). Below this, an 'Attributes' section lists various VCF fields like AC1, AF1, DP, DP4, FQ, MQ, VDB, alternative_alleles (T), description (SNV C -> T), reference_allele (C), and seq_id (ctgA). At the bottom, the 'Genotypes (1)' section contains a table summarizing variant counts and a detailed genotype table for the sample 'sample_data/raw/volv...sorted.bam'.

Primary Data

Type	SNV
Score	119
Description	SNV C -> T
Position	ctgA:17042..17042
Length	1 bp

Attributes

AC1	function values() { [native code] }
AF1	function values() { [native code] }
DP	function values() { [native code] }
DP4	function values() { [native code] }
FQ	function values() { [native code] }
MQ	function values() { [native code] }
VDB	function values() { [native code] }
alternative_alleles	T
description	SNV C -> T
reference_allele	C
seq_id	ctgA

Genotypes (1)

variant	1	100%
homozygous	1	100%
T variant	1	100%
Total	1	100%

Name	GT	PL	GQ
sample_data/raw/volv...sorted.bam	T / T	152 66 0	99

- Reads directly from VCF files compressed and indexed with bgzip and tabix.
- Shows all VCF data: alleles, genotypes, quality, etc.

Particular Strengths

- Web-based, but fast and smooth easy to set up
- Compressed NGS data: direct-to-browser BAM, BigWig, and VCF
- Optional faceted track selector – efficiently search thousands of tracks
- Open local files directly on client, no data transfer required
- Highly customizable, embeddable, integratable, programmable

WebApollo

<http://gmod.org/wiki/WebApollo>

- Based on JBrowse, using plugin system
- Next generation of the popular Apollo annotation editor
- Online annotation editing and curation!

Web  **pollo**

WebApollo

- Clients receive updates in real time (like Google Docs)
- Saves edits to a central Chado database
- Client side is a JBrowse plugin
- Extensive server-side Java
- <http://genomearchitect.org/WebApolloDemo/>

JBrowse Plugins

- Extend JBrowse with your own JavaScript code
- Can do pretty much anything
 - Add your own track visualizations
 - Add your own data backends
 - Add menu items
 - Subscribe to event notifications (pub/sub system)
 - Reach deep into the guts of JBrowse and (carefully!)
 - change anything at all!
- WebApollo client is a JBrowse plugin

Coming in 2.X

- MORE: data types, sorting options, speed
- Graphical configuration
- Multiple views, linked or independent
- Logins, uploading, track sharing
- Circular genomes

Big Thanks

Ian Holms (UC Berkeley)

Rob Buels

Mitch Skinner

Amelia Ireland

Lincoln Stein (OICR)

Julien Smith-Roberge

Erik Derohanian

Julie Moon

Natalie Fox

Adam Wright

Suzi Lewis (LBNL)

Nathan Dunn

Gregg Helt

Ed Lee

Justin Reese (UofMo)

Colin Diesh (UofMo)

NHGRI

Cold, hard cash

The End (on to the workshop)

<http://jbrowse.org/>

GMOD: <http://gmod.org/wiki/JBrowse>

Github: <http://github.com/GMOD/jbrowse>